MỤC LỤC

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG BIỂU

DANH MỤC CHỮ VIẾT TẮT

LSB	Least Significant Bit	Bit ít quan trọng nhất
DCT	Discrete Consine Transform	Phép biến đổi cosin rời rạc
IMG	Image	Ảnh đen trắng img
PCX	Personal Computer Exchange	Ảnh xám PCX
GIF	Graphics Interchange Format	Định dạng ảnh đồ họa GIF
BMP	Bitmap	Ảnh không nén Bitmap
PNG	Portable Network Graphics	Ảnh PNG
JPEG	Joint Photographic Expert Group	Ảnh nén JPEG
CCP	Close Colour Pair	Cặp màu gần nhau
SPA	Sample Pair Analysis	Phân tích cặp mẫu

MỞ ĐẦU

Ngày nay, khi Internet ngày càng phát triển mạnh mẽ và dần trở thành môi trường thế giới ảo được sử dụng trên toàn cầu. Cùng với 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 mới cho quá trình phát triển. Internet và mạng không dây đã trợ giúp cho việc chuyển phát một khối lượng thông tin rất lớn qua mạng giúp cho việc truyền thông và giao tiếp trở nên thuận lợi hơn. Tuy nhiên nó cũng làm tăng nguy cơ sử dụng trái phép, ăn cắp thông tin, xuyên tạc bất hợp pháp các thông tin được lưu chuyển trên mạng, đồng thời việc sử dụng một cách bình đẳng và an toàn các dữ liệu đa phương tiện cũng như cung cấp một cách kịp thời thông tin tới rất nhiều người dùng cuối và các thiết bị cuối cũng là một vấn đề quan trọng và còn nhiều thách thức. Hơn nữa sự phát triển của các phương tiện kỹ thuật số đã làm cho việc lưu trữ, sửa đổi và sao chép dữ liệu ngày càng đơn giản, từ đó việc bảo vệ bản quyền và chống xâm phạm trái phép các dữ liệu đa phương tiện (âm thanh, hình ảnh, tài liệu) cũng gặp nhiều khó khăn.

Một công nghệ mới được ra đời đã giải quyết phần nào một số khó khăn trên là giấu thông tin trong các nguồn đa phương tiện như các nguồn âm thanh, hình ảnh… Xét theo khía cạnh tổng quát thì giấu thông tin cũng là một hệ mật mã nhằm đảm bảo tính an toàn thông tin, những phương pháp này ưu điểm ở chỗ giảm được khả năng phát hiện ra sự tồn tại của thông tin trong các nguồn mạng. Không giống như mã hoá thông tin là để chống sự truy cập và sửa chữa một cách trái phép thông tin. Giấu và phát hiện thông tin là kỹ thuật còn tương đối mới và đang phát triển rất nhanh thu hút được sự quan tâm của cả giới khoa học và giới công nghiệp nhưng cũng còn rất nhiều thách thức.

Bản báo cáo này trình bày tổng quan về kỹ thuật giấu và phát hiện ảnh có giấu tin. Đồng thời trình bày một số kỹ thuật phát hiện thông tin giấu trên LSB của ảnh số, từ đó đưa ra các thực nghiệm và đánh giá cho việc phát hiện ảnh số có giấu tin được áp dụng.

Cấu trúc trình bày của đề án bao gồm :

- Chương I: Tổng quan kỹ thuật giấu tin và phát hiện ảnh có giấu tin.
- Chương II: Kỹ thuật phát hiện ảnh có giấu tin trên LSB
- Chương III: Cài đặt và thực nghiệm.

CHƯƠNG 1. TỔNG QUAN VỀ KỸ THUẬT GIẤU TIN VÀ PHÁT HIỆN ẢNH CÓ GIẤU TIN

1.1 TỔNG QUAN VỀ KỸ THUẬT GIẤU TIN

1.1.1 Định nghĩa kỹ thuật giấu tin

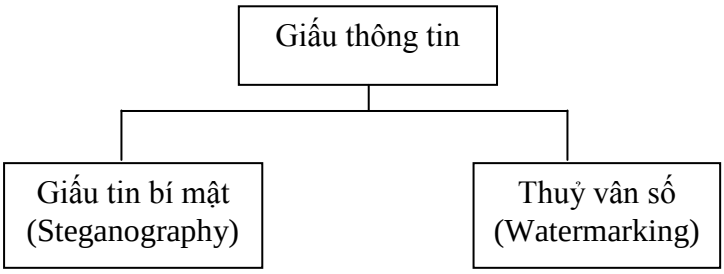
Giấu thông tin là một kỹ thuật nhúng (giấu) một lượng thông tin số nào đó vào trong một đối tượng dữ liệu số khác (giấu thông tin chỉ mang tính quy ước không phải là một hành động cụ thể).

1.1.2 Mục đích của giấu tin

Có hai mục đích của giấu tin:

- Trao đổi thông tin mật.
- Bảo đảm an toàn và phát hiện xuyên tạc thông tin cho chính các đối tượng chứa dữ liệu giấu trong đó.

Có thể thấy 2 mục đích này hoàn toàn trái ngược nhau và dần phát triển thành 2 lĩnh vực với những yêu cầu và tính chất khác nhau.



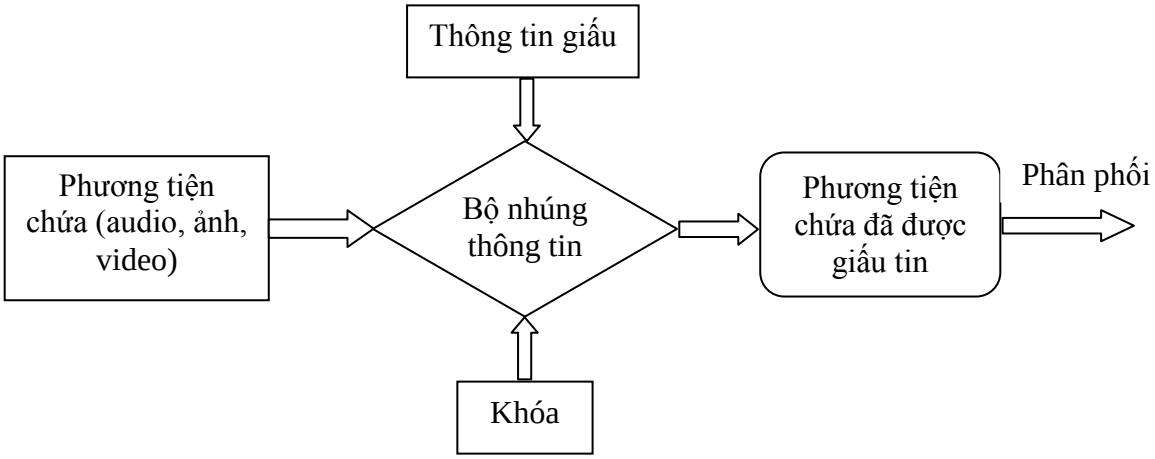
Hình 1.1. Hai lĩnh vực chính của kỹ thuật giấu thông tin

Kỹ thuật giấu thông tin bí mật (Steganography): với mục đích đảm bảo an toàn và bảo mật thông tin tập trung vào các kỹ thuật giấu tin để có thể giấu được nhiều thông tin nhất. Thông tin mật được giấu một cách vô hình trong một đối tượng khác sao cho người khác khó phát hiện được.

Kỹ thuật giấu thông tin theo kiểu đánh giấu – thủy vân (watermarking) với mục đích để bảo vệ bản quyền chính đối tượng dùng để chứa thông tin, thường tập trung đảm bảo một số các yêu cầu như đảm bảo tính bền vững… Đây là ứng dụng cơ bản nhất của kỹ thuật thuỷ vân số.

1.1.3 Mô hình kỹ thuật giấu thông tin cơ bản

Giấu thông tin vào phương tiện chứa và tách lấy thông tin là 2 quá trình trái ngược nhau và có thể mô tả qua sơ đồ khối của hệ thống như hình 1.2:



Hình 1.2 Lược đồ chung cho quá trình giấu tin

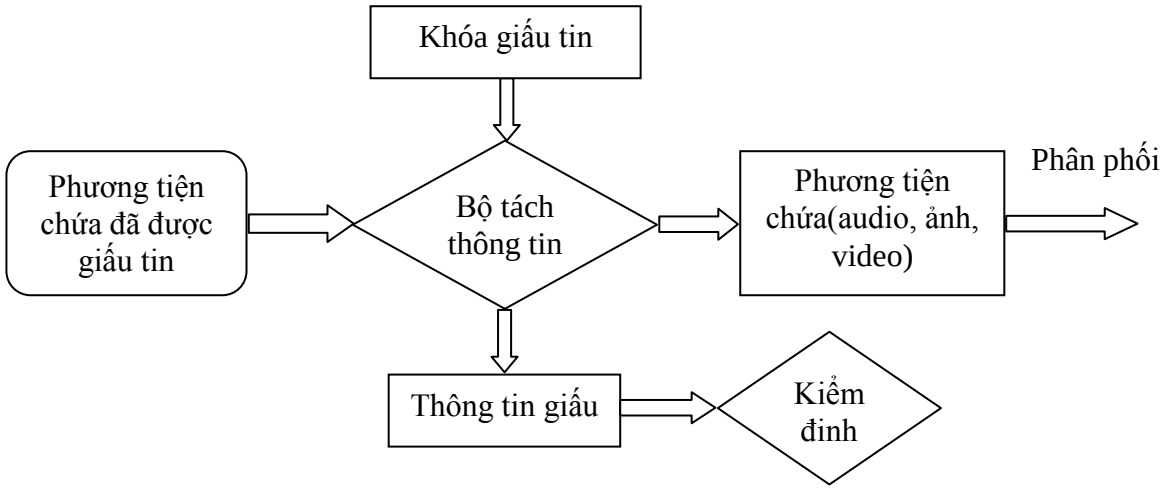
- Thông tin cần giấu tuỳ theo mục đích của người sử dụng, nó có thể là thông tin mật (với các tin bí mật) hay các logo, hình ảnh bản quyền.
- Phương tiện chứa: các file ảnh, text, audio… là môi trường để nhúng tin.

Bộ nhúng thông tin: là những chương trình thực hiện việc giấu tin

Đầu ra: là các phương tiện chứa đã có tin giấu trong đó

Tách thông tin từ các phương tiện chứa diễn ra theo quy trình ngược lại với đầu ra là các thông tin đã được giấu vào phương tiện chứa. Phương tiện chứa sau khi tách lấy thông tin có thể được sử dụng, quản lý theo những yêu cầu khác nhau.

1.1.4 Mô hình kỹ thuật tách thông tin cơ bản



Hình 1.3 Lược đồ chung cho quá trình tách thông tin

Hình 1.3 chỉ ra các công việc giải mã thông tin đã giấu. Sau khi nhận được đối tượng phương tiện chứa có giấu thông tin, quá trình giải mã được thực hiện thông qua một bộ giải mã tương ứng với bộ nhúng thông tin cùng với khoá của quá trình nhúng. Kết quả thu được gồm phương tiện chứa gốc và thông tin đã giấu. Bước tiếp theo thông tin đã giấu sẽ được xử lý kiểm định so sánh với thông tin ban đầu.

1.1.5 Yêu cầu thiết yếu đối với một hệ thống giấu tin

Có 3 yêu cầu thiết yếu đối với một hệ thống giấu tin:

- Tính vô hình: là một trong 3 yêu cầu của bất kì 1 hệ giấu tin nào.
- Tính bền vững: là yêu cầu thứ 2 của một hệ giấu tin. Tính bền vững là nói đến khả năng chịu được các thao tác biến đổi nào đó trên phương tiện nhúng và các cuộc tấn công có chủ đích.
- Khả năng nhúng: là yêu cầu thứ 3 của một hệ giấu tin. Khả năng nhúng chính là số lượng thông tin nhúng được nhúng trong phương tiện chứa.

1.1.6 Môi trường giấu tin

a. Giấu tin trong ảnh

- Giấu tin trong ảnh hiện đang rất được quan tâm. Nó đóng vai trò hết sức quan trọng trong hầu hết các ứng dụng bảo vệ an toàn thông tin như: nhận thực thông tin, xác định xuyên tạc thông tin, bảo vệ bản quyền tác giả…
- Một đặc điểm của giấu thông tin trong ảnh nữa đó là thông tin được giấu một cách vô hình, nó như là cách truyền thông tin mật cho nhau mà người khác không thể biết được bởi sau khi giấu thông tin chất lượng ảnh gần như không thay đổi đặc biệt đối với ảnh màu hay ảnh xám.

b. Giấu tin trong audio

- Khác với kỹ thuật giấu thông tin trong ảnh: phụ thuộc vào hệ thống thị giác của con người – HSV (Human Vision System), kỹ thuật giấu thông tin trong audio lại phụ thuộc vào hệ thống thính giác HAS (Human Auditory System). Bởi vì tai con người rất kém trong việc phát hiện sự khác biệt giữa các giải tần và công suất, có nghĩa là các âm thanh to, cao tần có thể che giấu đi được các âm thanh nhỏ, thấp một cách dễ dàng.
- Yêu cầu cơ bản và quan trọng nhất của giấu tin trong audio là đảm bảo tính chất ẩn của thông tin được giấu đồng thời không làm ảnh hưởng đến chất lượng của dữ liệu.

c. Giấu tin trong video

- Cũng giống như giấu thông tin trong ảnh hay trong audio, giấu tin trong video cũng được quan tâm và được phát triển mạnh mẽ cho nhiều ứng dụng như điều khiển truy cập thông tin, xác thực thông tin, bản quyền tác giả…
- Một phương pháp giấu tin trong video được đưa ra bởi Cox là phương pháp phân bố đều. Ý tưởng cơ bản của phương pháp là phân phối thông tin giấu dàn trải theo tần số của dữ liệu gốc.

d. Giấu thông tin trong văn bản dạng text

- Giấu tin trong văn bản dạng text khó thực hiện hơn do có ít các thông tin dư thừa, để làm được điều này người ta phải khéo léo khai thác các dư thừa tự nhiên của ngôn ngữ. Một cách khác là tận dụng các định dạng văn bản (mã hoá thông tin vào khoảng cách giữa các từ hay các dòng văn bản) => Kỹ thuật giấu tin đang được áp dụng cho nhiều loại đối tượng chứ không riêng dữ liệu đa phương tiện như ảnh, audio, video.

1.1.7 Một số đặc điểm của việc giấu tin trên ảnh

1.1.7.1 Tính vô hình của thông tin

Khái niệm này dựa trên đặc điểm của hệ thống thị giác của con người. Thông tin nhúng là không tri giác được nếu một người với thị giác bình thường không phân biệt được ảnh môi trường và ảnh kết quả (tức là không phân biệt được ảnh trược và sau khi giấu thông tin). Trong khi *image hiding (Steganography)* yêu cầu tính vô hình của thông tin ở mức độ cao thì *watermarking* lại chỉ yêu cầu ở một cấp độ nhất định. Chẳng hạn như người ta áp dụng watermarking cho việc gắn một biểu tượng mờ vào một chương trình truyền hình để bảo vệ bản quyền.

1.1.7.2 Khả năng nhúng tin

Lượng thông tin giấu so với kích thước ảnh môi trường cũng là một vấn đề cần quan tâm trong một thuật toán giấu tin. Rõ ràng là có thể chỉ giấu 1 bit thông tin vào mỗi ảnh mà không cần lo lắng về độ nhiễu của ảnh nhưng như vậy sẽ rất kém hiệu quả khi mà thông tin giấu có kích thước bằng Kb. Các thuật toán đều cố gắng đạt được mục đích làm thế nào giấu được nhiều thông tin nhất mà không gây ra nhiễu đáng kể.

1.1.7.3 Tính bảo mật

Thuật toán nhúng tin được coi là có tính bảo mật nếu thông tin được nhúng không bị tìm ra khi bị tấn công một cách có chủ đích trên cơ sở có hiểu biết đầy đủ về thuật toán nhúng tin và có bộ giải mã (trừ khóa bí mật), hơn nữa còn có được ảnh có mang thông tin (ảnh kết quả). Đây là một yêu cầu rất quan trọng đối với ảnh *image hiding.*

1.1.7.4 Ảnh môi trường đối với quá trình giải mã

Yêu cầu cuối cùng là thuật toán phải cho phép lấy lại được những thông tin đã giấu trong ảnh mà không có ảnh gốc. Điều này là một thuận lợi khi ảnh môi trường là duy nhất nhưng lại làm giới hạn khả năng ứng dụng của kỹ thuật giấu tin.

1.2 TỔNG QUAN VỀ KỸ THUẬT PHÁT HIỆN ẢNH CÓ GIẤU TIN

1.2.1 Khái niệm

Steganalysis là kỹ thuật phát hiện sự tồn tại của thông tin ẩn giấu trong nguồn đa phương tiện (multimedia). Giống như thám mã, mục đích của Steganalysis là phát hiện ra ảnh có mang thông tin mật và phá vỡ tính bí mật của vật mang tin ẩn.

Mục đích của kỹ thuật phát hiện là để phân loại một ảnh số bất kỳ có phải là ảnh gốc (cover image) hay ảnh có giấu tin (stego image) hay không, để từ đó có thể đưa ra bước xử lý tiếp theo.

1.2.2 Phân tích ảnh giấu tin thường dựa vào các yếu tố

- Phân tích dựa vào các đối tượng đã mang tin.

- Phân tích bằng so sánh đặc trưng: So sánh vật mang tin chưa được giấu tin với vật mang tin đã được giấu tin, đưa ra sự khác biệt giữa chúng.

- Phân tích dựa vào thông tin mật cần giấu để dò tìm.

- Phân tích dựa vào các thuật toán giấu tin và các đối tượng giấu đã biết: Kiểu phân tích này phải quyết định các đặc trưng của đối tượng giấu tin, chỉ ra công cụ giấu tin (thuật toán) đã sử dụng.

- Phân tích dựa vào thuật toán giấu tin, đối tượng gốc và đối tượng sau khi giấu tin.

1.2.3 Các phương pháp phân tích ảnh có giấu tin

- Phân tích trực quan: Thường dựa vào quan sát hoặc dùng biểu đồ tần suất (histogram) giữa ảnh gốc và ảnh chưa giấu tin để phát hiện ra sự khác biệt giữa hai ảnh căn cứ đưa ra vấn đề nghi vấn. Với phương pháp phân tích này thường khó phát hiện với ảnh có độ nhiễu cao và kích cỡ lớn.

- Phân tích theo dạng ảnh: Phương pháp này thường dựa vào các dạng ảnh bitmap hay là ảnh nén để đoán nhận kỹ thuật giấu hay sử dụng như các ảnh bitmap thường hay sử dụng giấu trên miền LSB, ảnh nén thường sử dụng kỹ thuật giấu trên các hệ số biến đổi như DCT, DWT, DFT.

- Phân tích theo thống kê: Đây là phương pháp sử dụng các lý thuyết thống kê và thống kê toán sau khi đã xác định được nghi vấn đặc trưng. Phương pháp này thường đưa ra độ tin cậy cao hơn và đặc biệt là cho tập ảnh lớn.

1.3 MỘT SỐ ẢNH ĐỊNH DẠNG BITMAP PHỔ BIẾN

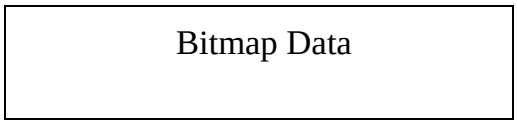
1.3.1 Cấu trúc ảnh Bitmap

Ảnh BMP (Bitmap) được phát triển bởi Microsoft Corporation, được lưu trữ dưới dạng độc lập thiết bị cho phép Windows hiển thị dữ liệu không phụ thuộc vào khung chỉ định màu trên bất kì phần cứng nào. Tên file mở rộng mặc định của một file ảnh Bitmap là ".BMP". Ảnh BMP được sử dụng trên Microsoft Windows và các ứng dụng chạy trên Windows từ version 3.0 trở lên.

Mỗi file ảnh Bitmap gồm 3 phần như bảng 1.1:

Bảng 1.1 Cấu trúc ảnh BitMap

Bitmap Header (54 byte)
Color Palette



1.3.1.1 Bitmap Header

Thành phần bitcount (Bảng 1.2) của cấu trúc Bitmap Header 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 tri "0" thì điểm ảnh là điểm đen, nếu bit mang giá trị "1" thì điểm ảnh là điểm trắng.
- 4: Bitmap là ảnh 16 màu, mỗi điểm ảnh được biểu diễn bằng 4 bit.
- 8: Bitmap là ảnh 256 màu, mỗi điểm ảnh được biểu diễn bằng 8 bit.
- 16: Bitmap là ảnh High Color, mỗi dãy 2 byte liên tiếp trong Bitmap biểu diễn cường độ tương đối của màu đỏ, xanh lá cây và xanh lơ (RGB) của điểm ảnh.
- 24: Bitmap là ảnh True Color, mỗi dãy 3 byte liên tiếp trong Bitmap biểu diễn cường độ tương đối của màu đỏ, xanh lá cây và xanh lơ (RGB) của điểm ảnh.

Thành phần Color Used của cấu trúc Bitmap Header xác định số lượng màu của Palete 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.

Bảng 1.2 Thông tin về Bitmap Header

Byte thứ	Ý nghĩa	Giá trị
1-2	Nhận dạng file	'BM' hay 19778
3-6	Kích thước file	Kiểu long trong Turbo C
7-10	Dự trữ	Thường mang giá trị 0
11-14	Byte bắt đầu vùng dữ liệu	Offset của byte bắt đầu vùng dữ liệu
15-18	Số byte cho vùng thông tin	4 byte
19-22	Chiều rộng ảnh BMP	Tính bằng pixel
23-26	Chiều cao ảnh BMP	Tính bằng pixel
27-28	Số Planes màu	Cố định là 1
29-30	Số bit cho 1 pixel (bitcount)	Có thể là: 1,4,8,16,24 tùy theo loại ảnh
31-34	Kiểu nén dữ liệu	0: Không nén

		1: Nén runlength 8bits/pixel 2: Nén runlength 4bits/pixel
35-38	Kích thước ảnh	Tính bằng byte
39-42	Độ phân giải ngang	Tính bằng pixel / metter
43-46	Độ phân giải dọc	Tính bằng pixel / metter
47-50	Số màu sử dụng trong ảnh	
51-54	Số màu được sử dụng khi hiển thị ảnh (Color Used)	

1.3.1.2 Palette màu

Bảng màu của ảnh. Chỉ những ảnh nhỏ hơn hoặc bằng 8 bit mới có bảng màu.

Bảng 1.3 Bảng màu của ảnh BITMAP

Địa chỉ (Offset)	Tên	Ý nghĩa
0	RgbBlue	Giá trị cho màu xanh blue
1	RgbGreen	Giá trị cho màu xanh Green
2	RgbRed	Giá trị cho màu đỏ
3	RgbReserved	Dự trữ

1.3.1.3 Bitmap data

Phần này nằm ngay sau phần Palete 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 trữ từ trái sang phải. Giá trị của mỗi điểm ảnh là một chỉ số trỏ tới phần tử màu tương ứng trong Palete màu.

1.3.2 Cấu trúc ảnh PNG

1.3.2.1 Lịch sử và phát triển

Động cơ thúc đẩy cho việc tạo ra định dạng PNG bắt đầu vào khoảng đầu năm 1995, sau khi Unisys công bố họ sẽ áp dụng bằng sáng chế vào thuật toán nén dữ liệu LZW- được sử dụng trong định dạng GIF. Thuật toán được bảo vệ bởi bằng công nhận độc quyền sáng tạo ở Mỹ và tất cả các nước trên thế giới. Tuy nhiên, cũng đã có một số vấn đề với định dạng GIF khi cần có một số thay đổi trên hình ảnh, nhất giới hạn của nó là 256 màu trong thời điểm máy tính có khả năng hiển thị nhiều hơn 256 màu đang trở nên phổ biến. Mặc dù định dạng GIF có thể thể hiện các hình ảnh động, song PNG vẫn được quyết định là định dạng hình ảnh đơn (chỉ có một hình duy nhất). Một người "anh em" của nó là MNG đã được tạo ra để giải

quyết vấn đề ảnh động. PNG lại tăng thêm sự phổ biến của nó vào tháng 8 năm 1999, sau khi hãng Unisys huỷ bỏ giấy phép của họ đối với các lập trình viên phần mềm miễn phí, và phi thương mại.

- Phiên bản 1.0 của đặc tả PNG được phát hành vào ngày 1 tháng 7 năm 1996, và sau đó xuất hiện với tư cách RFC 2083. Nó được tổ chức W3C khuyến nghị vào ngày 1 tháng 10 năm 1996.
- Phiên bản 1.1, với một số thay đổi nhỏ và thêm vào 3 thành phần mới, được phát hành vào ngày 31 tháng 12 năm 1998.
- Phiên bản 1.2, thêm vào một thành phần mở rộng, được phát hành vào ngày 11 tháng 8 năm 1999.
- PNG giờ đây là một chuẩn quốc tế (ISO/IEC 15948:2003), và cũng được công bố như một khuyến nghị của W3C vào ngày 10 tháng 11 năm 2003. Phiên bản hiện tại của PNG chỉ khác chút ít so với phiên bản 1.2 và không có thêm thành phần mới nào.

1.3.2.2 Thông tin kỹ thuật

a. Phần đầu của tập tin

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 hai dấu xuống dòng, ở giữa là sắp 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.

b. 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 thiết yếu, nếu không thì ít cần thiết hơn (ancillary). Chuỗi thiết yếu chứa thông tin cần thiết để đọc được tệp và nếu bộ giải mã không nhận dạng được chuỗi thiết yếu, việc đọc tệp phải được hủy.

c. Thành phần cơ bản

Một bộ giải mã (decoder) phải có thể thông dịch để đọc và hiển thị một tệp PNG.

- IHDR phải là thành phần đầu tiên, nó chứa đựng header
- PLTE chứa đựng bảng màu (danh sách các màu)
- IDAT chứa đựng ảnh. Ảnh này có thể được chia nhỏ chứa trong nhiều phần IDAT. Điều này làm tăng kích cỡ của tệp lên một ít nhưng nó làm cho việc phát sinh ảnh PNG mượt hơn (streaming manner).
- IEND đánh dấu điểm kết thúc của ảnh.

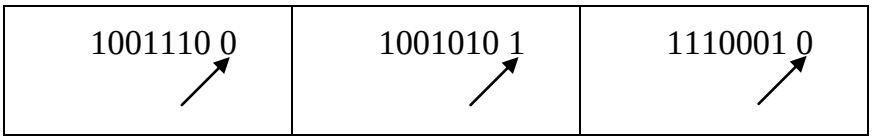
CHƯƠNG 2: KỸ THUẬT PHÁT HIỆN ẢNH CÓ GIẤU TIN TRÊN LSB

2.1 KỸ THUẬT GIẤU TIN TRÊN LSB

2.1.1 Khái niệm bit có trọng số thấp (LSB – least significant bit)

Bit có trọng số thấp là bit có ảnh hưởng ít nhất tới việc quyết định tới màu của mỗi điểm ảnh, vì vậy khi ta thay đổi bit ít quan trọng của một điểm ảnh thì màu của mỗi điểm ảnh mới sẽ tương đối gần với điểm ảnh cũ. Như vậy kỹ thuật tách bit trong xử lý ảnh được sử dụng rất nhiều trong quy trình giấu tin. Việc xác định LSB của mỗi điểm ảnh trong một bức ảnh phụ thuộc vào định dạng của ảnh và số bit màu dành cho mỗi điểm của ảnh đó. Ví dụ đối với ảnh 16 bit thì 15 bit là biểu diễn 3 màu RGB của điểm ảnh còn bit cuối cùng không dùng đến thì ra sẽ tách bit này ra ở mỗi điểm ảnh để giấu tin, hoặc với ảnh 256 màu thì bit cuối cùng trong 8 bit biểu diễn một điểm ảnh được coi là bit ít quan trọng nhất…

Ví dụ: Tách bit cuối cùng trong 8 bit biểu diễn mỗi điểm ảnh của ảnh 256 màu



Hình 2.1*:* Mỗi điểm ảnh biểu diễn bởi 8 bit, bit cuối cùng được coi là bit ít quan trọng nhất tức là bit bên phải nhất

Trong phép tách này ta coi bit cuối cùng là bit ít quan trọng nhất, thay đổi giá trị của bit này thì sẽ thay đổi giá trị của điểm ảnh lên hoặc xuống đúng một đơn vị, ví dụ như giá trị điểm ảnh là 234 thì khi thay đổi bit cuối cùng nó có thể mang giá trị mới là 235 nếu đổi bit cuối cùng từ 0 thành 1. Với sự thay đổi nhỏ đó ta hi vọng là cấp độ màu của điểm ảnh sẽ không bị thay đổi nhiều

2.1.2 Thuật toán giấu một chuỗi thông tin mật trên LSB

2.1.2.1 Ý tưởng thuật toán

+ Cho thông tin mật nhúng W, W có thể là:
  - Một chuỗi bit thông tin mật (vd: W = [0 0 1 1 0 1 0 1 1 1 1 0 0 1]).
  - Một chuỗi các kí tự (vd: W = HPU → phải đổi W sang hệ nhị phân).
+ Đổi W ra hệ nhị phân, tính độ dài của thông tin mật W sau đó thực hiện thay thế các bit thông tin mật W cần giấu vào các bit có giá trị thấp (LSB) của ảnh cho đến khi bit thông tin mật cần giấu không còn nữa thì ngừng.
+ Ảnh thu được là ảnh có giấu thông tin vào tất cả các bit LSB của ảnh lần lượt từ trái qua phải, từ trên xuống dưới.

2.1.2.2 Thuật toán giấu

Đầu vào:

Ảnh cover và thông tin mật cần nhúng.

Đầu ra:

Ảnh có giấu tin.

Các bước thực hiện:

B1: Chuyển dữ liệu ảnh sang mảng 2 chiều

B2: Đổi thông tin mật sang chuỗi nhị phân (bit)

B3: Thay thế các bit thông tin mật vào các bit có giá trị thấp (LSB) của ảnh đến khi các bit thông tin mật không còn nữa thì ngừng.

2.1.3 Thuật toán giấu thông tin mật theo tỷ lệ trên LSB

2.1.3.1 Ý tưởng thuật toán

+ Cho tỷ lệ p% (so với kích cỡ của ảnh) thông tin mật cần giấu, tạo một ma trận ngẫu nhiên các bit nhị phân có kích thước bằng p% ảnh cần giấu.
+ Thực hiện thay thế các bit thông tin mật trong ma trận ngẫu nhiên vào các bit có giá trị thấp (LSB) của ảnh cho đến khi bit thông tin mật trong ma trận không còn nữa thì ngừng.
+ Ảnh thu được là ảnh có giấu p% thông tin của ảnh vào tất cả các bit LSB của ảnh lần lượt từ trái qua phải, từ trên xuống dưới.

2.1.3.2 Thuật toán giấu

Đầu vào :

Ảnh cover và tỷ lệ p% thông tin mật cần nhúng.

Đầu ra :

Ảnh có giấu tin.

Các bước thực hiện :

B1: Chuyển dữ liệu ảnh sang mảng 2 chiều M*N

B2: Tính kích thước ma trận ngẫu nhiên cần tạo ra:

L=p*M*N/100

B3: Tạo một ma trận các bit nhị phân ngẫu nhiên có số hàng M và số cột
R=L/M

B4: Thay thế lần lượt các bit thông tin mật trong ma trận ngẫu nhiên vào các bit có giá trị thấp (LSB) của ảnh theo quy tắc từ trái sang phải từ trên xuống cho đến khi các bit thông tin mật trong ma trận ngẫu nhiên không còn thì dừng.

2.2 KỸ THUẬT PHÁT HIỆN ẢNH CÓ GIẤU TIN TRÊN LSB

2.2.1 Phương pháp phân tích cặp màu gần nhau (CCP - close colour pair)

2.2.1.1 Tổng quan về thuật toán

Trong ảnh tự nhiên 24-bit, mỗi điểm ảnh được đại diện bởi ba kênh màu (Red, Green và Blue), mỗi kênh rộng là 8 bit. Hầu hết các phương pháp che giấu thông tin trong một ảnh tự nhiên là dựa vào việc thay thế các LSB của các kênh màu bằng các bit thông tin. Như vậy, trung bình một nửa LSB được thay đổi và nó giả

định rằng nhúng thông tin theo cách này sẽ không ảnh hưởng các thông tin của ảnh cover. Giả định này là đúng nếu và chỉ nếu số lượng màu đặc biệt trong ảnh cover có thể so sánh với tổng số điểm ảnh trong ảnh.

Tuy nhiên quan sát cho thấy, trong một ảnh tự nhiên, tỷ lệ của số lượng màu đặc biệt với tổng số điểm ảnh là khoảng 01:06. Do đó sau khi nhúng LSB, mô hình LSB ngẫu nhiên sẽ tăng lên. Điều này tương đương với việc số lượng các cặp màu đặc biệt tăng lên, và được sử dụng làm dấu hiệu để phân biệt các loại ảnh.

Cặp màu gần nhau (P) và cặp màu đặc biệt (U) được định nghĩa như sau:
- Hai màu (R1, G1, B1) và (R2, G2, B2) là gần nhau nếu:

$$|R1 - R2| = 1, |G1 - G2| = 1 \text{ và } |B1 - B2| = 1$$

Hoặc

$$(R1 - R2)^2 + (G1 - G2)^2 + (B1 - B2)^2 <= 3 \qquad (2.2.1.1)$$

- Hai màu (R3, G3, B3) và (R4, G4, B3) là đặc biệt nếu bất kỳ một trong những điều sau đây là đúng sự thật :

$$|R1 - R2| = 1 \text{ hoặc } |G1 - G2| = 1 \text{ hoặc } |B1 - B2| = 1 \qquad (2.2.1.2)$$

Tỷ lệ R là một ý tưởng về số lượng tương đối của các cặp màu gần nhau với màu đặc biệt:

$$R=P/U \qquad (2.2.1.3)$$

Quan sát cho thấy rằng, đối với một hình ảnh không có bất kỳ thông tin nhúng, giá trị của R là lớn hơn so với một hình ảnh trong đó có thông tin đã được nhúng trong nó. Điều này xảy ra khi nhúng thông tin, làm tăng số lượng của cặp màu đặc biệt (U).

Sau khi thử nghiệm với các loại ảnh khác nhau, quan sát những giá trị thử nghiệm cho phép chúng ta phân biệt một ảnh giả mạo từ một ảnh không giả mạo. Đó là, nếu có một ảnh thử nghiệm đã được nhúng thông tin, tiếp tục nhúng thêm các bit thông tin giá trị R thay đổi không đáng kể. Cách khác, nếu một ảnh thử nghiệm là không giả mạo, tỷ lệ R giảm đáng kể khi tiếp tục nhúng thêm các bit thông tin. Chúng ta tạo ảnh thử nghiệm thông qua một phần mềm nhúng tin trên LSB. Nếu U' và P' là số lượng các cặp màu đặc biệt và các cặp màu gần nhau tương ứng sau đó.

Tỷ lệ R' thu được:

$$R'=P'/U' \qquad (2.2.1.4)$$

Nếu ảnh đã có một thông tin mật lớn ẩn trong nó, hai tỷ lệ này sẽ là gần như bằng nhau R = R'. Nhưng nếu ảnh không có nhúng thông tin mật thì dự kiến R'>R. Sự thay đổi trong tỷ lệ được đo bằng m, trong đó m là tỷ lệ phần trăm thay đổi trong R được đưa ra :

$$m=((R-R')*100)/R \qquad (2.2.1.5)$$

m có thể được coi là ngưỡng để phân biệt một ảnh.

2.2.1.2 Thuật toán phát hiện ảnh có giấu tin CCP

Đầu vào:

Ảnh màu 24-bit C.

Đầu ra:

Phân loại ảnh C là ảnh stego hay cover.

Các bước thực hiện:

B1: Tạo ra một ảnh stego C’ bằng cách nhúng thông tin vào ảnh C bằng kỹ thuật nhúng trên LSB với tỷ lệ thông tin nhúng là 20% so với kích cỡ của ảnh C.

B2: Tính tổng số các cặp màu đặc biệt U và các cặp màu gần nhau P trong ảnh C theo phương trình (2.2.1.1) và (2.2.1.2).

B3: Tính tỷ lệ R = P/U.

B4: Tính tổng số các màu đặc biệt U’ và cặp màu gần nhau P’ trong ảnh C’ theo phương trình (2.2.1.1) và (2.2.1.2).

B5: Tính tỷ lệ R’ =P’/U’.

B6: Tính giá trị m =(R-R’)*100/R.

B7: Tính tỷ lệ β = R/R’.

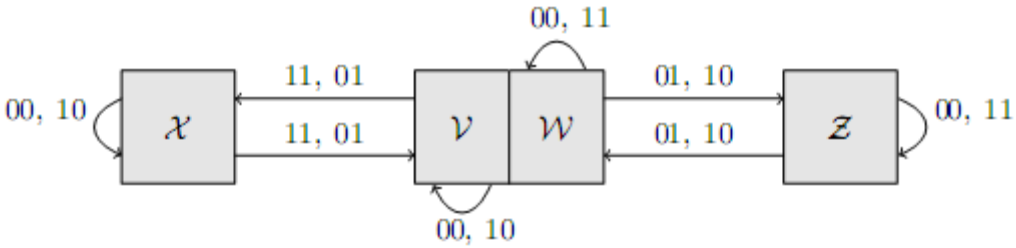
B8: Nếu (β<1) hoặc (m< δ) “Ảnh là Stego” Ngược lại “Ảnh là Cover”.

2.2.2 Phương pháp phân tích cặp mẫu (SPA - sample pair analysis)

2.2.2.1 Tổng quan về thuật toán

Bài viết thực hiện phân tích nhúng thông tin trên LSB một cách tốt hơn thực hiện bởi Wu [4] và tiếp tục cải thiện bởi Lu [3]. Ý tưởng của phương pháp này là để phát triển một biện pháp bảo vệ các ảnh tự nhiên và phát hiện các hình ảnh stego được tạo ra bởi các thuật toán giấu LSB. Điều này được thực hiện bằng cách phân tích các giá trị trên cặp điểm ảnh.

Chúng ta bắt đầu bằng cách chia ảnh thành từng cặp điểm ảnh lân cận và biểu diễn bằng cặp (u, v)$\in$P, trong đó P là tập hợp của tất cả các cặp điểm ảnh trong hình ảnh. Kích thước của P thiết lập là n/2, trong đó n là số lượng điểm ảnh. Tiếp theo, chúng ta chia P thành ba tập con P = X$\cup$Y$\cup$Z.



Hình 2.2 Sơ đồ phân tích sự chuyển đổi của các cặp điểm ảnh

dựa trên định nghĩa ta có :

$(u, v) \in X \Leftrightarrow (u < v$ và v là lẻ) hoặc $(u > v$ và v là chẵn)

$(u, v) \in Y \Leftrightarrow (u < v$ và v là chẵn) hoặc $(u > v$ và v là lẻ)

$(u, v) \in Z \Leftrightarrow u = v$ (2.2.2.1)

và tiếp tục phân chia tập $Y = V \cup W$, trong đó :

$(u, v) \in W \Leftrightarrow (u, v) = (2k, 2k + 1)$ hoặc $(u, v) = (2k + 1, 2k)$

$(u, v) \in V \Leftrightarrow (u, v) \notin W$ (2.2.2.2)

Mặc dù định nghĩa có vẻ phức tạp, các tập có tính chất quan trọng được chứng minh bằng cách sử dụng hình 2.2. Trong hình này, sơ đồ quá trình chuyển đổi có thể được giải thích theo cách sau. Các cặp điểm ảnh $(u,v) \in X$ có thể thay đổi tập, nếu mô hình LSB thay đổi là 11, hoặc 01 (cả hai điểm ảnh hoặc chỉ có điểm ảnh thứ hai được thay đổi). Giả sử mô hình 11 từ định nghĩa của tập X chúng ta có thể thấy rằng v là lẻ hoặc v là chẵn. Khi v lẻ $(u > v)$, sau đó bằng cách đảo LSB của số lẻ, chúng ta nhận được số nhỏ hơn là số chẵn và bằng cách đảo LSB của u, chúng ta có thể có được $v + 1$, do đó bất đẳng thức $u > v$ vẫn giữ và chuyển đổi (u,v) thuộc W. Sử dụng một phương pháp tiếp cận tương tự, chúng ta có thể chứng minh sơ đồ chuyển đổi hoàn toàn. Một khía cạnh quan trọng mà chúng ta có thể nhìn thấy từ biểu đồ, đó là bộ $X \cup V$ và $W \cup Z$ để nhúng LSB tùy ý.

Để thể hiện độ dài tin nhắn tương đối α sử dụng cho một ảnh stego, chúng ta sử dụng X, Y, V, W, Z để biểu thị các bộ được định nghĩa từ ảnh cover và X', Y', V', W', Z' để biểu thị các bộ tính toán từ ảnh stego. Mục tiêu của chúng ta là chính xác α trong bộ nguyên tố, bởi vì các bộ này có thể được tính toán. Khi chúng ta nhúng thông tin ngẫu nhiên, mỗi điểm ảnh truy cập được thay đổi trong quá trình nhúng, do đó khả năng thấy các mô hình thay đổi 11 và 00 trong các ảnh stego là $(\alpha/2)^2$, $(1-\alpha/2)^2$ tương ứng. Sử dụng kết quả này và sơ đồ chuyển đổi từ hình 2.2, chúng ta có thể tính kích thước dự kiến của bộ X', V', W':

$|X'| = |X|(1 - \alpha/2) + |V| \alpha/2$ (2.2.2.3)

$|V'| = |V|(1 - \alpha/2) + |X| \alpha/2$ (2.2.2.4)

$|W'| = |W|(1 - \alpha + \alpha^2/2) + |Z| \alpha (1 - \alpha/2)$ (2.2.2.5)

Đối với ảnh tự nhiên, không có lý do tại sao kích thước của bộ X và Y khác nhau. Do đó, chúng ta có :

$|X| = |Y| \Rightarrow |X| = |V| + |W|$ (2.2.2.6)

trừ phương trình (2.2.2.3) và (2.2.2.4) chúng ta có được :

$|X'| - |V'| = (|X| - |V|)(1 - \alpha)$ (2.2.2.7)

Khi chúng ta thay thế phương trình (2.2.2.6), chúng ta có thể viết lại phương trình cuối cùng :

$|X'| - |V'| = |W|(1 - \alpha)$ (2.2.2.8)

Ở đây, chúng ta phải tìm một phương trình cho $|W|$. Sử dụng (2.2.2.5), ta có thể viết:

$$|W'| = |W|(1 - \alpha + \alpha^2/2) + |Z|\,\alpha(1- \alpha/2)$$
$$= |W|(1 - \alpha + \alpha^2/2) + (\gamma - |W|)\,\alpha(1 - \alpha/2)$$
$$= |W|(1 - \alpha)^2 + \gamma\alpha(1 - \alpha/2)$$

Trong đó $\gamma = |W|+|Z| = |W'|+|Z'|$ là một giá trị đã biết. Cuối cùng, bằng cách thay thế (2.2.2.8) vào phương trình cuối cùng, chúng ta có được phương trình sau đây tính tương đối chiều dài tin nhắn α.

$$1/2\ \gamma\alpha^2 + (2|X'| - |P|)\alpha + |Y'| - |X'| = 0 \qquad (2.2.2.9)$$

Tất cả các hệ số có thể được tính toán từ các hình ảnh stego. Để có được ước tính chính xác của α tham số, chúng ta đã lấy một phần nhỏ thực sự từ phương trình (2.2.2.9).

2.2.2.2 Thuật toán phát hiện ảnh giấu tin SPA

Đầu vào: 1 ảnh cấp xám cần kiểm tra.

Đầu ra: Chính xác chiều dài tin nhắn α.

Các bước thực hiện:

B1: Tính tổng số cặp mẫu trong ảnh P = (M*N/2)

B2: Tính kích thước của mỗi bộ cặp mẫu *X, Y, W, Z* theo (2.2.2.1) và (2.2.2.2).

B3: Giải phương trình (2.2.2.9).

p1=(-2*X+P+sqrt((2*X-P)^2-2*(W+Z)*(Y-X)))/(W+Z)

p2=(-2*X+P-sqrt((2*X-P)^2-2*(W+Z)*(Y-X)))/(W+Z)

B4: Chính xác độ dài tin nhắn α

α=max(0,min(p1,p2))

CHƯƠNG 3: CÀI ĐẶT VÀ THỰC NGHIỆM

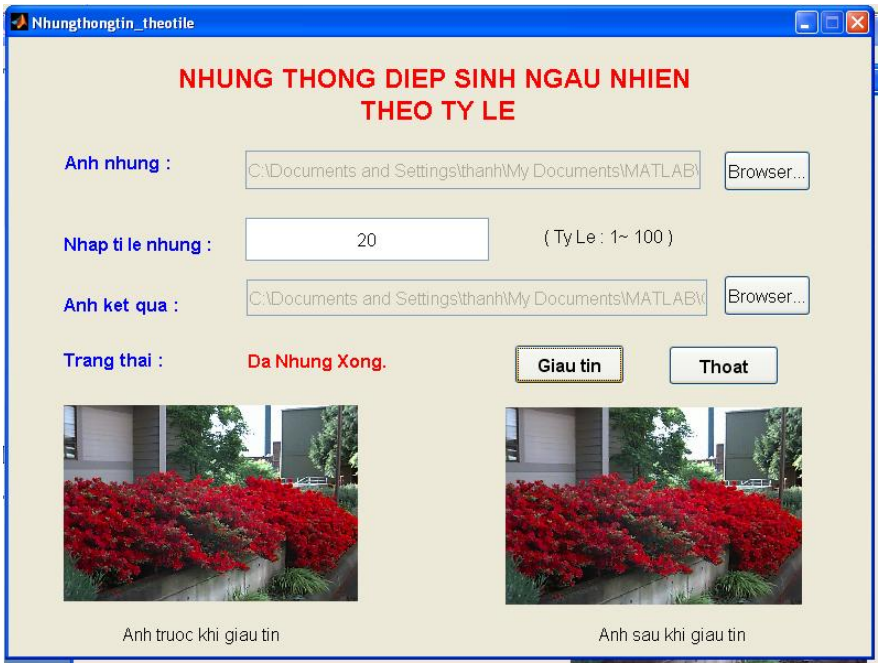
3.1 MÔI TRƯỜNG CÀI ĐẶT

- Ngôn ngữ cài đặt, môi trường soạn thảo và chạy chương trình được thực hiện trên ngôn ngữ lập trình Matlab 7.8.0 (R2009a).
- Hệ điều hành Window XP và môi trường NetFarme Work 2.0
- Yêu cầu cấu hình:

3.2 GIAO DIỆN CHƯƠNG TRÌNH

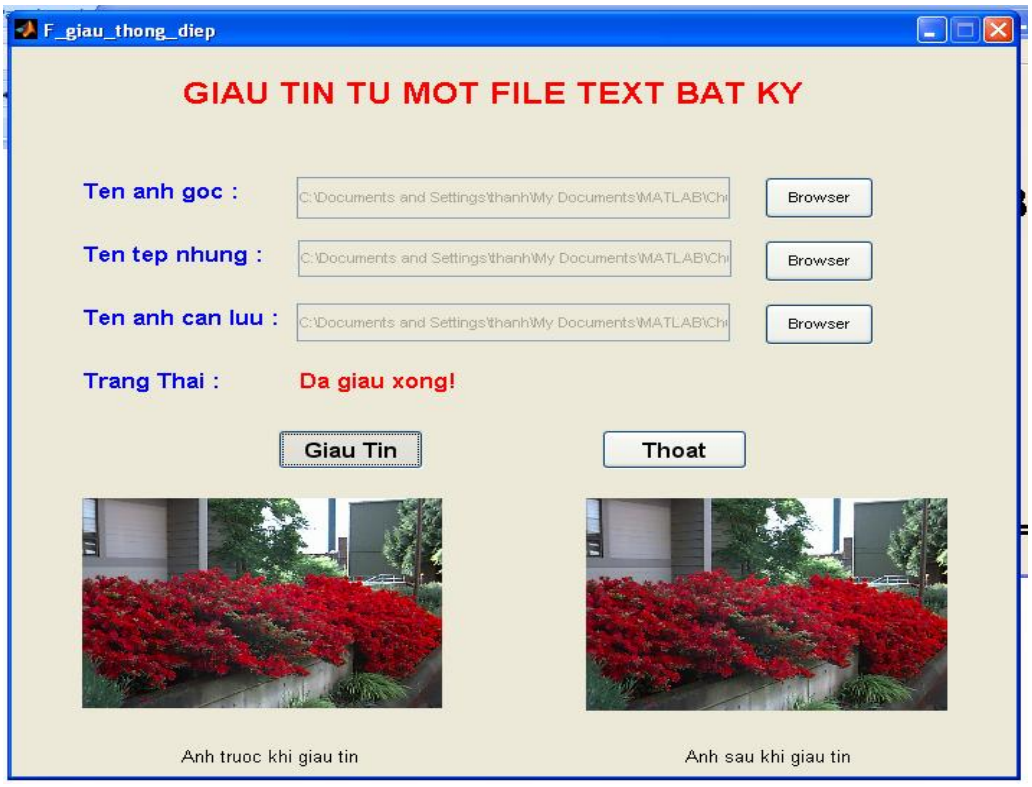
- Giao diện chính :

- Giao diện giấu theo tỷ lệ ảnh :

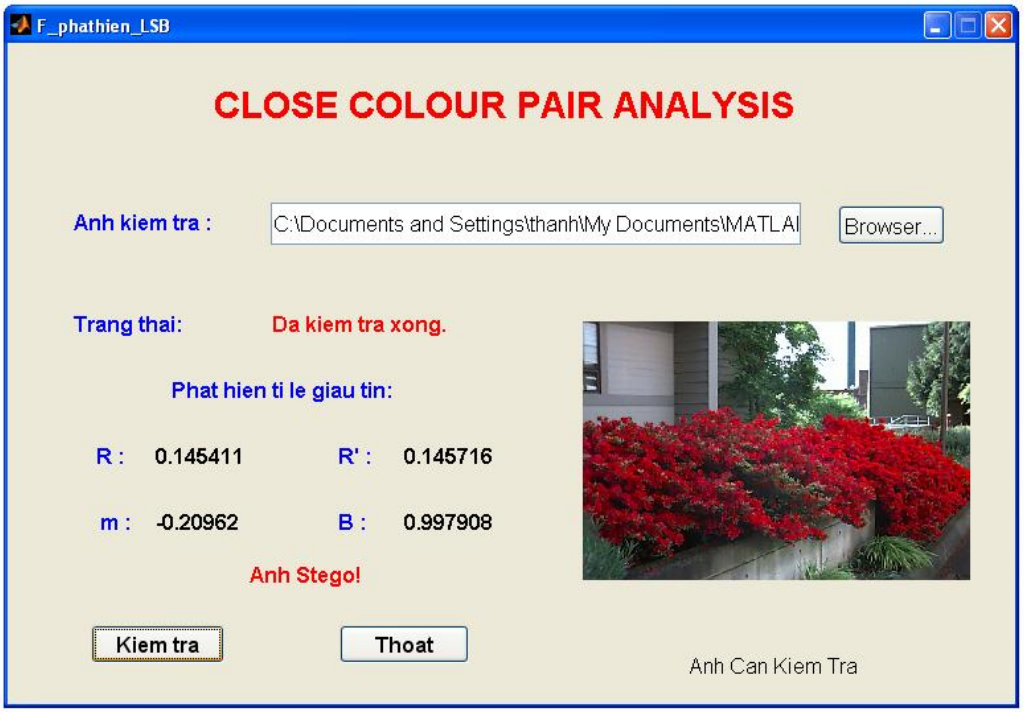


- Giao diện giấu chuỗi ký tự :

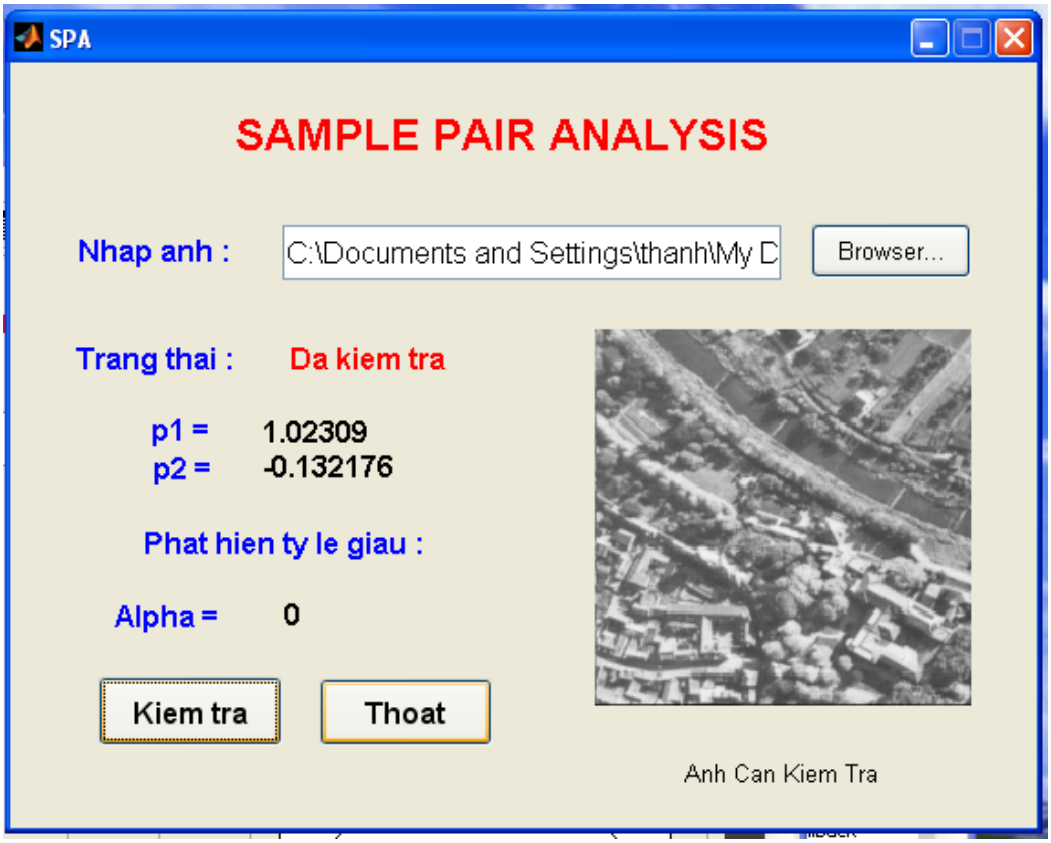
- Giao diện giấu tệp văn bản :



- Giao diện phát hiện (CCP) :



- Giao diện phát hiện (SPA) :



3.3 ĐÁNH GIÁ KỸ THUẬT PHÁT HIỆN THEO F-MEASURE

Trong những thử nghiệm này, em sử dụng các độ đo đánh giá là: *precision, recall và f-measure* thường được áp dụng trong phân loại dữ liệu. *Precision* là độ đo tính chính xác và đúng đắn của việc phân loại. *Recall* là độ đo tính toàn vẹn của việc phân lớp.

Cụ thể cho bài toán phân loại ảnh có giấu tin và ảnh chưa giấu tin, giả sử ta có một tập ảnh đầu vào E (gồm cả ảnh nhúng tin và ảnh chưa nhúng tin) cần phân thành 2 tập con E_1 (ảnh không nhúng tin) và E_2 (ảnh có nhúng tin). Sau khi thực hiện phân lớp chúng ta được bảng sau:

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Kết quả phân lớp đúng</td></tr>
<tr><td>E_1</td><td>E_2</td></tr>
<tr><td rowspan="2">Kết quả phân lớp đạt được</td><td>E_1</td><td>tp
(true positive)</td><td>fp
(false positive)</td></tr>
<tr><td>E_2</td><td>fn
(false negative)</td><td>tn
(true negative)</td></tr>
</table>

Khi đó precision và recall được tính toán theo công thức sau:

Precision = tp/(tp + fp) (3.3.1)

Recall = tp/(tp + fn) (3.3.2)

Mặc dù *precsion* và *recall* là những độ đo được dùng rộng rãi và phổ biến nhất, nhưng chúng lại gây khó khăn khi phải đánh giá các bài toán phân loại vì hai độ đo trên lại không tăng/giảm tương ứng với nhau. Bài toán đánh giá có *recall* cao có thể có *precision* thấp và ngược lại. Hơn nữa, việc so sánh mà chỉ dựa trên một mình precision và *recall* không phải là một ý hay. Với mục tiêu này, độ đo *F-measure* được sử dụng để đánh giá tổng quát các bài toán phân loại. *F-measure* là trung bình điều hoà có trọng số của *precision* và *recall* và có công thức:

$$F_\beta = (1+\beta^2)\frac{precision * recall}{\beta^2 * precision + recall}$$

trong đó β là một tham số có giá trị nằm giữa 0 và 1. Nếu β = 1, *F-measure* bằng với precision và nếu β = 0, *F-measure* bằng với *recall*. Giữa đoạn đó, giá trị β càng cao, độ quan trọng của *precision* càng cao so với *recall*. Ta sử dụng giá trị thường được dùng là β = 0.5, nghĩa là:

$$F = 2\frac{precision * recall}{precision + recall} \qquad (3.3.3)$$

3.4 KẾT QUẢ THỬ NGHIỆM

3.4.1 Kết quả thử nghiệm phương pháp cặp màu gần nhau (CCP)

3.4.1.1 Tập ảnh thử nghiệm

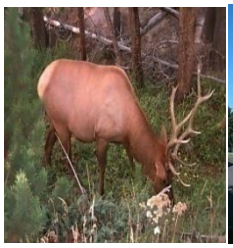

1.jpg


2.jpg


3.jpg


4.jpg


5.jpg


6.jpg


7.jpg


8.jpg


9.jpg


10.jpg


11.jpg


12.jpg


13.jpg


14.jpg

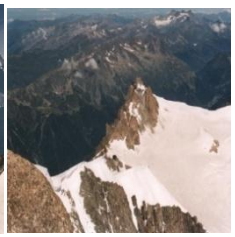

15.jpg


16.jpg


17.jpg


18.jpg


19.jpg

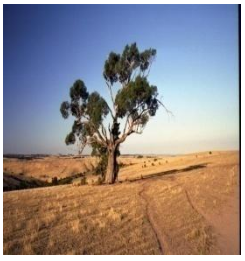

20.jpg


21.jpg


22.jpg


23.jpg


24.jpg


25.jpg


26.jpg


27.jpg


28.jpg


29.jpg


30.jpg


34.jpg


35.jpg


36.jpg


37.jpg

3.4.1.2 Thử nghiệm và đánh giá kết quả

Bảng 3.1 kết quả thử nghiệm của 31 hình ảnh và các hình ảnh được nhúng trên LSB với tỷ lệ tương ứng 20% và 50%.

Tên ảnh (*.jpg)	R	R’	m=(R-R1)/R*100	B=R/R’	Loại ảnh
1	0.2079	0.1453	30.1168	1.4310	Cover
20_1	0.1458	0.1446	0.8241	1.0083	Stego
50_1	0.1002	0.1003	-0.0539	0.9995	Stego
2	0.3794	0.2819	25.6903	1.3457	Cover
20_2	0.2819	0.2830	-0.3836	0.9962	Stego
50_2	0.1868	0.1877	-0.4777	0.9952	Stego
3	0.2479	0.1984	19.9794	1.2497	Cover
20_3	0.1980	0.1989	-0.4645	0.9954	Stego
50_3	0.1517	0.1515	0.1362	1.0014	Stego
4	0.3112	0.2245	27.8547	1.3861	Cover
20_4	0.2239	0.2246	-0.3312	0.9967	Stego
50_4	0.1559	0.1562	-0.1745	0.9983	Stego
5	0.3515	0.2661	24.2817	1.3207	Cover
20_5	0.2661	0.2653	0.3244	1.0033	Stego
50_5	0.1686	0.1688	-0.1519	0.9985	Stego
6	0.1781	0.1406	21.0918	1.2673	Cover
20_6	0.1408	0.1408	0.0405	1.0004	Stego

50_6	0.0906	0.0900	0.6041	1.0061	Stego
7	0.3351	0.2301	31.3222	1.4561	Cover
20_7	0.2288	0.2300	0.2300	0.9950	Stego
50_7	0.1362	0.1374	-0.8882	0.9912	Stego
8	0.2920	0.2041	30.0944	1.4305	Cover
20_8	0.2033	0.2042	-0.4138	0.9959	Stego
50_8	0.1327	0.1316	0.8057	1.0081	Stego
9	0.2187	0.1661	24.0634	1.3169	Cover
20_9	0.1675	0.1652	1.4020	1.0142	Stego
50_9	0.1361	0.1357	0.2805	1.0028	Stego
10	0.3544	0.2360	33.4171	1.5019	Cover
20_10	0.2363	0.2362	0.0680	0.0680	Stego
50_10	0.1361	0.1367	-0.4510	0.9955	Stego
11	0.2905	0.2505	13.7706	1.1597	Cover
20_11	0.2501	0.2499	0.0607	1.0006	Stego
50_11	0.1871	0.1877	-0.3298	0.9967	Stego
12	0.1296	0.1149	11.3032	1.1274	Cover
20_12	0.1144	0.1147	-0.2225	0.9978	Stego
50_12	0.0913	0.0913	0.0901	1.0009	Stego
13	0.1702	0.1298	23.7247	1.3110	Cover
20_13	0.1279	0.1290	-0.8400	0.9917	Stego
50_13	0.0977	0.0968	0.9036	1.0091	Stego
14	0.1767	0.1510	14.5771	1.1706	Cover
20_14	0.1503	0.1506	-0.2106	0.9979	Stego
50_14	0.1216	0.1206	0.8321	1.0084	Stego
15	0.3655	0.2667	27.0360	1.3705	Cover
20_15	0.2662	0.2653	0.3365	1.0034	Stego
50_15	0.1469	0.1467	0.0893	1.0009	Stego
16	0.7919	0.4450	43.8059	1.7795	Cover
20_16	0.4462	0.4472	-0.2283	0.9977	Stego
50_16	0.2491	0.2483	0.3303	1.0033	Stego
17	0.3673	0.2634	28.2951	1.3946	Cover

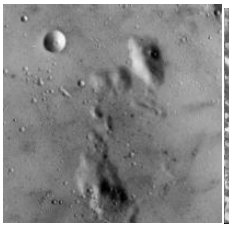
20_17	0.2634	0.2643	-0.3351	0.9967	Stego
50_17	0.1768	0.1770	-0.1222	0.9988	Stego
18	0.6194	0.4103	33.7608	1.5097	Cover
20_18	0.4118	0.4098	0.4829	1.0049	Stego
50_18	0.2349	0.2347	0.0799	1.0008	Stego
19	0.8479	0.5675	33.0710	1.4941	Cover
20_19	0.5678	0.5683	-0.0763	0.9992	Stego
50_19	0.3628	0.3622	0.1654	1.0017	Stego
20	0.6833	0.5540	18.9188	1.2333	Cover
20_20	0.5542	0.5527	0.2781	1.0028	Stego
50_20	0.3399	0.3397	0.0658	1.0007	Stego
21	0.2214	0.1689	23.7275	1.3111	Cover
20_21	0.1695	0.1693	0.1418	1.0014	Stego
50_21	0.1222	0.1219	0.2738	1.0027	Stego
22	0.2127	0.1941	8.7380	1.0957	Cover
20_22	0.1947	0.1935	0.6111	1.0061	Stego
50_22	0.1628	0.1630	-0.1455	0.9985	Stego
23	0.2625	0.2085	20.5575	1.2588	Cover
20_23	0.2072	0.2078	-0.2834	0.9972	Stego
50_23	0.1477	0.1473	0.2444	1.0025	Stego
24	0.6176	0.3325	46.1676	1.8576	Cover
20_24	0.3303	0.3311	-0.2596	0.9974	Stego
50_24	0.2200	0.2182	0.8200	1.0083	Stego
25	0.7126	0.3901	45.2556	1.8267	Cover
20_25	0.3906	0.3909	-0.0690	0.9993	Stego
50_25	0.1696	0.1693	0.1900	1.0019	Stego
26	0.6128	0.3448	43.7352	1.7773	Cover
20_26	0.3461	0.3453	0.2244	1.0022	Stego
50_26	0.1939	0.1949	-0.4781	0.9952	Stego
27	0.4246	0.4246	38.6453	1.6299	Cover
20_27	0.2599	0.2605	-0.2486	0.9975	Stego
50_27	0.1268	0.1277	-0.7405	0.9926	Stego

28	0.4749	0.2343	50.6658	2.0270	Cover
20_28	0.2349	0.2351	-0.1160	0.9988	Stego
50_28	0.1268	0.1277	-0.7405	0.9926	Stego
29	0.5552	0.3067	44.7570	1.8102	Cover
20_29	0.3072	0.3059	0.4117	1.0041	Stego
50_29	0.2321	0.2339	-0.7676	0.9924	Stego
30	0.4651	0.3779	18.7332	1.2305	Cover
20_30	0.3777	0.3771	0.1564	1.0016	Stego
50_30	0.2237	0.2240	-0.1563	0.9984	Stego
31	0.3204	0.2313	27.8195	1.3854	Cover
20_31	0.2319	0.2313	0.2730	1.0027	Stego
50_31	0.1617	0.1629	-0.7403	0.9927	Stego
32	0.4146	0.2603	37.2231	1.5929	Cover
20_32	0.2605	0.2621	-0.6300	0.9937	Stego
50_32	0.1784	0.1778	0.3065	1.0031	Stego
33	0.4048	0.1896	53.1626	2.1350	Cover
20_33	0.1925	0.1925	0.1925	0.1925	Stego
50_33	0.1275	0.1270	0.3451	1.0035	Stego
34	0.5714	0.3313	42.0227	1.7248	Cover
20_34	0.3306	0.3320	-0.4140	0.9959	Stego
50_34	0.2898	0.2896	0.0702	1.0007	Stego

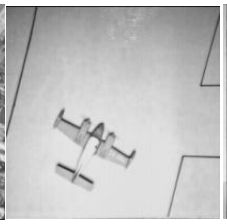
Thuật toán thử nghiệm với 31 hình ảnh cover theo bảng 3.1 cho thấy sự thay đổi giữa giá trị R và R’ (khi nhúng 20% tin nhắn thử nghiệm) là rất lớn, cho nên giá trị của ngưỡng m để so sánh được tính cho ảnh cover cũng lớn. Đối với ảnh stego_20% và stego_50% thì sự thay đổi của giá trị R và R’ là không đáng kể có khi R’>R. Qua đó chúng ta thấy thuật toán áp dụng cho ảnh có giấu tin với tỷ lệ càng lớn thì càng dễ phát hiện.

3.4.2 Kết quả thử nghiệm phương pháp cặp mẫu (SPA)

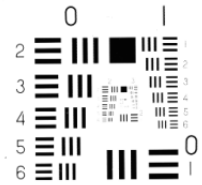
3.4.2.1 Tập ảnh thử nghiệm


1.png


2.png


3.png


4.png



5.png

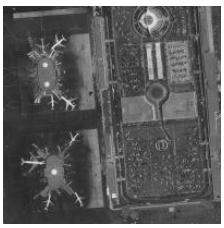

6.png


7.png


8.png


9.png


10.png


11.png


12.png


13.png


14.png


15.png


16.png


17.png


18.png


19.png


20.png


21.png

3.4.2.2 Kết quả thử nghiệm và đánh giá

Bảng 3.2 Kết quả thử nghiệm cho 21 ảnh xám với ảnh nhúng LSB tỷ lệ 20% và 50%

Tên ảnh (*.png)	(p1,p2)	Tỷ lệ nhúng α	Loại ảnh
1	(1.0474,-0.0452)	0	Cover
20_1	(1.0690,0.2593)	0. 2593	Stego_20%
50_1	(1.0463,0.6592)	0. 6592	Stego_50%
2	(1.0231,-0.1322)	0	Cover
20_2	(0.9974,0.1253)	0. 1253	Stego_20%
50_2	(0.9365,0.4453)	0. 4453	Stego_50%
3	(1.0876,-0.0049)	0	Cover
20_3	(1.0961,0.1402)	0. 1402	Stego_20%
50_3	(1.1358,0.4012)	0. 4012	Stego_50%
4	(1.1396,-0.0114)	0	Cover
20_4	(1.1532,0.1797)	0. 1797	Stego_20%
50_4	(1.1132,0.5503)	0. 5503	Stego_50%
5	(1.9857,0.0280)	0. 0280	Cover
20_5	(1.8573,0.1522)	0. 1522	Stego_20%
50_5	(1.6965,0.3207)	0. 3207	Stego_50%
6	(0.9988,0.0105)	0. 0105	Cover
20_6	(1.0339,0.2006)	0. 2006	Stego_20%
50_6	(0.9442,0.4993)	0. 4993	Stego_50%
7	(1.0912,-0.1505)	0	Cover
20_7	(1.0747,0.1139)	0. 1139	Stego_20%
50_7	(1.0467,0.2894)	0. 2894	Stego_50%
8	(1.0779,0.0078)	0. 0078	Cover
20_8	(1.0769,0.2368)	0. 2368	Stego_20%
50_8	(1.0197,0.4645)	0. 4645	Stego_50%
9	(1.0988,0.1558)	0. 1558	Cover
20_9	(1.1156,0.1845)	0. 1845	Stego_20%
50_9	(1.0416,0.6875)	0. 6875	Stego_50%
10	(1.0762,0.1392)	0. 1392	Cover

20_10	(1.0947,0.2744)	0. 2744	Stego_20%
50_10	(1.0479,0.4577)	0. 4577	Stego_50%
11	(1.0368,-0.0807)	0	Cover
20_11	(1.0139,0.1123)	0. 1123	Stego_20%
50_11	(1.0360,0.4661)	0. 4661	Stego_50%
12	(1.0572,0.0488)	0. 0488	Cover
20_12	(1.0663,0.1621)	0. 1621	Stego_20%
50_12	(1.0747,0.3100)	0. 3100	Stego_50%
13	(1.1572,-0.0067)	0	Cover
20_13	(1.1562,0.1643)	0. 1643	Stego_20%
50_13	(1.1564,0.4157)	0. 4157	Stego_50%
14	(0.9955,0.0750)	0. 0750	Cover
20_14	(0.9910,0.3120)	0. 3120	Stego_20%
50_14	(1.0156,0.5127)	0. 5127	Stego_50%
15	(1.0525,0.0425)	0. 0425	Cover
20_15	(1.0387,0.1983)	0. 1983	Stego_20%
50_15	(1.0555,0.5161)	0. 5161	Stego_50%
16	(1.0473,-0.2122)	0	Cover
20_16	(1.0612,0.0916)	0. 0916	Stego_20%
50_16	(1.0444,0.4139)	0. 4139	Stego_50%
17	(1.0385,-0.1147)	0	Cover
20_17	(1.0433,0.0524)	0. 0524	Stego_20%
50_17	(1.0822,0.4061)	0. 4061	Stego_50%
18	(0.9913,0.0098)	0. 0098	Cover
20_18	(1.0059,0.2655)	0. 2655	Stego_20%
50_18	(1.0000,0.3578)	0. 3578	Stego_50%
19	(1.0487,-0.0286)	0	Cover
20_19	(1.0348,0.0961)	0. 0961	Stego_20%
50_19	(1.0515,0.3905)	0. 3905	Stego_50%
20	(1.0622,-0.0015)	0	Cover
20_20	(1.0702,0.3354)	0. 3354	Stego_20%
50_20	(1.0183,0.5873)	0. 5873	Stego_50%

21	(1.0021,-0.1091)	0	Cover
20_21	(1.0047,0.2136)	0. 2136	Stego_20%
50_21	(0.9574,0.5381)	0. 5381	Stego_50%

Đánh giá kỹ thuật phát hiện theo *F-measure*. Sau khi thực hiện phân lớp trên hai tập thử nghiệm E_20% và E_50% ta được kết quả như bảng 3.3 và bảng 3.4.

Bảng 3.3 Tổng hợp kết quả từ bảng 3.2 của tập thử nghiệm E_20%

		Kết quả phân lớp đúng	
		E1	E2
Kết quả phân lớp đạt được	E1	19	3
	E2	1	20

Áp dụng công thức (3.3.1) và (3.3.2) và (3.3.3) ta có:

$$\text{Precision} = \frac{19}{19+3} = 0.90$$

$$\text{Recall} = \frac{19}{19+1} = 0.95$$

$$\text{F-measure} = 2 * \frac{0.9 * 0.95}{0.9 + 0.95} = 0.92$$

Bảng 3.4 Tổng hợp kết quả từ bảng 3.2 của tập thử nghiệm E_50%

		Kết quả phân lớp đúng	
		D1	D2
Kết quả phân lớp đạt được	D1	19	3
	D2	0	21

Áp dụng công thức (3.3.1), (3.3.2) và (3.3.3) ta có:

$$\text{Precision} = \frac{19}{19+3} = 0.9$$

$$\text{Recall} = \frac{29}{29+0} = 1$$

$$\text{F-measure} = 2 * \frac{0.9*1}{0.9+1} = 0.95$$

Bảng 3.5 Bảng thử nghiệm trên hai tập ảnh E_20% và E_50%

Độ đo / Kỹ thuật	Precision	Recall	F-measure
Kỹ thuật phát hiện cho lượng giấu 20%	0.90	0.95	0.92
Kỹ thuật phát hiện cho lượng giấu 50%	0.90	1	0.95

Thông qua thử nghiệm và thuật toán đánh giá F-measure cho phương pháp phân tích cặp mẫu. Ta thấy phương pháp phát hiện này được đánh giá rất cao, với ảnh có tỷ lệ thông tin nhúng càng lớn thì càng dễ phát hiện. Phương pháp này có thể đưa ra ước lượng chính xác tỷ lệ thông tin nhúng trong ảnh nên có thể áp dụng cho việc phát hiện những ảnh có tỷ lệ nhúng rất nhỏ. Theo tài liệu [1] thông điệp nhỏ nhất có thể phát hiện một cách đáng tin cậy bằng cách sử dụng phương pháp phân tích cặp mẫu là khoảng α = 0,05 cho hình ảnh gốc và thậm chí còn nhỏ hơn đối với các nguồn ảnh khác. Qua đó ta thấy rằng kết quả của thuật toán phát hiện ảnh có giấu tin trên LSB này là rất cao và chúng ta không nên tránh sử dụng phương pháp này.

KẾT LUẬN

Phát hiện thông tin mật ẩn giấu trong dữ liệu đa phương tiện, đặc biệt là trong ảnh số là một vấn đề đang được quan tâm hiện nay trong nhiều lĩnh vực. Để phát hiện và phân biệt một ảnh số nào đó có mang tin mật hay không đòi hỏi rất nhiều yếu tố và kỹ thuật phức tạp.

Trong đồ án này đã đưa ra một cái nhìn tổng quan về giấu tin trên miền LSB và phát hiện ảnh có giấu tin sử dụng kỹ thuật giấu LSB.

Trong thời gian làm đồ án em đã nghiên cứu được những vấn đề sau:

- Nghiên cứu tổng quan kỹ thuật giấu tin trong ảnh.
- Nghiên cứu cấu trúc ảnh BITMAP và PNG.
- Tìm hiểu chi tiết kỹ thuật giấu tin trên miền LSB của ảnh.
- Nghiên cứu kỹ thuật phát hiện ảnh có giấu tin trên LSB.
- Cài đặt và thử nghiệm bằng matlab 7.8.0(R2009a).

Trong quá trình làm đồ án, do kiến thức còn thiếu sót, hạn chế về thời gian nên việc nghiên cứu đề tài không thể tránh khỏi những thiếu sót. Rất mong nhận được sự đóng góp ý kiến của các thầy, cô và toàn thể các bạn đồng môn để báo cáo của em được hoàn thiện hơn.

Em xin chân thành cảm ơn!

TÀI LIỆU THAM KHẢO

[1] A. Ker. "*A general framework for structural analysis of LSB replacement*". In Proceedings 7th Information Hiding Workshop, Barcelona, Spain, June 6–8, 2005.

[2] Jhonson, N. F and Jajodia, S. "*Steganalysis of Images Using Current Steganography Software*", 2nd International Workshop on Information Hiding, April 1988.

[3] P. Lu, X. Luo, Q. Tang, and L. Shen. "*An improved sample pairs method for detection of LSB embedding*". In J. Fridrich, editor, Information Hiding, 6th International Workshop, volume 3200 of LNCS, pages 116–127. Springer-Verlag, Berlin, 2004.

[4] S. Dumitrescu, X. Wu, and Z. Wang. "*Detection of lsb steganography via sample pair analysis*". In IH' 02: Revised Papers from the 5th International Workshop on Information Hiding, pages 355–372, London, UK, 2003. Springer-Verlag.