BỘ GIÁO DỤC VÀ ĐÀO TẠO
TR- ỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG
-------O0O-------

ISO 9001 : 2008

BÁO CÁO KHOA HỌC

NGÀNH CÔNG NGHỆ THÔNG TIN

TÌM HIỂU KỸ THUẬT PHÁT HIỆN THÔNG TIN GIẤU TRONG ẢNH GIF

Chủ nhiệm đề tài: Trịnh Thị Thu Hà
Thành viên: Mạc Như Hiển

Giáo viên h□ớng dẫn: ThS. Hồ Thị H□ơng Thơm

HẢI PHÒNG 08-2009

BÁO CÁO KHOA HỌC

Đề tài: Tìm hiểu kỹ thuật phát hiện thông tin giấu trên ảnh GIF

Chủ nhiệm đề tài: ***Trịnh Thị Thu Hà*** **Lớp CT901**
Thành viên: ***Mạc Như Hiển*** **Lớp CT901**

Giáo viên hướng dẫn: ***Ths. Hồ Thị Hương Thơm***

MỤC LỤC

LỜI CẢM ƠN

Chúng em xin chân thành cảm ơn hội đồng khoa Công Nghệ Thông Tin, hội đồng khoa học trường Đại Học Dân Lập Hải Phòng đã tạo điều kiện để chúng em thực hiện tốt đề tài nghiên cứu khoa học này.

Chúng em xin chân thành cảm ơn cô giáo: Ths. Hồ Thị Hương Thơm – giảng viên khoa công nghệ thông tin trường ĐHDL Hải Phòng, đã tận tình hướng dẫn và chỉ đạo chúng em trong suốt thời gian nghiên cứu đề tài.

Cuối cùng, chúng mình xin cảm ơn tất cả các bạn đồng môn đã động viên, góp ý và trao đổi hỗ trợ cho chúng mình trong suốt thời gian nghiên cứu vừa qua.

Vì thời gian nghiên cứu chỉ có hạn, trình độ hiểu biết của bản thân chúng em còn nhiều hạn chế. Cho nên trong đề tài không tránh khỏi những thiếu sót, chúng em rất mong được sự góp ý quý báu của tất cả các thầy cô giáo cũng như các bạn để đề tài của chúng em được hoàn thiện hơn.

Chúng em xin chân thành cảm ơn!

Hải Phòng, ngày 10 tháng 08 năm 2009

Nhóm thực hiện

Trịnh Thị Thu Hà

Mạc Như Hiển

GIỚI THIỆU

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. Tuy nhiên nó cũng làm tăng nguy cơ sử dụng trái phép, 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 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 phần nào giải quyết được các khó khăn trên là giấu thông tin trong các nguồn đa phương tiện như các nguồn âm thanh, hình ảnh, ảnh tĩnh…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 về giấu và phát hiện ảnh có giấu tin. Đồng thời trình bày một số kỹ thuật giấu và phát hiện thông tin trên ảnh GIF, từ đó đưa ra các thực nghiệm và đánh giá cho việc phát hiện thông tin ẩn giấu trong ảnh GIF.

CHƯƠNG I. TỔNG QUAN VỀ KỸ THUẬT GIẤU TIN VÀ GIẤU TIN TRONG ẢNH

1.1 Định nghĩa kỹ thuật giấu tin

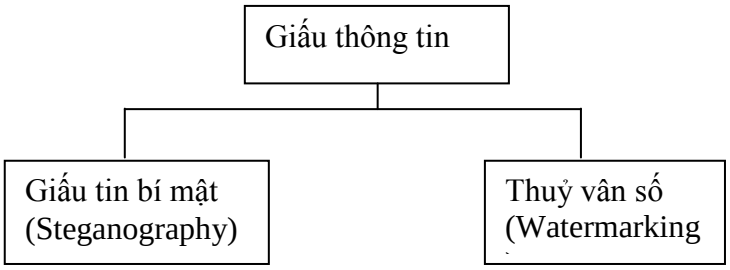
Giấu thông tin là một kỹ thuật nhúng (giấu) một lượng thông tin số nào đó vào trong một đối tượng dữ liệu số khác (giấu thông tin chỉ mang tính quy ước không phải là một hành động cụ thể).

1.2 Mục đích của giấu tin

Có hai mục đích của giấu tin:

- Bảo mật cho những dữ liệu được giấu
- Bảo đảm an toàn (bảo vệ bản quyền) 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: Hai lĩnh vực chính của kỹ thuật giấu thông tin

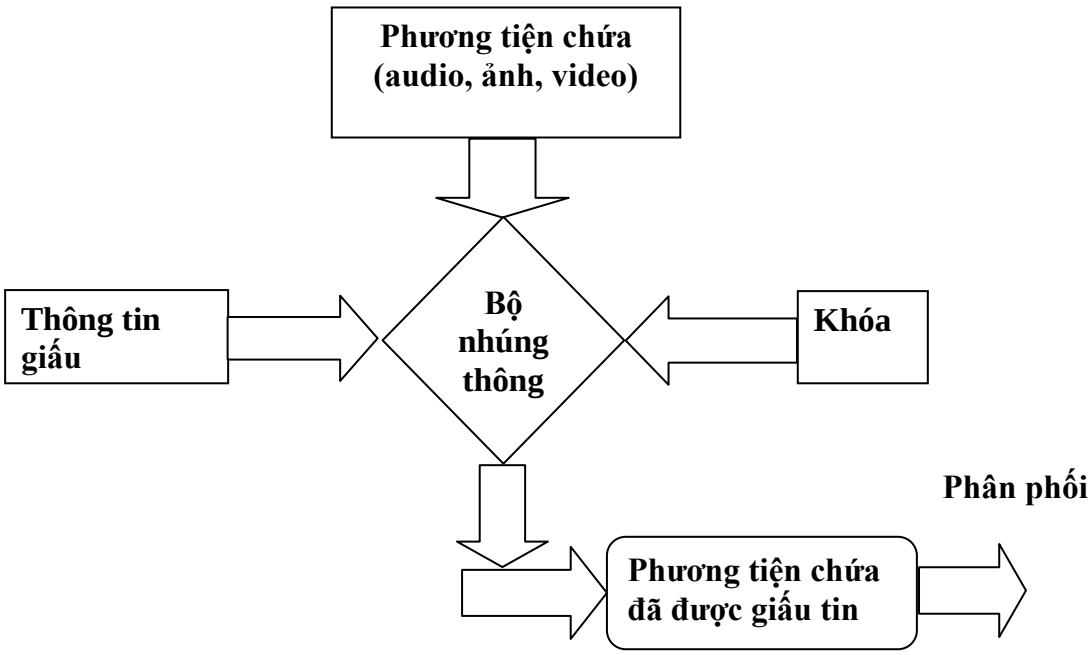
Kỹ thuật giấu thông tin bí mật (Steganography): với mục đích đảm bảo an toàn và bảo mật thông tin tập trung vào các kỹ thuật giấu tin để có thể giấu được nhiều thông tin nhất. Thông tin mật được giấu kỹ trong một đối tượng khác sao cho người khác không phát hiện được.

Kỹ thuật giấu thông tin theo kiểu đánh giấu (watermarking) để bảo vệ bản quyền của đối tượng chứa thông tin thì lại 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.

1.2.1 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 sau:



Hình 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 điệp (với các tin bí mật) hay các logo, hình ảnh bản quyền.

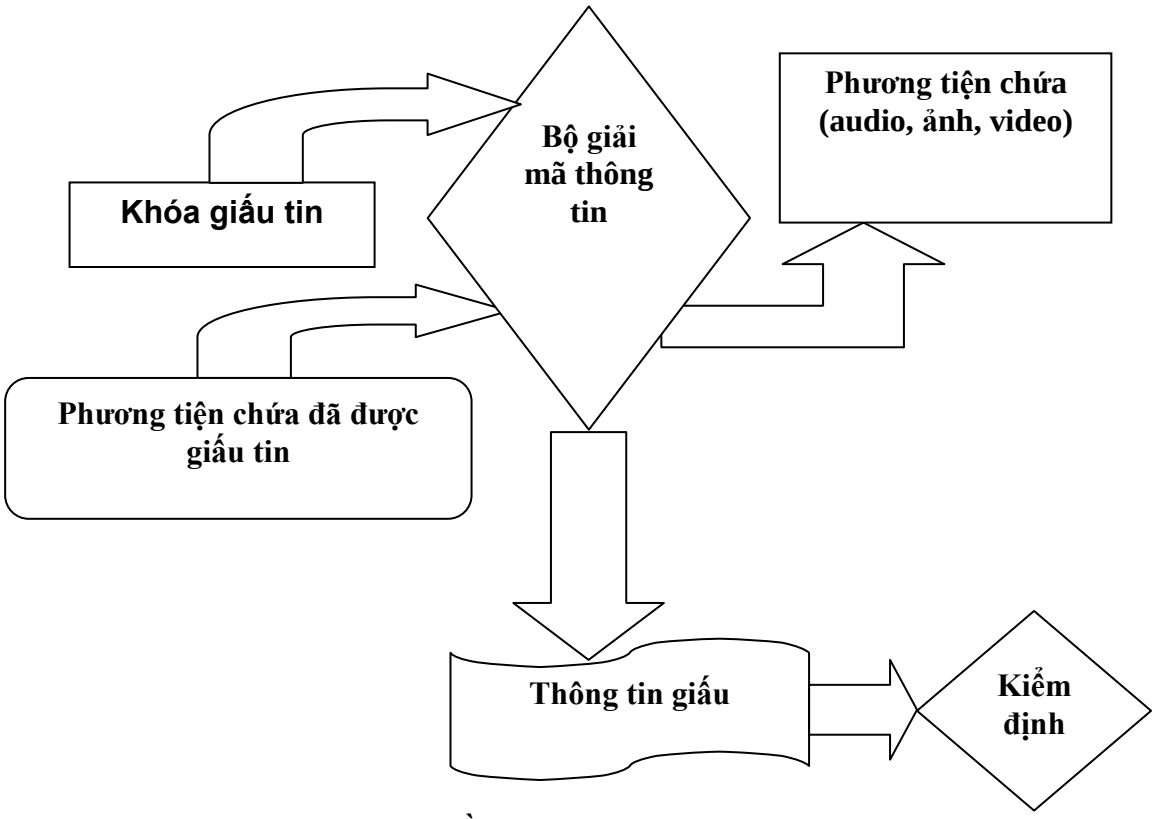
- Phương tiện chứa: các file ảnh, text, audio… là môi trường để nhúng tin.

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.

1.2.2 Mô hình kỹ thuật tách thông tin cơ bản



Hình 3: Lược đồ chung cho quá trình giải mã thông tin

Hình vẽ trên 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.

1.3. Môi trường giấu tin

1.3.1 Giấu tin trong ảnh

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.

1.3.2. Giấu tin trong audio

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.

1.3.3. 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, nhận thức 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.

1.3.4 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.

CHƯƠNG II. ẢNH GIF

2.1 Cấu trúc ảnh GIF

Ảnh GIF (Graphics Interchange Format) là một định dạng tập tin hình ảnh bitmap cho các hình ảnh dùng ít hơn 256 màu sắc khác nhau và các hoạt hình dùng ít hơn 256 màu cho mỗi khung hình. Gif thường dùng cho sơ đồ, hình vẽ, nút bấm và các hình màu. GIF là định dạng nén dữ liệu đặc biệt hữu ích cho việc truyền hình ảnh qua đường truyền lưu lượng nhỏ. Đây là một giải pháp tốt cho hình ảnh trên mạng, cho các hoạt hình nhỏ và ngắn.

GIF sử dụng thuật toán nén LOSS LESS (Không mất dữ liệu), điều đó cho phép chúng tạo ra kích thước nhỏ mà không bị mất hoặc mờ bất kỳ chi tiết nào của ảnh dữ liệu.

GIF note
GIF header (7 byte)
Globel Palette
Header Image (10 byte)
Palette of Image (nếu có)
Data of Image 1
‘,’ ký tự liên kết
……..
‘;’ terminator

Hình 4. Cấu trúc ảnh Gif

- Chữ ký của ảnh.
- Bộ mô tả hiển thị.
- Bản đồ màu tổng thể.

2.2 Mô tả một đối tượng của ảnh.

- Dấu phân cách.
- Bộ mô tả ảnh.

- Bản đồ màu cục bộ.
- Dữ liệu ảnh. Phần mô tả này lặp lại n lần nếu ảnh chứa n đối tượng.
- Phần đầu cuối ảnh GIF (terminator).

+ Chữ ký của ảnh GIF có giá trị là GIF87a. Nó gồm 6 ký tự, 3 ký tự đầu chỉ ra kiểu định dạng, 3 ký tự sau chỉ ra version của ảnh.

+ Bộ hình thị: chứa mô tả các thông số cho toàn bộ ảnh GIF:

Độ rộng hình raster theo pixel: 2 byte.

Độ cao hình raster theo pixel: 2 byte.

Các thông tin và bản đồ màu, hình hiển thị,…

Thông tin màu nền: 1 byte.

Phần chưa dùng: 1 byte.

+ Bản đồ màu tổng thể: mô tả bộ màu tối ưu đòi hỏi khi bit M=1.

Khi bộ màu tổng thể được thể hiện, nó sẽ xác định ngay bộ mô tả hiển thị ở trên và bằng 2^m, với m là lượng bit trên một pixel, 3 byte (biểu diễn cường độ màu của 3 màu cơ bản Red-Green-Blue). Cấu trúc của khối này như sau:

Bit	Thứ tự byte	Mô tả
Màu Red	1	**Giá trị màu đỏ theo index 0**
Màu Green	2	Giá trị màu xanh lục theo index 0
Màu Blue	3	Giá trị màu xanh lơ theo index 0
Màu Red	4	Giá trị màu đỏ theo index 1
Màu Green	5	Giá trị màu xanh lục theo index 1
Màu Blue	6	Giá trị màu xanh lơ theo index 0

Hình 5. Cấu trúc của khối bản đồ màu tổng thể

+ Bộ mô tả ảnh: định nghĩa vị trí thực tế và phần mở rộng của ảnh trong phạm vi không gian ảnh đã có trong phần mô tả hiển thị. Nếu ảnh biểu diễn theo ánh xạ màu cục bộ thì cờ định nghĩa phải được thiết lập. Mỗi bộ mô tả ảnh được chỉ ra bởi ký tự kết nối ảnh. Ký tự này chỉ được dùng khi định dạng GIF có từ 2 ảnh trở lên. Ký tự này có các giá trị 0x2c (ký tự dấu phẩy). Khi ký tự này được đọc qua, bộ mô tả ảnh sẽ được kích hoạt. Bộ mô tả ảnh gồm 10 byte và có cấu trúc như sau:

Các bit	Thứ tự byte	Mô tả
0010110	1	Ký tự liên kết ảnh (‘)
Căn trái ảnh	2,3	Pixel bắt đầu ảnh tính từ trái hình hiển thị
Căn đỉnh trên	4,5	Pixel cuối ảnh bắt đầu tính từ đỉnh trên hình hiển thị
Độ rộng ảnh	6,7	Độ rộng ảnh tính theo pixel
Độ cao ảnh	8,9	Chiều cao ảnh tính theo pixel
MI000pixel	10	Khi bit M=0 sử dụng bảng màu tổng thể. M=1 sử dụng bản đồ màu cục bộ. I = 0: định dạng ảnh theo thứ tự liên tục. I = 1: định dạng ảnh theo thứ tự xen kẽ pixel + 1: số bit/pixel của ảnh này.

Hình 6. Cấu trúc bộ mô tả ảnh

+ Bản đồ màu cục bộ: chỉ được chọn khi bit M của byte thứ 10 là 1. Khi bản đồ màu được chọn, bản đồ màu sẽ chiếu theo bộ mô tả ảnh mà lấy vào cho đúng. Tại phần cuối ảnh, bản đồ màu sẽ lấy lại phần xác lập sau bộ mô tả hiển thị. Các tham số này không những chỉ cho biết kích thước ảnh theo pixel mà còn chỉ ra số thực thể bản đồ màu của nó.

+ Dữ liệu ảnh: chuỗi các giá trị có thứ tự của các pixel màu tạo nên ảnh. Các pixel được xếp liên tục trên một dòng ảnh, từ trái qua phải. Các dòng ảnh được viết từ trên xuống dưới.

+ Phần kết thúc ảnh: cung cấp tính đồng bộ cho đầu cuối của ảnh GIF. Cuối của ảnh sẽ xác định bởi kí tự “;” (0x3b).

CHƯƠNG III. KỸ THUẬT GIẤU TIN THUẬN NGHỊCH TRONG ẢNH GIF

3.1 Khái niệm thuận nghịch

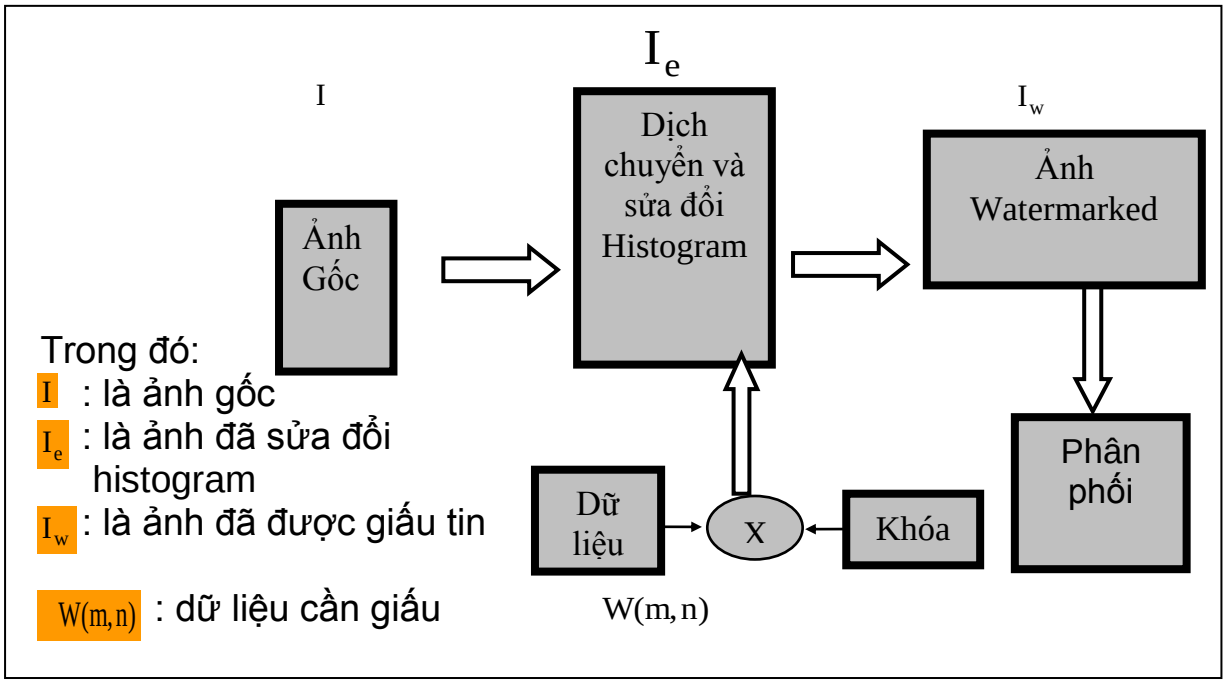
Giấu tin thuận nghịch là kỹ thuật giấu thông điệp sau khi khôi phục thông điệp ta có thể khôi phục lại xấp xỉ ảnh gốc ban đầu.

3.2 Kỹ thuật giấu thuận nghịch dựa trên DIH

Kỹ thuật giấu thuận nghịch dựa trên Difference Image Histogram (DIH) được đề xuất bởi Sang – Kwang Lee, Young – Ho Suh, và Yo – Sung Ho năm 2004[1].

Ý tưởng: Kỹ thuật này nhúng thông điệp cần giấu dựa vào histogram của ảnh sau khi đã được sửa đổi. Chuỗi thông điệp giấu được giấu vào các pixel mà Difference Image Histogram có giá trị 1 hoặc -1 trong ảnh đã sửa đổi. Số lượng Difference Image Histogram có giá trị 1 hoặc -1 thể hiện khả năng giấu lượng bit thông điệp vào ảnh gốc.

3.2.1 Quá trình giấu thông tin



Hình 7. Lược đồ quá trình giấu tin DIH

Các bước thực hiện:

Bước 1: Tính giá trị sai khác của ảnh D

+ Với mỗi hình ảnh *I* kích thước M × N pixel, ta tính được sự sai khác của ảnh *D(i, j)* với kích thước M × N / 2 như sau:

$$D(i, j) = I(i, 2j + 1) - I(i, 2j),\ 0 \le i \le M - 1,\ 0 \le j \le N/2 - 1 \qquad \textbf{(1)}$$

Trong đó I(i, 2j + 1) và I(i, 2j) là các trường lẻ và chẵn tương ứng *(odd line field and the even line field)*

Bước 2: Dịch chuyển và thay đổi Histogram

+ Trước khi nhúng thông điệp, ta làm rỗng các vùng -2 và 2 bằng việc thay đổi một vài giá trị điểm ảnh trong ảnh khác. Nếu các giá trị trong ảnh khác lớn hơn hoặc bằng 2, ta cộng thêm 1 vào những điểm hàng lẻ. Nếu các giá trị trong ảnh khác nhỏ hơn hoặc bằng -2, ta trừ 1 trong những điểm hàng lẻ

$$\tilde{D}(i, j) = \tilde{I}(i, 2j + 1) - I(i, 2j) \qquad (2)$$

$$\tilde{I}(i, 2j+1) = \begin{cases} I(i,2j+1) + 1 & \text{if } D(i,j) \ge 2 \\ I(i,2j+1) - 1 & \text{if } D(i,j) \le -2 \\ I(i,2j) \end{cases}$$

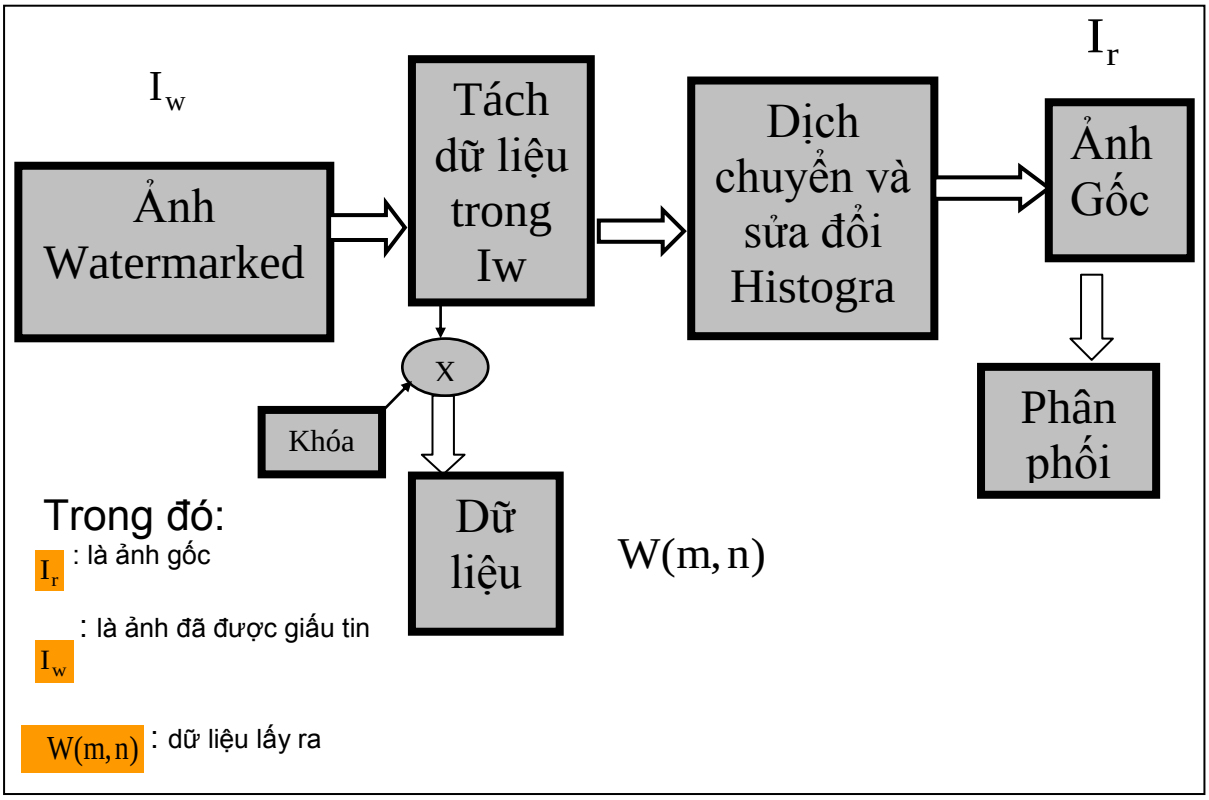
Bước 3: Thực hiện giấu thông điệp

+ W(m, n) là thông điệp cần giấu. Sau khi ta gặp một điểm ảnh $\tilde{D}(i,j)$ có giá trị -1 hoặc 1,ta kiểm tra watermark để nhúng vào. Nếu $\tilde{D}(i,j) = 1$ và $W(m,n) = 1$ thì $I_w(i,2j+1) = I_e(i,2j+1) + 1$. Nếu $\tilde{D}(i,j) = -1$ và $W(m,n) = 1$ thì $I_w(i,2j+1) = I_e(i,2j+1) - 1$. Còn các bit được nhúng vào là 0, ta bỏ qua các điểm ảnh của ảnh khác cho đến khi ta gặp một điểm ảnh có giá trị -1 hoặc 1. Trong trường hợp này, không có sự thay đổi trong biểu đồ. Do đó $I_w(i, 2j + 1)$ và $I_w(i, 2j)$ được tạo lên:

$$I_w(i,2j+1) = \begin{cases} \tilde{I}(i,2j+1)+1 \text{ if } \tilde{D}(i,j)=1 \text{ and W(m,n) = 1} \\ \tilde{I}(i,2j+1)-1 \text{ if } \tilde{D}(i,j)=-1 \text{ and W(m,n) = 1} \\ \tilde{I}(i,2j+1) \qquad \text{otherwise} \end{cases} \tag{3}$$

$$I_w(i,2j) = I\ (i,2j)$$

3.2.2 Quá trình lấy thông tin



Hình 8. Lược đồ quá trình lấy tin DIH

Sau khi có được ảnh watermarked $I_e(i, j)$, Áp dụng công thức (1) ta được $D_e(i, j)$. Nếu gặp các điểm ảnh có giá trị -1 hoặc 1, thì bit 0 được lấy. Nếu gặp các điểm ảnh có giá trị -2 hoặc 2 thì bit 1 được lấy. Bằng cách này $W_e(m,n)$ có thể được lấy ra:

$$W_e(m,n) = \begin{cases} 0 \text{ if } D_e(i,j) = -1 \text{ or } 1 \\ 1 \text{ if } D_e(i,j) = -2 \text{ or } 2 \end{cases} \tag{4}$$

Để khôi phục được ảnh gốc, ta dịch chuyển một số pixel trong I_e như sau: nếu D_e (I,j) có giá trị ≤ *-2* thì tăng thêm *1* vào I_e *(I,2j+1),* nếu D_e có giá trị ≥ *2* thì giảm *1* tại I_e *(I,2j+1).* Cuối cùng ta sẽ thu được ảnh gốc ban đầu:

$$I_r(i,2j+1) = \begin{cases} I_e(i,2j+1) - 1 \text{ if } \mathrm{D}_e(i,j) \geq 2 \\ I_e(i,2j+1) + 1 \text{ if } \mathrm{D}_e(i,j) \leq -2 \\ I_e(i,2j+1) \text{ otherwise} \end{cases} \quad (5)$$

$$I_r(i,2j) = \mathrm{I}_e(i,2j)$$

Phương pháp giấu DIH có thể không trả về được ảnh gốc hoàn toàn đúng như ban đầu bởi việc mất mát thông tin xảy ra trong quá trình cộng trừ tại biên của vòng xám (mức xám là từ 0 ÷ 255). Để khắc phục vấn đề này, họ đưa ra modulo số học cho các phép cộng và trừ thủy vân. Đối với trường lẻ I(i,2j+1), phép cộng modulo c như sau:

$$I(i,2j+1) +_c 1 = ((i,2j+1) + 1) \bmod c \quad (6)$$

Với c là độ dài của vòng giá trị màu. Đối với phép trừ modul *c* được định nghĩa như sau:

$$I(i,2j+1) -_c 1 = ((i,2j+1) + 1) \bmod c \quad (7)$$

Những vấn đề thuận nghịch được phát sinh từ sự thừa, thiếu hụt pixel. Vì vậy, ta sử dụng $+_c$ và $-_c$ thay vì + và – chỉ khi bỏ bớt do thừa hay thiếu hụt xảy ra. Nói cách khác, ta chỉ để xem xét $255 +_c 1$ và $0 -_c 1$.

Khi nhận được, ta cần phân biệt giữa các trường hợp, ví dụ: $\boldsymbol{I_e(i, 2j+1) = 255}$ có được như: $\boldsymbol{I(i, 2j+1) + 1}$ **và** $\boldsymbol{I(i, 2j+1) -_{256} 1}$. Nếu có một sự khác biệt đáng kể giữa $\boldsymbol{I_e(i, 2j+1)}$ **và** $\boldsymbol{I_e(i, 2j)}$, ta ước lượng $\boldsymbol{(i, 2j+1)}$ vận dụng modulo số học.

$$\begin{cases} I(i,2j+1) + 1 & \text{if } \left| I_e(i,2j+1) - I_e(i,2j) \right| \leq \tau \\ I(i,2j+1) -_{256} 1 & \text{otherwise} \end{cases} \quad (8)$$

Trong đó τ là giá trị ngưỡng. Tương tự $\boldsymbol{I_e(i, 2j + 1) = 0}$ được ước lượng bằng cách:

$$I(i,2j+1) = \begin{cases} I(i,2j+1)-1 & \text{if } \left|I_e(i,2j+1)-I_e(i,2j)\right| \le \tau \\ I(i,2j+1)+_{256}1 & \text{otherwise} \end{cases} \quad (9)$$

CHƯƠNG IV. KỸ THUẬT PHÁT HIỆN THÔNG TIN ẨN GIẤU TRONG ẢNH GIF

4.1 Tổng quan về kỹ thuật phát hiện tin ẩn giấu trong ảnh (Steganalysis)

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

Phân tích tin ẩn giấu thường dựa vào các yếu tố sau:

- 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 điệp 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.

Các phương pháp phân tích có thể phân thành 3 nhóm:

- Phân tích trực quan: Thường dựa vào quan sát hoặc dùng biểu đồ 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 các ảnh dữ liệu lớn.

4.2 Kỹ thuật phát hiện ảnh có giấu tin dựa trên DIH

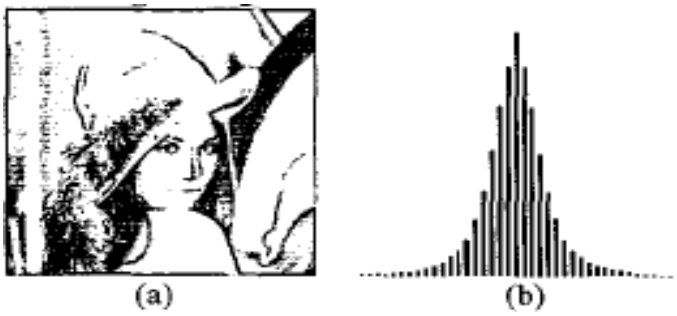
Theo một tài liệu của hai tác giả Tao Zhang và Xijian Ping[2] của học viện khoa học thông tin, Đại học công nghệ thông tin Zhengzhou, P.R Trung Quốc nói:

Hầu hết các hệ thống giấu tin đều không hoàn toàn an toàn và có thể nhận biết được qua nhận dạng vân tay hay một số hình thức khác. Xem xét các thuộc tính của giấu tin trên miền LSB, chúng tôi chọn ***difference image histogram*** như là một công cụ phân tích thống kê. Giá trị cường độ của ảnh I tại vị trí (i, j) – I(i, j), và sự khác biệt của ảnh được định nghĩa là:

$$D(ij)=I(ij)-I(i,j+l). \qquad (10)$$

Difference Image Histogram được định nghĩa như là histogram của ảnh khác biệt D. Nhìn chung, nó được tin rằng ảnh khác biệt được chấp nhận như một phân bổ Gaussian tổng quát có hàm mật độ xác suất có thể được tính như sau:

$$p_{\nu,\beta}(x)=\frac{\nu}{2\beta\Gamma(\frac{1}{\nu})}\exp\left\{-\left(\frac{|x|}{\beta}\right)^{\nu}\right\} \qquad (11)$$



Hình 9. (a). Ảnh chuẩn "Lena"; Difference Image Histogram của ảnh "Lena" (b).

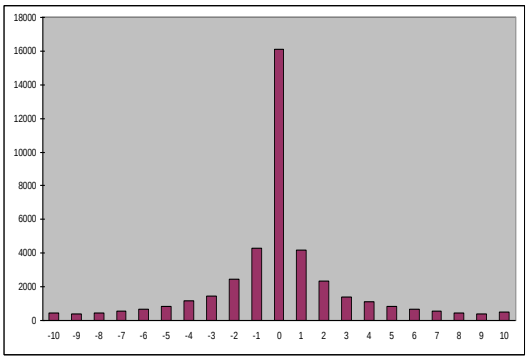
Từ đó nhóm tác giả đề xuất kỹ thuật phát hiện dựa trên DIH, ước lượng histogram của ảnh cover và ảnh stego và thống kê sự khác biệt đó. Và đề xuất này được đánh giá bằng thực nghiệm (5.4).

Đối với ảnh không giấu tin, tổng số histogram :

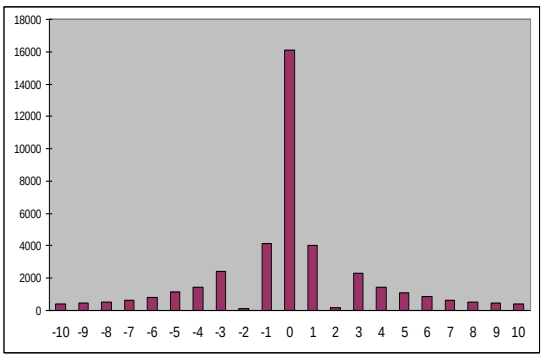
$$h_1 + h_{-1} > h_2 + h_{-2} > h_3 + h_{-3} > \ldots\ldots > h_{10} + h_{-10} > \ldots\ldots \qquad (12)$$

Đối với ảnh có gấu tin, ta có:

$$h_2 + h_{-2} \leq h_3 + h_{-3} \qquad (13)$$



Ảnh gốc



Ảnh có giấu tin

Chúng tôi nhận thấy rằng, sau khi nhúng thông điệp bằng thuật toán DIH sẽ làm thay đổi tổng số histogram $h_{\pm 2}$ của ảnh.

Ý tưởng như sau:

Xét tỉ lệ của $(h_2 + h_{-2})$ với $(h_3 + h_{-3})$. Xét tỷ lệ này với T. Nếu tỉ lệ này nhỏ hơn hoặc bằng thì ảnh này có giấu tin ngược lại ảnh không giấu tin.

Thuật toán:

Input: Cho một ảnh GIF Q

Output: Kiểm tra xem ảnh Q có giấu tin hay không giấu tin

Các bước thực hiện như sau:

Bước 1. Tính độ sai khác (DIH) giữa các pixel của ảnh giống như quy trình giấu tin. Sau đó tính tần số của các giá trị sai khác này ký hiệu là h_i.

Bước 2. So sánh tỷ lệ giữa $h_{\pm 2}$ và $h_{\pm 3}$:

Nếu $(h_2 + h_{-2})/(h_3 + h_{-3}) <= T_1$ và $(h_1 + h_{-1})/(h_2 + h_{-2}) <= T_2$ thì thực hiện Bước 3. Ngược lại thực hiện Bước 4.

Bước 3. Ảnh có giấu tin. Ước lượng xấp xỉ độ dài thông điệp giấu như sau:

- Gọi L là độ dài xấp xỉ thông điệp ẩn giấu trên ảnh được tính theo công thức:

$$L=2*(h_2 + h_{-2}) \quad (14)$$

- Gọi [p, q] là kích thước ảnh. Ước lượng tỷ lệ phần trăm của ảnh có chứa thông điệp ẩn giấu như sau:

$$E = [L/(p*q)] * 100 \quad (15)$$

Bước 4. Ảnh không giấu tin

CHƯƠNG V. KẾT QUẢ THỰC NGHIỆM

5.1. Môi trường thử nghiệm

Cài đặt chương trình trên môi trường Java, sử dụng bộ soạn thảo JCreator_Pro_v4.5 và thông dịch JDK-6u10.

Cầu hình máy tính tối thiểu để chạy chương trình: Hệ diều hành Windown XP hoặc các hệ điều hành tương tự, Chip PIII 500 trở lên, Ram từ 128, ổ cứng còn trống 400 Mb.

Chương trình gồm các chức năng sau:

+ **Giấu tin**: Quá trình thực hiện như sau:

Chọn file ảnh GIF → Giấu tin → Chọn vị trí lưu file ảnh output.gif → Chọn file text cần giấu.

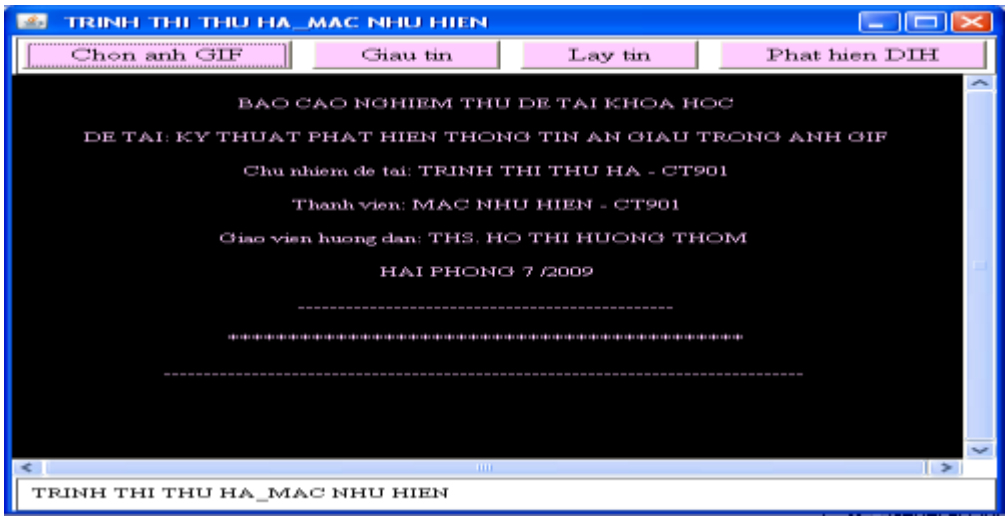
+ **Lấy tin:** Quá trình thực hiên như sau:

Chọn file ảnh GIF → Lấy tin → Chọn vị trí lưu file text output.txt → Chọn vị trí lưu file ảnh gốc anhgoc.gif

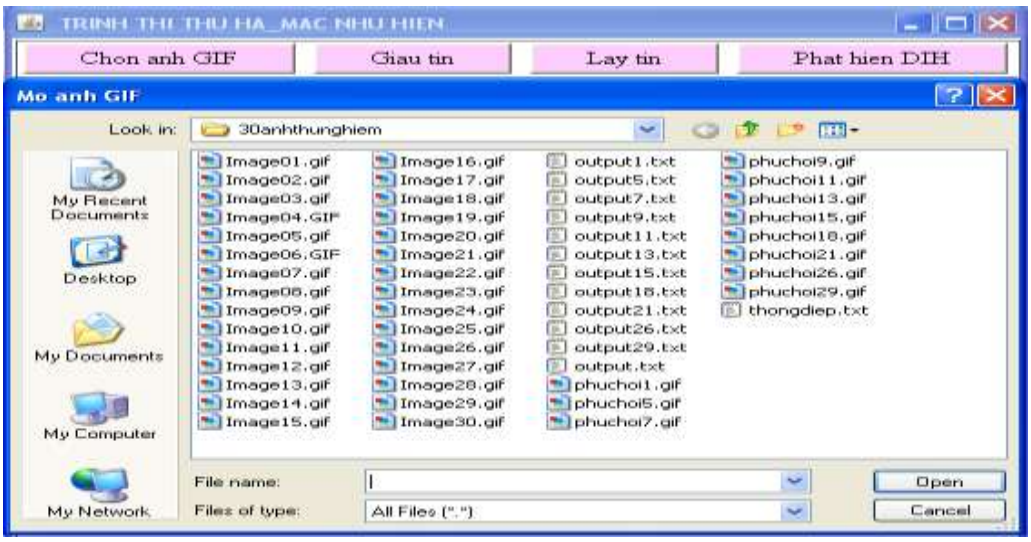
+ **Phát hiện DIH:** Quá trình thực hiên như sau:

Chọn file ảnh GIF → Phát hiện DIH → Đưa ra thông báo trên thanh công cụ về ảnh kiểm tra

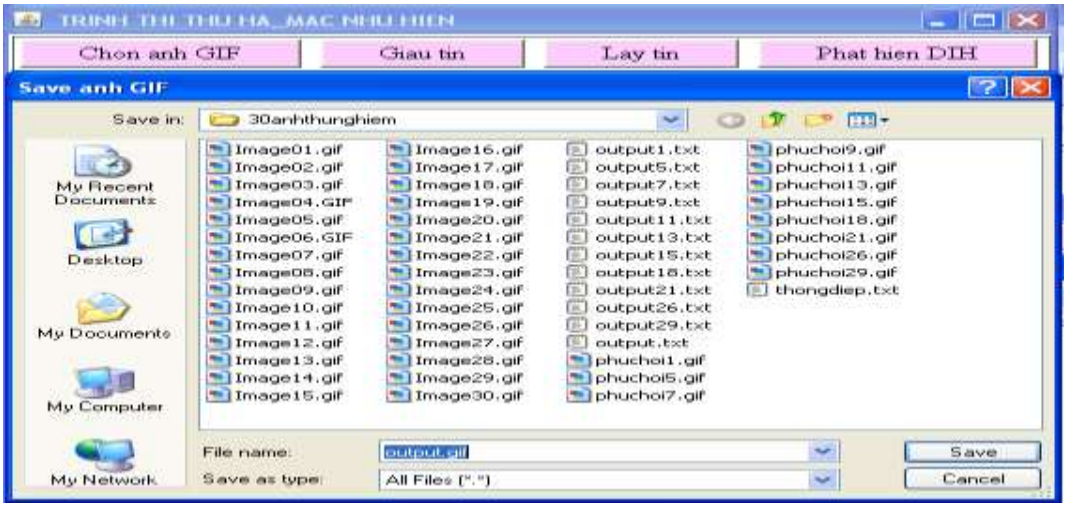
Giao diện chương trình:



Hình 10. Giao diện chính của chương trình



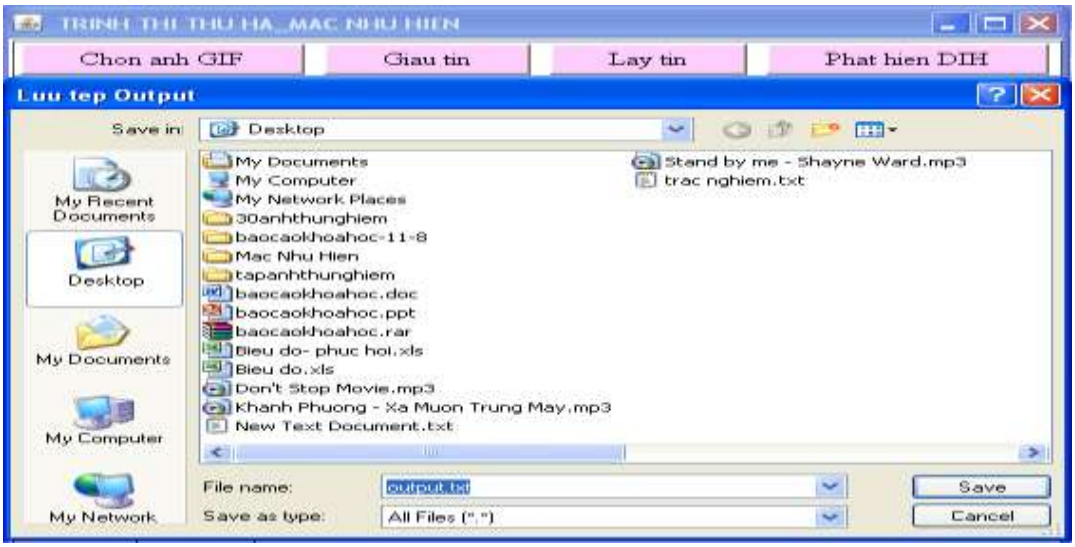
Hình 11. Chọn file ảnh GIF cần giấu



Hình 12. Chọn vị trí lưu file ảnh mới output.gif

Hình 13. Chọn file text cần giấu

Hình 14. Chọn vị trí lưu ảnh phục hồi recovered.gif



Hình 15. Chọn vị trí lưu file text được lấy ra output.txt

Hình 16. Kiểm tra ảnh

5.2. Thử nghiệm thuật toán giấu thông điệp

5.2.1 Cơ sở dữ liệu thử nghiệm

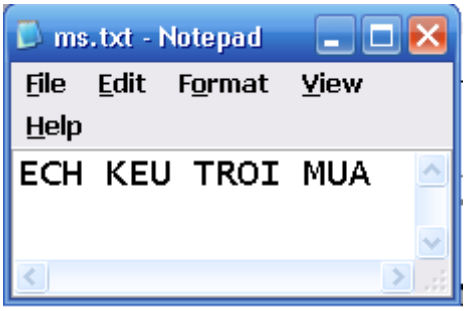
Có một tập cơ sở dữ liệu ảnh gồm 6 ảnh GIF chuẩn được download từ [5] và [6] có kích cỡ 512x512 pixel.

Hình 17. Các hình ảnh GIF thử nghiệm

5.2.2 Kết quả thử nghiệm và đánh giá thuật toán bằng (PSNR)

Để đánh giá hiệu quả hoạt động của phương pháp đề xuất, chúng ta thực hiện trên nhiều máy tính mô phỏng trên một vài ảnh GIF kích thước 512 × 512 pixels.

Chuỗi ký tự cần giấu:



Hình 18. Chuỗi kỹ tự cần giấu

Kết quả thực nghiệm:

Ảnh gốc	Ảnh Wartermared
 Airplane.gif	 Airplane.gif
 Sailboat.gif	 DIH_Sailboat.gif
 Baboon.gif	 DIH_Baboon.gif
 Lena.gif	 DIH_Lena.gif
 Tiffany.gif	 DIH_Tiffany.gif
 Peppers.gif	 DIH_Peppers.gif

Hình 19. Ảnh trước và sau khi giấu tin

Đánh giá thuật toán bằng PSNR

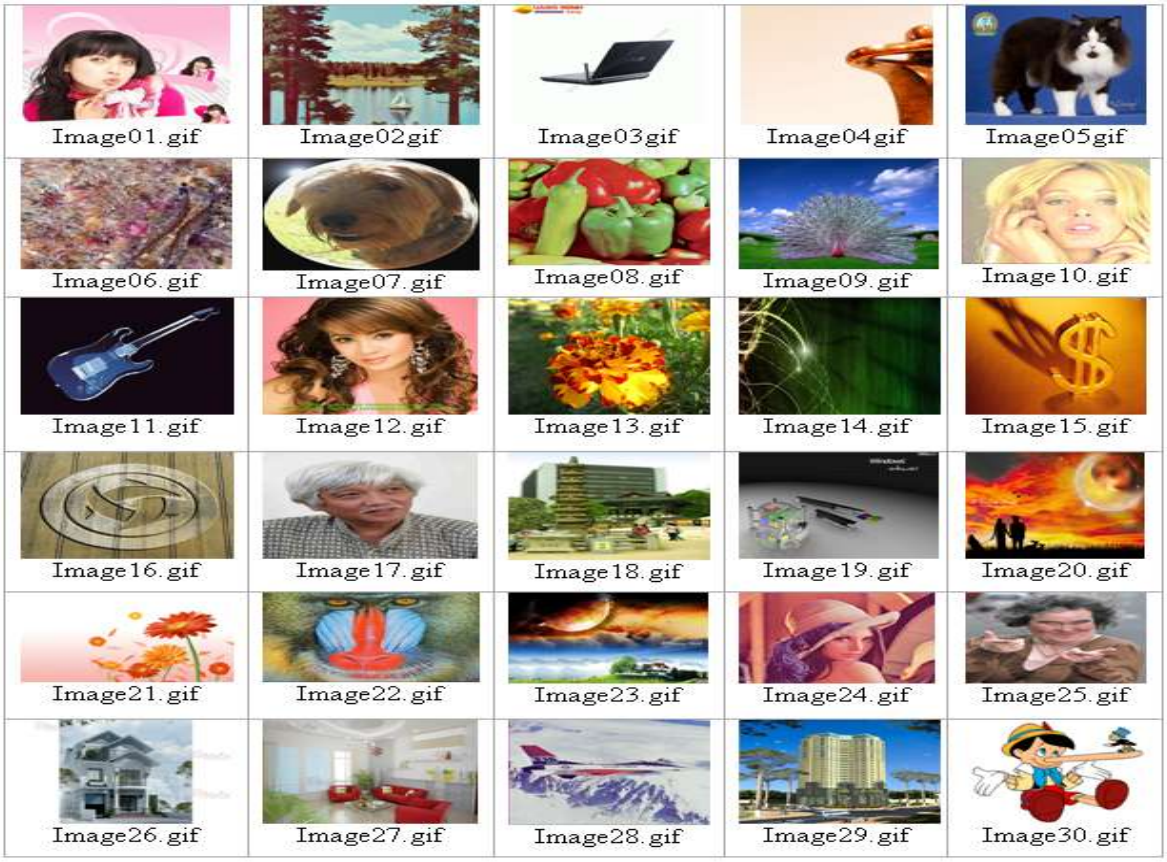
Bảng 1 cho thấy rằng các giá trị PSNR của tất cả các hình ảnh watermarked đang ở trên 51,14 dB. Khả năng dao động từ 8 kbits đến 30 kbits của 512 × 512 × 8 bits.

Ảnh (512x512x8)	PSNR (dB)	Khả năng giấu (bit)	Vượt ngưỡng (pixels)
Airplane	58.78	13,551	0
Baboon	51.49	14,111	2
Lena	55.63	16.379	0
Peppers	55.74	23,725	2
Sailboat	55.55	17,719	14
Tiffany	55.20	20,497	1

Bảng 1. Bảng tóm tắt kết quả thực nghiệm.

5.3 Cài đặt thuật toán phát hiện

Cho một tập ảnh thử nghiệm gồm 30 ảnh với kích thước bất kỳ (weight, height nhỏ hơn 2000). Trong đó có 10 ảnh đã có giấu tin bằng kỹ thuật giấu thuận nghịch dựa trên DIH (gồm: Image01.gif, image05.gif, image07.gif, image09.gif, image11.gif, image13.gif, image15.gif, image18.gif, image26.gif, image29.gif). Dưới đây là tập ảnh thử nghiệm chương trình phát hiện DIH.



Bảng 2. Tập ảnh thử nghiệm

Tên ảnh	Kiểm tra ảnh bằng phương pháp Phát hiện DIH
Imge 01.gif	**C**
Imge 02.gif	**K**
Imge 03.gif	**K**
Imge 04.gif	**K**
Imge 05.gif	**C**
Imge 06.gif	**K**
Imge 07.gif	**C**
Imge 08.gif	**K**
Imge 09.gif	**C**
Imge 10.gif	**K**
Imge 11.gif	**C**
Imge 12.gif	**K**
Imge 13.gif	**C**
Imge 14.gif	**K**
Imge 15.gif	**C**
Imge 16.gif	**K**
Imge 17.gif	**K**
Imge 18.gif	**C**
Imge 19.gif	**K**
Imge 20.gif	**K**
Imge 21.gif	**C**
Imge 22.gif	**K**
Imge 23.gif	**K**
Imge 24.gif	**K**
Imge 25.gif	**K**
Imge 26.gif	**C**
Imge 27.gif	**K**
Imge 28.gif	**K**
Imge 29.gif	**C**
Imge 30.gif	**K**

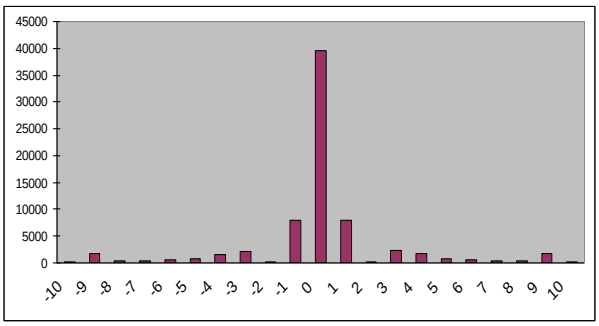
Bảng 3. Kết quả kiểm tra ảnh

Chú thích: trong bảng 3 ký hiệu K là "ảnh không giấu tin", ký hiệu C là "ảnh có giấu tin".

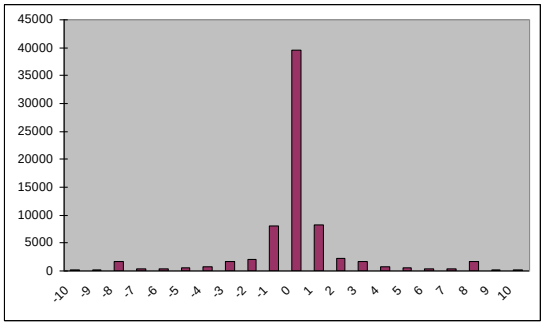
5.4 Đánh giá các kết quả thực nghiệm

Từ bảng kết quả kiểm tra ảnh trên cho thấy, chương trình phát hiện DIH phát hiện có 11 ảnh có giấu tin mật, lệch với số ảnh giấu tin trên thực tế ban đầu là 1 ảnh.

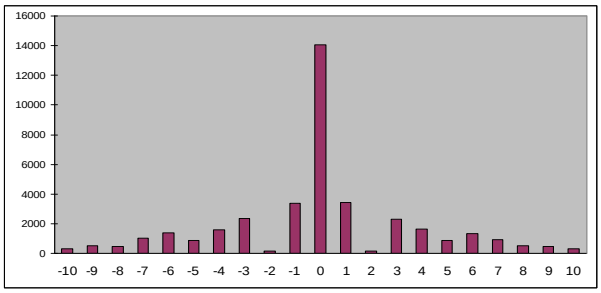
Để kiểm tra tính tin cậy của tư tưởng thuật toán phát hiện, nhóm tác giả lấy 11 ảnh được chương trình phát hiện là có thông tin ẩn giấu để thống kê histogram của ảnh. Sau đó sử dụng chương trình lấy thông tin và khôi phục ảnh gốc để lấy ra thông tin mật và thống kê lại histogram của ảnh đã khôi phục.



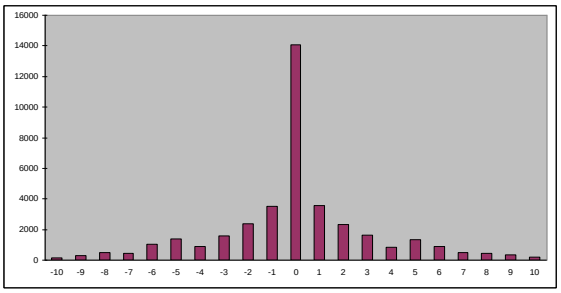
Histogram của Image01.gif



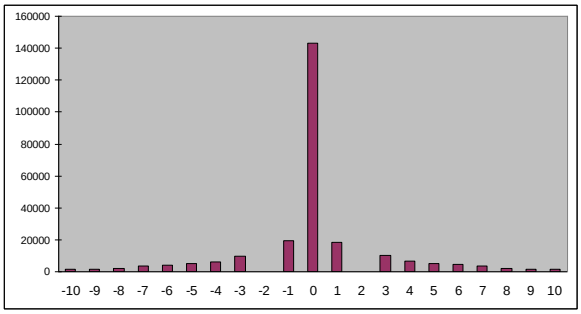
Histogram của phuchoi01.gif



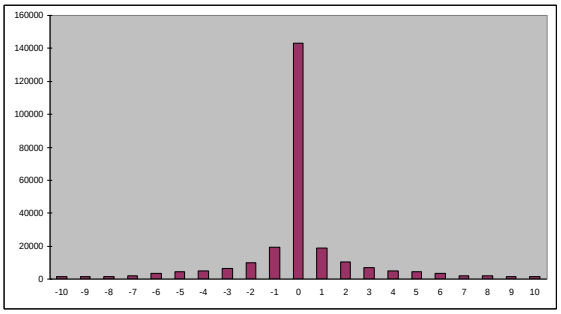
Histogram của Image05.gif



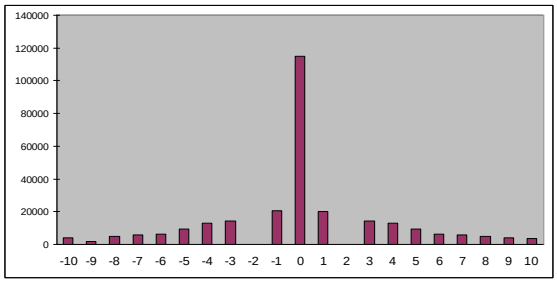
Histogram của phuchoi05.gif



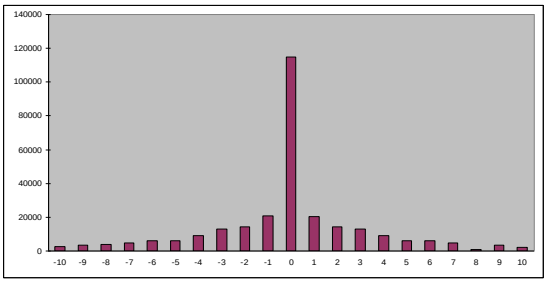
Histogram của Image07.gif



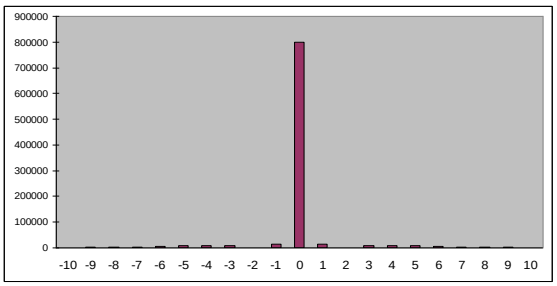
Histogram của phuchoi07.gif



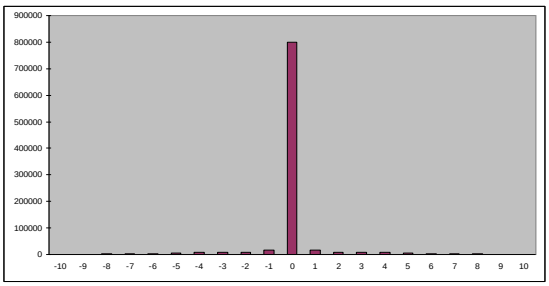
Histogram của Image09.gif



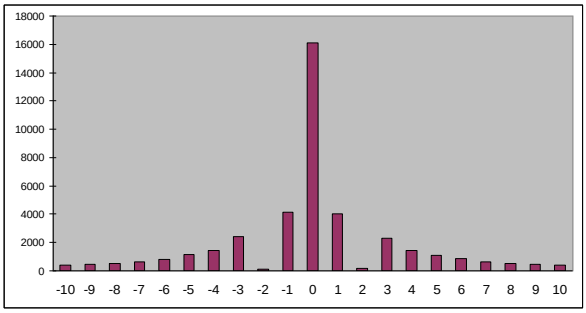
Histogram của phuchoi09.gif



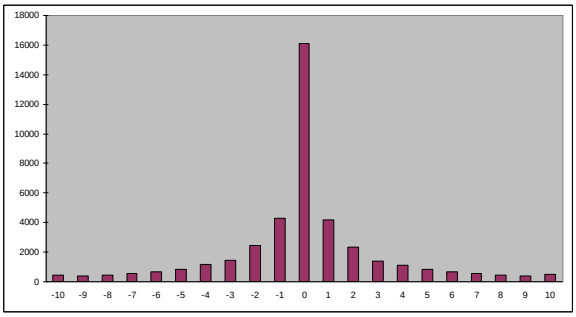
Histogram của Image11.gif



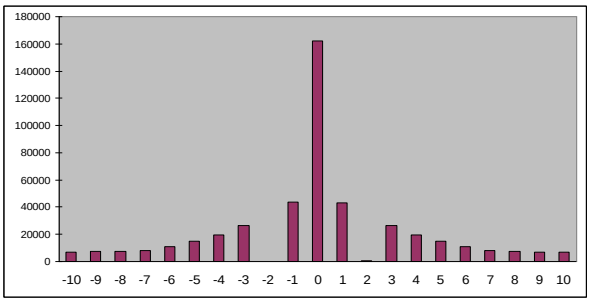
Histogram của phuchoi11.gif



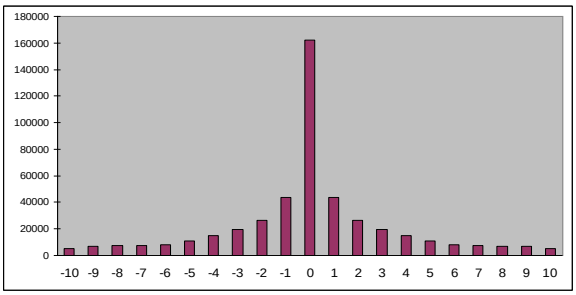
Histogram của Image13.gif



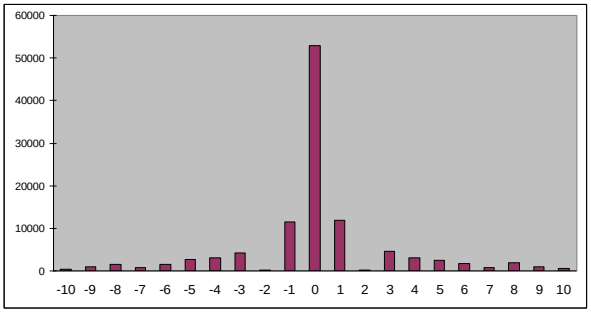
Histogram của phuchoi13.gif



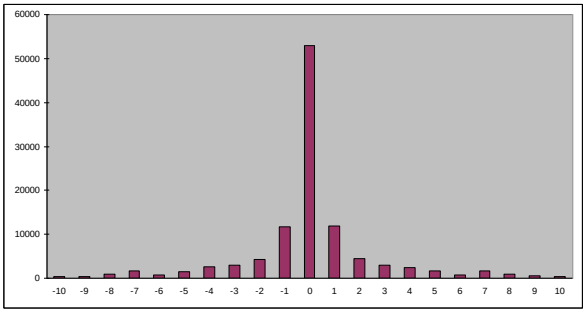
Histogram của Image15.gif



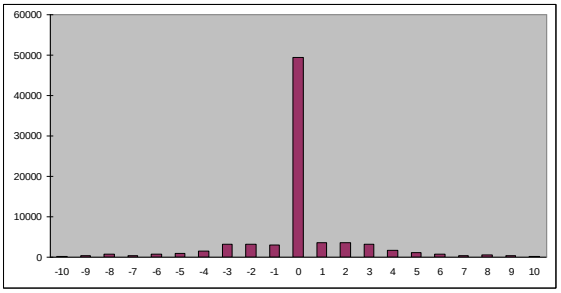
Histogram của phuchoi15.gif



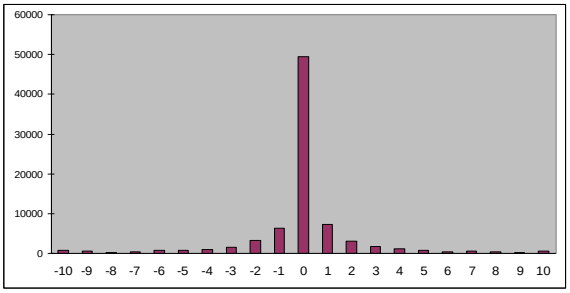
Histogram của Image18.gif



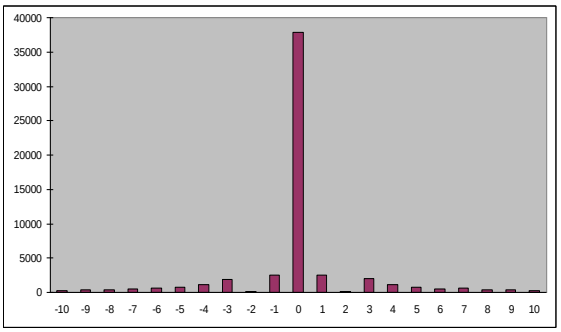
Histogram của phuchoi18.gif



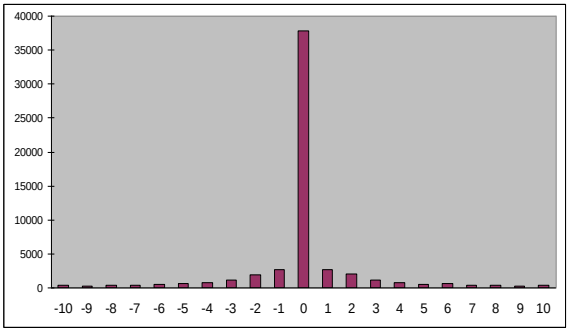
Histogram của Image21.gif



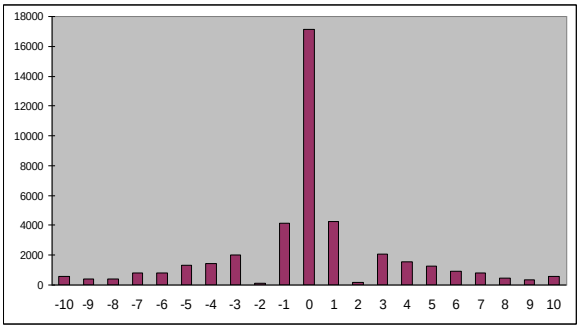
Histogram của phuchoi21.gif



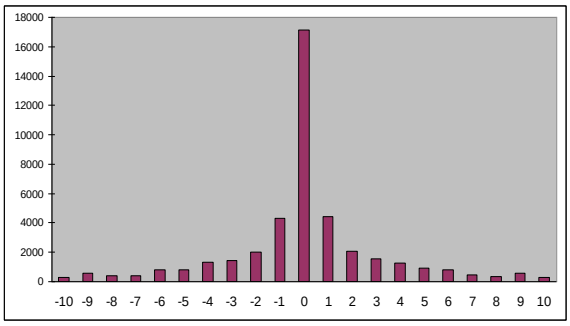
Histogram của Image26.gif



Histogram của phuchoi26.gif



Histogram của Image29.gif



Histogram của phuchoi29.gif

Và dưới đây là histogram của các ảnh mà chương trình phát hiện DIH cho kết quả là không giấu tin mật.

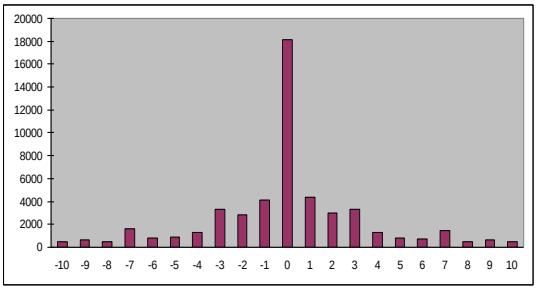


Image02.gif

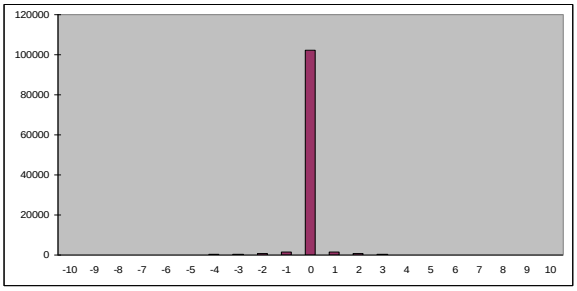


Image03.gif

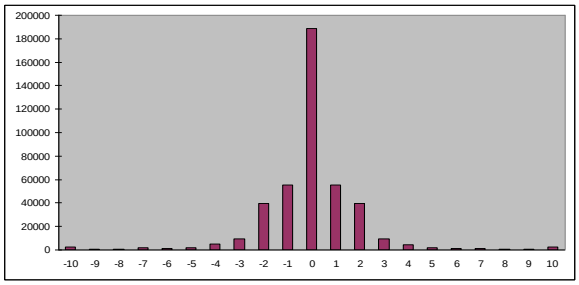


Image04.gif

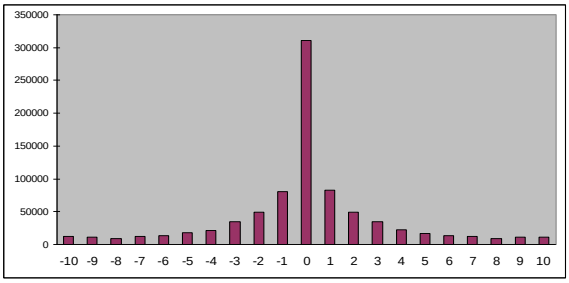


Image06.gif

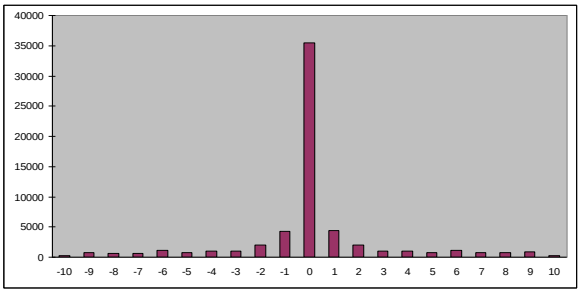


Image08.gif

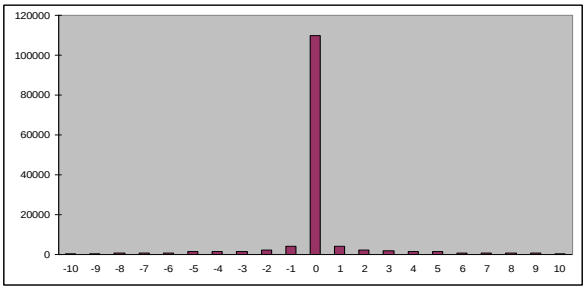


Image10.gif

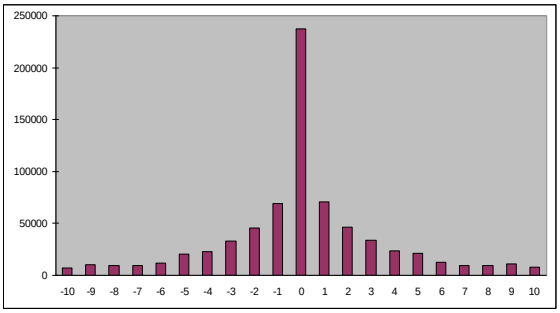


Image14.gif

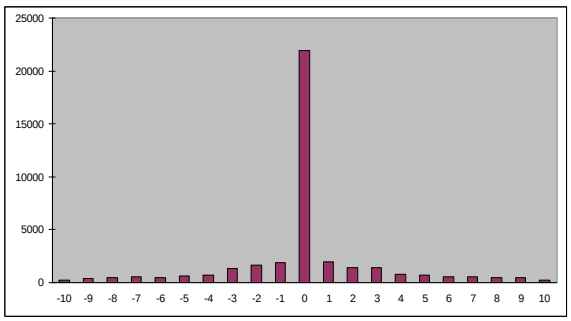


Image17.gif

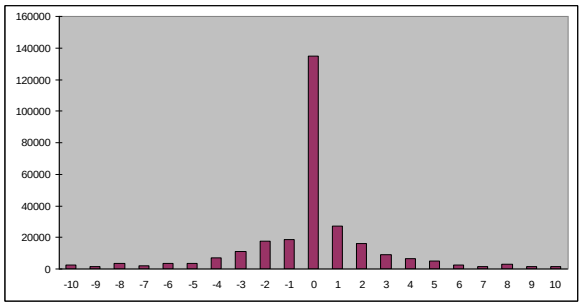


Image19.gif

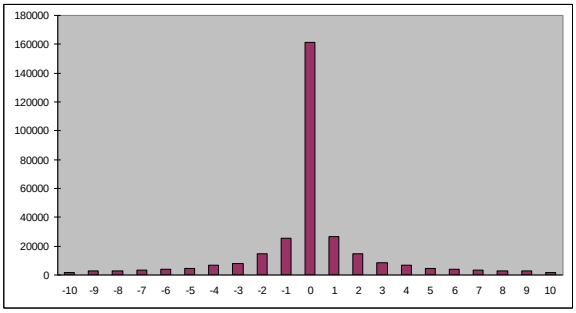


Image20.gif

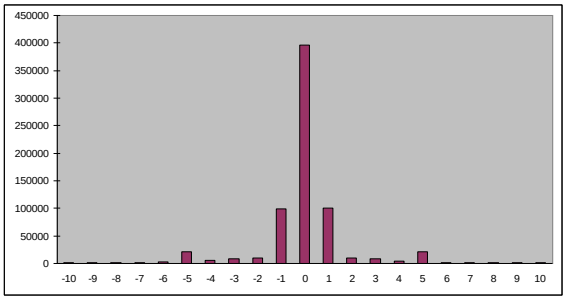


Image22.gif

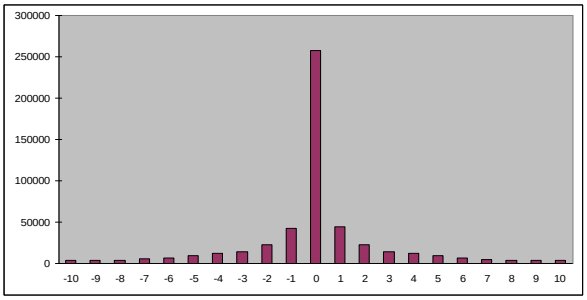


Image23.gif

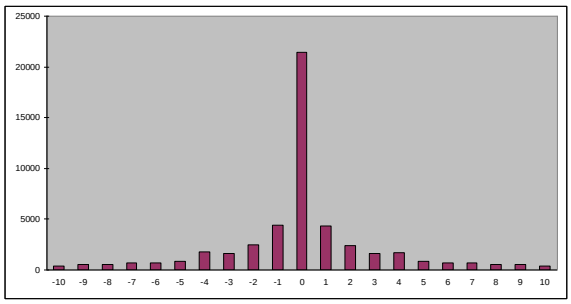


Image24.gif

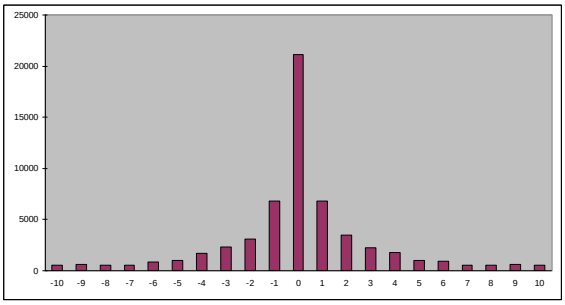


Image25.gif

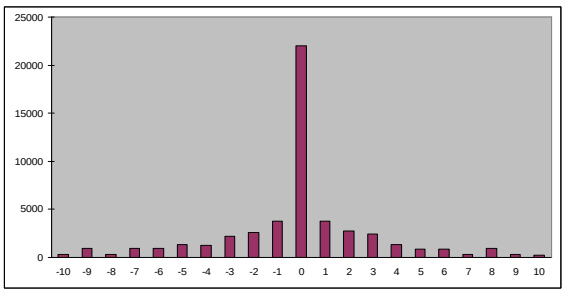


Image27.gif

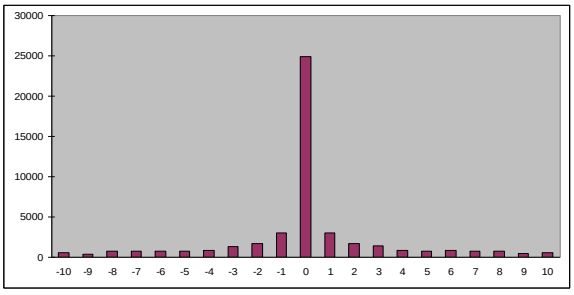


Image28.gif

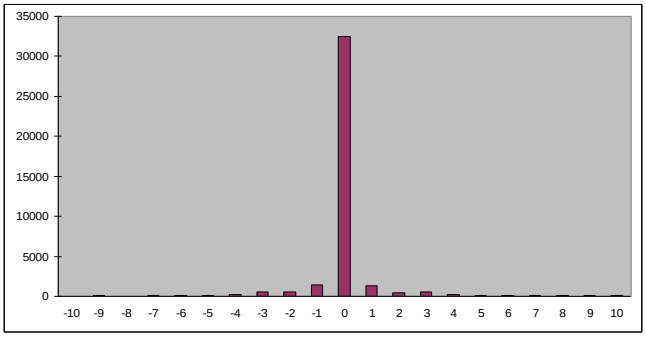


Image30.gif

Nhóm tác giả thực hiện đánh giá kết quả trên theo hai hệ số: Precision và Recall. Trong đó, hệ số Precision (ký hiệu là P) phản ánh độ chính xác của việc phát hiện ảnh có giấu tin, được tính theo công thức:

P = Số ảnh giấu tin chính xác/ Số ảnh tìm được là có giấu tin.

Và hệ số Recall (ký hiệu là R) được tính theo công thức:

R = Số ảnh tìm được đúng có giấu tin/ Số ảnh đúng có giấu tin ban đầu.

Như vậy từ kết quả thử nghiệm trên, ta có giá trị của các hệ số như sau:

P = 10/11 = 0.9090909090….

R = 10/10 = 1.

KẾT KUẬN

Phát hiện thông tin 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 được một ảnh số nào đó có mang tin mật hay không đòi hỏi rất nhiều yếu tố và kỹ thuật phức tạp.

Sau khoảng thời gian không nhiều (6 tháng) nhóm tác giả đã làm việc hăng say, tích cực, nghiêm túc, nhóm tác giả đã đạt được các kết quả sau:

- Nắm rõ được khái niệm tổng quan về kỹ thuật giấu tin trong ảnh nói chung, giấu tin trong ảnh GIF nói riêng
- Tìm hiểu và cài đặt được kỹ thuật giấu tin thuật nghịch dựa trên tần số ảnh sai khác (Difference Image Histogram-DIH) trong ảnh GIF.
- Đề xuất kỹ thuật phát hiện ảnh GIF có giấu tin dựa trên DIH.
- Xây dựng chương trình giấu và phát hiện ảnh GIF có giấu tin với giao diện thân thiện và dễ sử dụng.

Tuy kỹ thuật phát hiện đề xuất chỉ phát hiện được đối với ảnh giấu tin bằng kỹ thuật giấu tin thuận nghịch dựa trên DIH và mới chỉ được chứng minh qua thử nghiệm nhưng đã mở ra một hướng đi mới cho các nghiên cứu tiếp theo trong lĩnh vực steganalysis.

Hướng nghiên cứu tiếp theo sẽ nghiên cứu tiếp việc áp dụng kỹ thuật giấu và phát hiện trên với ảnh GIF động, bởi những kết quả thực nghiệm trên mới chỉ thực hiện trên ảnh GIF tĩnh và không cho kết quả tốt với ảnh GIF động.

TÀI LIỆU THAM KHẢO

[1]. Sang-Kwang Lee, Young-Ho Suh, and Yo-Sung Ho, *Lossless Data Hiding Based on Histogram Modification of Difference Images,* Advances in Multimedia Information Processing - PCM 2004, pp.340-347. November/ December, 2004.

[2]. Tao Zhang, Xijian Ping: *RELIABLE DETECTION OF LSB STEGANOGRAPHY BASED ON THE DIFFERENCE IMAGE HISTOGRAM*. ICASSP 2003. Vol. I, pp.545-548.

[3]. CBIR image database, University of Washington, available at:http://www.cs.washington.edu/research/imagedatabase/groundtruth/.

[4].USC-SIPI Image Database,

http://sipi.usc.edu/services/database/Database.html.