Um estudo sobre abordagens computacionais de esteganografia e esteganálise em imagens

Adriele Alvarenga Secundino

JUIZ DE FORA

ABRIL, 2013

Um estudo sobre abordagens computacionais de esteganografia e esteganálise em imagens

Adriele Alvarenga Secundino

Universidade Federal de Juiz de Fora

Instituto de Ciências Exatas

Departamento de Ciência da Computação

Bacharelado em Ciência da Computação

Orientador: Stênio Sã Rosário Furtado Soares

JUIZ DE FORA

ABRIL, 2013

Um estudo sobre abordagens computacionais de esteganografia e esteganálise em imagens

Adriele Alvarenga Secundino

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Stênio Sã Rosário Furtado Soares
DSc. em Ciência da Computação - DCC UFJF

Eduardo Pagani Julio
MSc. em Ciência da Computação - DCC UFJF

Marcelo Caniato Renhe
MSc. em Engenharia de Sistemas e Computação - DCC UFJF

JUIZ DE FORA

04 DE ABRIL, 2013

Aos meus familiares pelo o amor e carinho.

Aos meus amigos pelo o apoio.

Resumo

As técnicas de ocultar informações em arquivos digitais são definidas como esteganografia. Dentre as várias conhecidas, aquela que faz uso de imagens vem despertando bastante interesse da comunidade científica, dada a popularização dos diversos dispositivos de captura de imagens, bem como a facilidade com que este tipo de arquivo pode ser distribuído na Web. Tendo em vista a possibilidade de que mensagens ocultas possam ser enviadas com propósitos ilícitos, como para planejamento de atentados terroristas, o desenvolvimento da esteganálise (abordagens com propósito de identificar este tipo de manipulação de arquivos) ganha força tanto no cenário acadêmico como no que se refere às suas aplicações. Esta monografia apresenta o desenvolvimento das principais abordagens utilizadas na detecção de informações ocultas, conforme a evolução das abordagens de esteganografia.

Palavras-chave: Esteganografia, esteganálise, análise digital.

Abstract

The techniques to hide information in digital files are defined as steganography. Among the many known, one that makes use of images has attracted a lot of interest from the scientific community, given the popularity of various devices capture images as well as the ease with which this type of file can be distributed on the Web. In order to possibility that hidden messages can be sent with illicit purposes, such as for planning terrorist attacks, the development of steganalysis (purposeful approaches to identify this type of file manipulation) gains strength both in the academic setting and in relation to their applications . This monograph presents the development of the main approaches used in the detection of hidden information, as the evolution of approaches to steganography.

Keywords: Steganography, steganalysis, digital analysis.

Agradecimentos

Aos meus pais, por tudo que fizeram por mim desde que eu nasci. Aos meus amigos de longa data pelo encorajamento e apoio. A todos os amigos que fiz durante essa jornada que me acompanharam tanto nos momentos mais difíceis quanto nos momentos de diversão e não me deixaram desistir.

Ao professor Stênio Sã Rosário Furtado Soares pela orientação, amizade e principalmente, pela paciência, sem a qual este trabalho não se realizaria.

Aos professores do Departamento de Ciência da Computação por tudo que me ensinaram e aos funcionários do Instituto de Ciências Exatas, principalmente a Sheila Valentim por sua paciência extra comigo.

A Universidade Federal de Juiz Fora pela oportunidade.

A Deus por toda força que me proporcionou para realização deste trabalho.

"Se a gente não se esforçar pra mudar o mundo, depois é o mundo que vai mudar a gente!".

Malfada Quino

Sumário

Lista de Figuras

1 Introdução

Segredos sempre foram valiosos para a humanidade. Ao longo dos anos, diversas formas de esconder informações surgiram como um meio de restringir o acesso às mesmas. A esteganografia é a ciência que estuda dos meios para ocultar informações(Popa, 1998).

Geralmente, técnicas de esteganografia usam uma mensagem sem importância, que pode ser um texto, um áudio, uma imagem e escondem a mensagem que realmente importa em sua estrutura. Existem várias formas de ocultar uma mensagem e também são diversos os objetivos, como proteção de direitos autorais (marca d'água), fins militares, troca de informações entre empresas etc.

Atualmente a esteganografia é usada no meio digital e qualquer mídia pode ser usada para esconder uma mensagem. A imagem digital é uma das mídias usadas na esteganografia.

Assim como existem pessoas que querem esconder informações existem pessoas tentando desvendá-las. Um exemplo seria no caso em que a esteganografia é usada para propósitos ilícitos. Assim, neste contexto, é de suma importância o desenvolvimento de técnicas para detectar as informações escondidas. A detecção de mensagens ocultas é denominada esteganálise.

A esteganálise faz o caminho inverso à esteganografia, tentando revelar a presença da informação escondida. As técnicas de esteganálise procuram por possíveis alterações causadas pela mensagem escondida.

A detecção de mensagens ocultas em imagens não é uma operação trivial, pois as alterações nas mesmas são muitas vezes imperceptíveis ao olho humano. Entretanto, as técnicas de esteganografia possuem falhas, ou deixam padrões que podem ser detectados (Popa, 1998). É a partir dessas falhas que são baseadas as técnicas de esteganálise em imagens.

A proposta desta monografia é um estudo do desenvolvimento das principais abordagens utilizadas na detecção de mensagens escondidas em imagens digitais, conforme a evolução das abordagens de esteganografia.

Esta monografia está organizada em 6 capítulos. No Capítulo 2 são apresentados os fundamentos da esteganografia em geral e da esteganografia em imagens digitais. Também no Capítulo 2 são reunidas algumas das principais técnicas de esteganografia em imagens digitais. No Capítulo 3 são apresentados os fundamentos da esteganálise em imagens digitais e são apresentadas algumas das principais técnicas de esteganálise em imagens digitais e suas limitações. Por fim, o trabalho é concluído no Capítulo 4.

2 Esteganografia

A esteganografia é uma arte antiga. Suas origens remontam à antiguidade.

A palavra esteganografia deriva do grego (estegano = "esconder"e grafia ="escrita") (Petitcolas et al, 1999).

Ao longo dos anos foram apresentadas diversas formas de esteganografia. Um exemplo aparece em uma das estórias de Heródoto, datada por volta de 440 a.C. Nesta história um homem chamado Harpagus matou uma lebre e escondeu uma mensagem em sua barriga. Depois, Harpagus entregou a lebre a um mensageiro disfarçado de caçador. Já na obra literária "Arte do amor" de Ovídio, existe uma menção do uso de um leite para produzir textos invisíveis. Para "decodificar"a mensagem, o destinatário polvilhava fuligem ou fumo no papel que fixaria ao resíduo do leite (Popa, 1998).

Entre os séculos 16 e 17 surgiu uma vasta literatura a respeito da esteganografia. O livro *Schola Steganographica* de Gaspar Schott (1608-1666) mostra como esconder mensagens em partituras onde cada nota corresponde a uma letra. Outro exemplo seria o método apresentado por John Wilkins (1614-1672) que para esconder uma mensagem em um desenho geométrico usava pontos, linhas e triângulos para representar várias letras (Petitcolas et al, 1999).

Mensagens ocultas também foram usadas durantes as guerras. Na 1ª Guerra Mundial os alemães colocavam pontinhos quase invisíveis acima das letras em revistas. Eles também colocavam letras pontilhadas com tintas invisíveis e quando o papel era aquecido as letras podiam ser vistas (Petitcolas et al, 1999).

Já na 2ª Guerra Mundial, o espião nazista George Dasch escrevia mensagens em seu lenço usando uma solução de sulfato de cobre, que permaneciam invisíveis até serem expostas a vapores de amônia.

Também durante a 2ª Guerra Mundial, foi desenvolvida pelos nazistas um meio de esconder informação que ficou conhecido como tecnologia de microponto. Os nazistas fotografavam suas mensagens, que eram reduzidas ao tamanho de um ponto final de uma frase. Estes "pontos" eram introduzidos em uma mensagem totalmente inocente.

Atualmente, a esteganografia migrou para o meio digital aperfeiçoando técnicas clássicas, surgindo, assim, a esteganografia digital, que abre mão de tintas invisíveis, borrifar em revistas etc, e usa diferentes mídias para esconder a informação. Com isso, a esteganografia ganhou muito em termos de praticidade, facilidade e durabilidade (Rocha, 2006).

Existem diversas maneiras de se esconder uma mensagem, mas em geral é preciso um hospedeiro, ou seja, o objeto digital (áudio, vídeo, texto ou imagem) que irá levar a mensagem que é realmente importante. Esse objeto é conhecido como mensagem de cobertura (*cover-message*) e a mensagem oculta é chamada de dado embutido (*embedded data*). Quando a mensagem é inserida no objeto-digital, tem-se o chamado estego-objeto (stego-object) (Rocha et al, 2003).

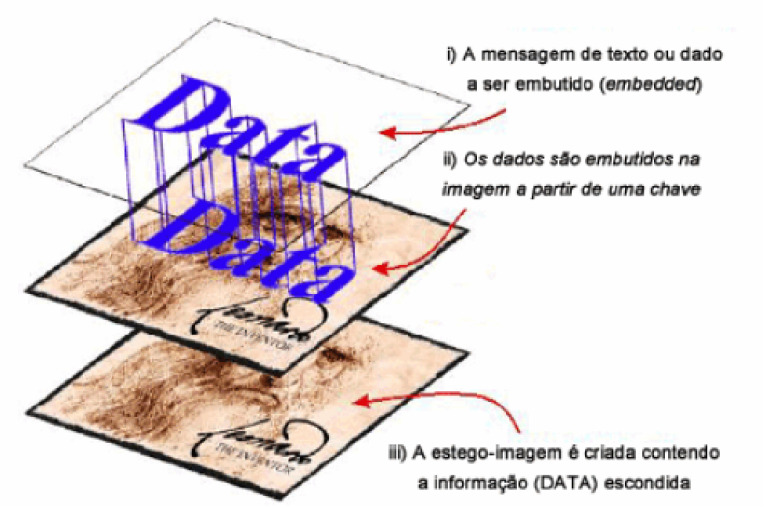

Figura 2.1: O processo de esteganografia (Rocha et al, 2003)

Os principais algoritmos de esteganografia digital utilizam a substituição de componentes de ruído de um objeto digital por uma mensagem secreta pseudo-randômica (Popa, 1998). Os objetos de cobertura podem ser divididos em duas classes (Rocha, 2006):

- *Stream Cover*: é um fluxo de dados contíguos como em uma linha telefônica, podendo ser, por exemplo, um arquivo de áudio. Neste caso normalmente não é possível dizer quando termina e quando começa a mensagem com antecedência. Este fluxo

é gerado a partir de um *keystream generator*, algo como uma chave de criptografia que diz a ordem de inserção dos bits que serão recuperados;

- *Random Access Cover*: as informações são inseridas em qualquer ordem particular e com definição prévia do seu tamanho e conteúdo, sendo que seu comprimento é menor que o fluxo de dados.

As aplicações para ocultar informações são variadas e algumas delas são citadas em (Petitcolas et al, 1999). Em agências militares e de inteligência a discrição é garantida em suas comunicações usando, além da criptografia, técnicas como transmissão por modulação de dispersão de espectro para tornar os sinais mais difíceis de serem detectados pelo inimigo.

Para bular tentativas de alguns governos de limitar a liberdade de expressão online, são usadas técnicas de comunicações anônimas na rede, incluindo *Anonymous Remailer*[1] e web *proxies*. Movimentações financeiras pela Internet e esquemas de eleição precisam de um alto nível de segurança. Assim o uso da esteganografia é um fator importante para garantir a integridade dessas ações. Para empresas, a esteganografia é útil no processo de marcação (*watermarking*) e identificação por digitais (*fingerprinting*), que protegem os direitos autorais e buscam restringir a pirataria (Julio et al, 2007).

Por outro lado, a esteganografia também é usada por criminosos como modo de esconder informações das autoridades, como em registros ocultos de atividades fraudulentas ou de dados relacionados a espionagem industrial. Em 2000, foi divulgado largamente pela imprensa norte-americana que terroristas estariam supostamente usando esteganografia na Internet para trocar informações (Provos, 2003).

Depois do ataque terrorista ao World Trade Center em 11 de setembro de 2001 foi divulgado que o grupo terrorista Al Qaeda, que era liderado por Osama bin Laden, estaria se comunicando com suas células espalhadas pelo mundo através de mensagens escondidas em imagens digitais. Tais imagens estariam sendo distribuídas através de chats, grupos de discussão, e-mails, leilões eletrônicos entre outros meios (Rocha et al, 2003).

Um caso recente é do traficante Juan Carlos Abadia que passava mensagens de

[1] serviço que envia mensagens de correio eletrônico anônimas a diferentes destinatários e apaga toda a informação referente ao usuário antes de enviar a mensagem

texto e de voz escondidas em imagens da "Hello Kitty"[2]. Tais mensagens seriam as ordens para movimentação de cocaína entre os países onde o traficante atuava. Há também casos em que a esteganografia tem sido utilizada para divulgar imagens de pornografia infantil na Internet (Julio et al, 2007).

Neste contexto de ações impróprias à sociedade, surgiu a área conhecida como Forense Digital:

> *Um conjunto de técnicas e procedimentos que utilizam conhecimento científico para coletar, analisar e apresentar evidências que possam ser utilizadas em um tribunal* (Pereira et al, 2009).

Nesta área, cabe ao perito analisar os dados em busca de possíveis evidências ocultas. A busca por meios de se descobrir a existência de mensagens escondidas por esteganografia é denominada esteganálise, que é um dos objetos de estudo deste trabalho, analisada no capítulo 3.

2.1 Esteganografia em imagens digitais

A imagem é a mais popular cobertura usada para se esconder uma mensagem (Julio et al, 2007a). Os diversos formatos desse arquivo digital permitiram o surgimento de várias técnicas de esteganografia para cada um. A seguir uma breve descrição da imagem digital.

2.1.1 Imagem Digital

A imagem pode ser representada por uma função bidimensional de intensidade da luz $f(x, y)$, onde x e y denotam as coordenadas espaciais e o valor de f em qualquer ponto (x,y) é proporcional ao brilho (ou tons de cinza) da imagem (Gonzalez, 2002).

A imagem digital é a representação discreta, tanto em coordenadas espaciais quanto em brilho, desta função bidimensional. Uma imagem digital pode ser considerada como sendo uma matriz cujos os índices de linha e coluna identificam um ponto da imagem, e o correspondente valor do elemento da matriz identifica o brilho. Os elementos dessa matriz são chamados pixels (Gonzalez, 2002). A Figura 2.2 mostra (a): uma imagem

[2]personagem de desenho animado

digital com um grupo de pixels selecionados, (*b*): o zoom da janela de 9×9 pixels e (*c*): a matriz com os valores da escala em tons de cinza.

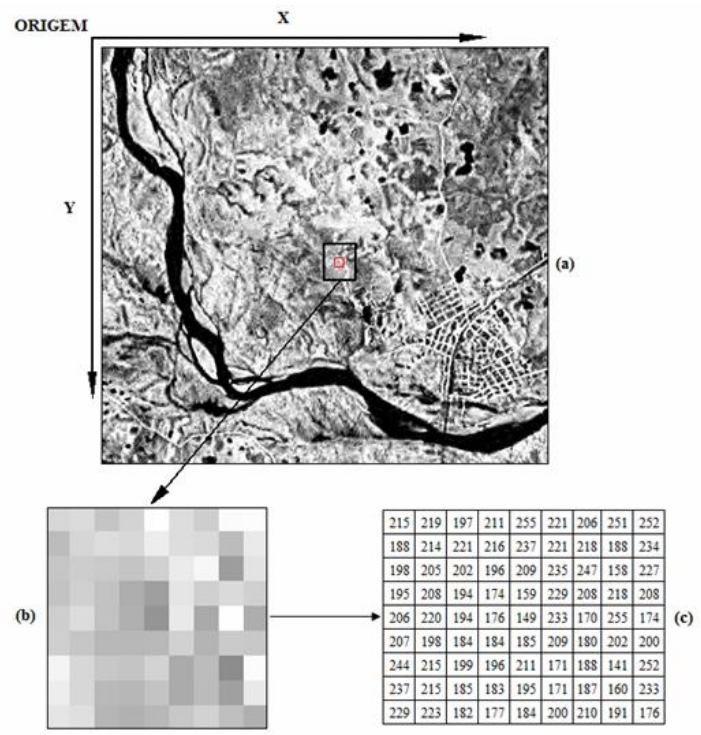

Figura 2.2: Exemplo de imagem digital [3]

Cada pixel tem um número de bits (*bit depth*) também conhecido como profundidade de cor. Uma imagem pode ser representada com diferentes sistemas de cores. Em imagens com o sistema de cores RGB, cada pixel pode ter 24 bits com cada *byte* representando um canal das cores primárias vermelho (*Red*), verde (*Green*) e azul (*Blue*) ou com 32 bits com adição do canal alfa (*alpha transparency*). Uma imagem digital também pode ser composta em escala em tons de cinza. Neste caso, cada pixel tem 8 bits.

Além disso, existem vários formatos de arquivos de imagem, como por exemplo o GIF *(Graphical Interchange Format)*, JPEG *(Joint Photographic Experts Group)* e PNG *(Portable Network Graphics)*, que fazem uso de métodos de compressão (Gonzalez, 2002).

A compressão pode ser com perda de dados, que descarta o excesso de dados da imagem, removendo detalhes menos sensíveis ao olho humano, ou sem perdas, que não descarta nenhum dado e sim os representa matematicamente, reduzindo o tamanho do arquivo. Dentre as principais técnicas de compressão, pode-se citar redução de domínio,

[3]Fonte: http://www.ufrgs.br/engcart/PDASR/estrim.html

redução do espaço de cor, codificação preditiva, codificação por transformadas (Gonzalez, 2002).

O JPEG é um dos formatos mais usados na Internet. Sua codificação funciona da seguinte forma: primeiro a imagem é dividida em blocos de pixels com dimensão 8×8. Caso a imagem esteja no sistema de cores RGB, pode ser feita a conversão para um outro sistema de cores chamado YCbCr, com Y como a luminância (quantidade de luz branca), Cb e Cr com nível de crominância (matiz e saturação). O objetivo dessa conversão é diminuir a proporção de crominância a qual o olho humano é menos sensível. Então, para cada bloco é feito o calculo da transformada discreta do cosseno.

A DCT (*Discrete Cosine Transform*) é uma transformada matemática muito utilizada no processamento de imagens. A transformada DCT passa a imagem para o domínio da frequência.

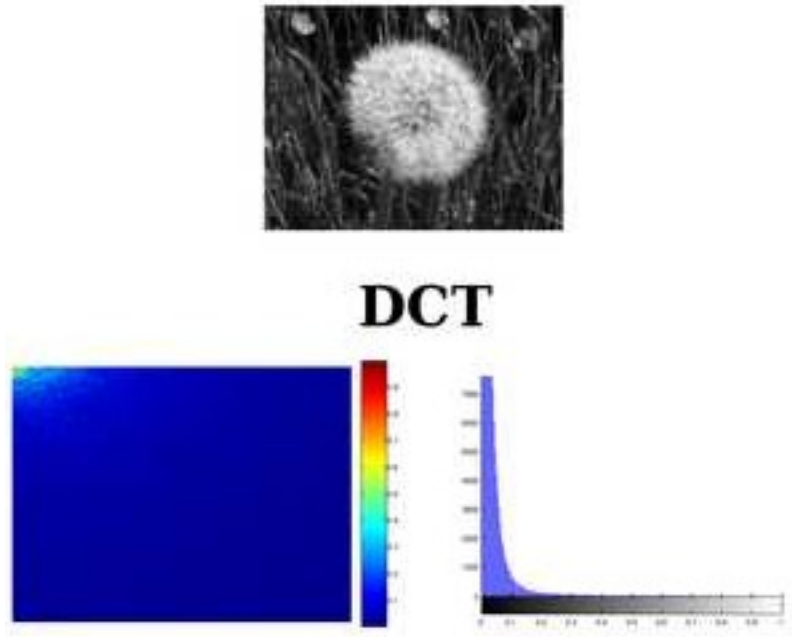

Figura 2.3: Coeficientes DCT da imagem (Julio et al, 2007a)

Para cada bloco 8×8 são gerados 64 coeficientes através da transformada DCT. Os coeficientes DCT são classificados como: o coeficiente DC, que é o valor médio do sinal da imagem e os coeficientes AC que são os demais coeficientes diferentes de zero. Assim, os 64 coeficientes de cada bloco são quantizados e os valores destes coeficientes são arredondados. Depois, são selecionados os coeficientes em que o olho humano é mais

sensível e o demais são descartados. Por fim, é aplicada a codificação de Huffman para os coeficientes restantes (Gonzalez, 2002).

Na Figura 2.3 pode se observar o acúmulo dos coeficientes mais significativos no canto superior direito da imagem da DCT, proporcionando melhor capacidade de compressão (Julio et al, 2007a).

Outro formato que também é popular na Internet é o GIF, que diferentemente do JPEG, usa compressão sem perdas. Utilizando uma paleta de 256 cores (*Global color table* e *Local color table*), em que cada cor tem 24 bits, os pixels passam a ser representados pelos índices da paleta, que tem somente 8 bits o que resulta em uma compressão de 3:1.

Algumas versões do formato GIF suportam um recurso chamado de entrelaçamento. O entrelaçamento salva a imagem em linhas entrelaçadas, dessa forma as informações das próximas linhas são salvas primeiro e, em seguida, são salvas as informações da linha atual. A ideia é obter uma noção geral da imagem sem ter que esperar até toda a imagem ser carregada (Popa, 1998).

Quando uma imagem é armazenada no computador há possibilidade de falhas durante este processo. Essas falhas acrescentadas a imagem são denominadas ruídos. Existem diversos tipos de ruído e para cada tipo existem meios de eliminá-los ou amenizá-los. Os tratamentos desses ruídos são chamados filtros (Gonzalez, 2002).

Há também alguns tipos de ruído que não causam alterações visíveis na imagem. Isso acontece porque, o sistema visual humano (HVS) possui algumas limitações. Essas limitações são exploradas pela esteganografia para garantir que as modificações feitas na imagem também passem despercebidas.

Com isso, a mensagem pode ser inserida tanto, no domínio espacial quanto no domínio da frequência da imagem usando qualquer pixel ou blocos de pixels, que podem ser pequenas mudanças na intensidade de luminância, mudanças no coeficientes da transformada, alterações no brilho.

Em geral, os bits da mensagem escondida são inseridos nos bits menos significativos da imagem. Estes bits têm algumas propriedades estatísticas. A alteração de alguns destes bits pode resultar na perda destas propriedades. Assim se o propósito é ocultar uma mensagem, esta deve causar o mínimo de perturbação possível na imagem (Wayner,

2002).

Para garantir, então, que o dado embutido não cause alterações significativas na imagem, pode-se usar, por exemplo, um método de seleção que gera um grande número de estego-objetos e escolhe aquele com a menor variação nas propriedades dos bits menos significativos. Outra opção seria um método construtivo que usa uma função que tem como objetivo fazer com que os bits da mensagem oculta fiquem o mais próximo possível da forma dos bits da mensagem de cobertura (Rocha et al, 2003).

Na próxima seção são citadas técnicas de esteganografia que usam uma imagem como cobertura gerando a chamada estego-imagem. Tais técnicas podem ser específicas para formatos, sistemas cores ou qualquer outra característica da imagem.

2.2 Técnicas de Esteganografia em Imagens

Existem várias formas inserir informações em uma imagem . Isso porque sua estrutura e seus dados redundantes resultam em numerosas possibilidades. Algumas delas podem ser realizadas manualmente, como no caso da técnica de geração da imagem (*Image Generation Technique*) em que a mensagem é escondida modificando os elementos da imagem, como por exemplo, mudando a cor dos olhos ou do cabelo de uma pessoa na fotografia. Isto pode ser realizado usando qualquer programa de edição de fotos (Hamid et al, 2012). Já outras geram a estego-imagem a partir de algoritmos que escolhem regiões da imagem e procuram o melhor meio de encaixar a mensagem.

Há muitas maneiras de se classificar os métodos de esteganografia, e uma encontrada na literatura divide as abordagens em três abordagens as citadas a seguir (Rocha et al, 2003):

- LSB (*Least Signifcant Bit*): usa os bits menos significativos de cada pixel da imagem que pode ser codificada com 24 bits, ou com 32 bits com adição do canal alfa. Contudo, estas técnicas acabam sendo vulneráveis à manipulação de imagens que alteram o plano LSB, como por exemplo a compressão.

  Na Figura 2.4 são ilustrados três pixels de uma imagem de cobertura. Como exem-

(00100111 11101001 11001000 10001100) [A,R,G,B]
(10100111 11001000 11101001 11101000) [A,R,G,B]
(11001000 00100111 11101001 00100111) [A,R,G,B]

Figura 2.4: Três pixels de uma imagem de cobertura(Silva et al, 2010)

(00100110 11101001 11001000 10001100) [A,R,G,B]
(10100110 11001000 11101000 11101001) [A,R,G,B]
(11001000 00100111 11101001 00100111) [A,R,G,B]

Figura 2.5: Os pixels após inserção no LSBs (Silva et al, 2010)

plo, a letra A codificada em ASCII[4] como 01000001 é inserida nestes pixels. O resultado da inserção no LSBs é mostrado na Figura 2.5. Os bits em preto representam o LSBs e os bits em vermelho as modificações realizadas para inserir a letra A.

- Técnicas de Mascaramento: Estas técnicas são mais robustas e geram estego-imagens resistentes à compressão e recorte. Devem ser aplicadas em imagens em tons de cinza, não sendo eficazes em imagens coloridas, pois como usam bits de pixels mais significativos, a detecção torna-se mais fácil (Popa, 1998). Estas técnicas escondem a informação através da criação de uma imagem semelhante às marcações de *copyright* em papel (Rocha et al, 2003).

- Algoritmos de Transformação: os algoritmos de transformação geralmente trabalham com formas mais sofisticadas de manuseio de imagens como brilho, saturação e compressão das imagens. Utilizam transformadas matemáticas, como a transformada discreta do cosseno, transformada discreta de Fourier e transformada Z, para converter a imagem no domínio da frequência e inserem a mensagem neste domínio. Quando inserida no domínio da frequência, a mensagem pode ser espalhada por toda a imagem.

Algumas da técnicas mais populares desenvolvidas ao longo dos anos são descritas a seguir:

[4]codificação de caracteres de oito bits baseada no alfabeto inglês

2.3 Método LSB

O método LSB, conhecido também com técnica de substituição, é um método simples e popular (Hamid et al, 2012). Como já dito neste Capítulo, o método LSB insere a mensagem nos bits menos significativos dos pixels da imagem. Mas uma mensagem costuma ser menor que o plano LSB da imagem, então, é preciso inserir os bits da informação de forma que produza o mínimo de alterações na imagem.

2.3.1 Método LSB usando uma chave secreta

A ideia básica deste método é a utilização de uma permutação pseudo-aleatória dos bits da imagem de cobertura. Os dados da mensagem são inseridos através da modificação dos bits selecionados. Seja N o número de bits disponíveis da imagem e seja P_0^N uma permutação dos números de 0 a $N-1$. Tendo uma mensagem de tamanho n, é possível inserir esta mensagem nos bits da imagem de cobertura usando os elementos da permutação: $P_0^N(0), P_0^N(1), ..., P_0^N(n-1)$. A função de permutação deve ser pseudo-aleatória, tendo que selecionar os bits em uma ordem aparentemente aleatória. Como resultado, os bits escondidos serão espalhados por toda imagem de cobertura (Popa, 1998).

Além disso, a função de permutação deve depender de uma chave secreta K. Portanto, é necessário um gerador de permutação pseudo-aleatório, que tem como entrada a chave K e produz como saída diferentes sequências de $0, ..., N-1$. Este gerador deve ser seguro para que ninguém possa adivinhar em quais bits de cobertura os dados da mensagem foram incorporados sem conhecer a chave K.

O gerador pode usar uma função pseudo-aleatória. Esta função produz diferentes funções imprevisíveis de acordo com valor de cada chave secreta. Ela pode ser facilmente construída a partir de qualquer função *hash* H segura concatenando o argumento i ($i = 0, ..., n-1$) com a chave secreta K, como mostra a equação 2.1

$$f_K(i) = H(K \circ i) \tag{2.1}$$

, onde $(K \circ i)$ é a concatenação da chave K com o argumento i. Assim, obtém-se uma função pseudo-aleatória. $f_K(i)$ que depende do parâmetro K.

Este método, entretanto, não pode inserir mensagens longas sem degradar a imagem. Para cada imagem de cobertura há um limite para o tamanho da mensagem. Isto depende do tamanho da imagem e se a imagem tem ou não áreas de textura. Outro problema do método LSB é a possibilidade da informação ser perdida no processo de compressão (Popa, 1998).

2.3.2 Métodos LSB adaptativos

Para garantir que as propriedades estatísticas da imagem não sejam alteradas significativamente, foram desenvolvidas técnicas que procuram ao invés de simplesmente inserir a mensagem de forma pseudo-aleatória, insere da melhor forma os dados da informação, analisando primeiramente as propriedades e estrutura da imagem ((Johnson, 2000) :(Hamid et al, 2012)).

Um exemplo deste tipo de abordagem seria o método desenvolvido em (Vieira et al, 2010) para otimizar a substituição de bits no plano LSB.

Esta abordagem trabalha com uma matriz de substituição que mapeia os bits menos significativos da imagem que foram substituídos pelos bits da mensagem. Encontrar a matriz de substituição que melhor se encaixa no plano LSB de cada imagem não é trivial, pois existe uma grande quantidade de possibilidades. Com isso, o método tenta encontrar a matriz ótima através do uso de Algoritmo Genético (Back, 1996) com *path relinking* (Glover, 1997).

Outro exemplo é a técnica SLSB, vista em Roque (2009), que ao invés de inserir os bits da mensagem nos três componentes RGB de cada pixel da imagem, seleciona-se somente um dos componentes resultando em uma menor distorção no plano LSB da imagem. O método primeiramente faz uma análise preliminar das cores e seleciona a cor que possui mais cores próximas porque tal cor apresenta maior diversidade o que leva a mudanças mais imperceptíveis.

Depois,o método procura adicionar o bit da mensagem no componente de cor de cada pixel de forma que cause a menor alteração possível. Para isso, é feito o cálculo da distância entre a cor original e cor esteganográfica. No caso da distância ser maior que o limiar (determinado pelo número de bits a ser escondido) a cor é decrementada

para chegar a uma cor final mais próxima da original, o quem implica numa redução da distorção causada pela mensagem inserida.

Por exemplo, usando o byte 11001000 para esconder os 3 bits (111) da mensagem. Com o método LSB simples o resultado será 11001111, em que a distância entre o este byte e o byte original é igual 7. Agora usando o método proposto em Roque (2009) neste exemplo o 4º bit é decrementado pois 3 bits foram inseridos, chegando ao resultado 11000111 que possui a distância igual a 1.

2.4 O algoritmo Patchwork

Patchwork foi um dos primeiros métodos de esteganografia para marca d'água. Este método é aplicado no domínio espacial e utiliza um método estatístico para adicionar a mensagem através de codificação de um padrão redundante (Morkel et al, 2005). O algoritmo acrescenta redundância a mensagem escondida e depois a espalha ao longo da imagem.

O Patchwork escolhe duas regiões na imagem, A e B com N pixels (que é determinado dependo da chave secreta) de forma aleatória na imagem. Os pixels de A tem o valor de brilho incrementado com um pequeno valor de x unidades e os pixels de B tem o valor de brilho decrementado com o mesmo valor de x de unidades. Em outras palavras, a intensidade dos pixels da região A da imagem são aumentadas por um valor constante, ao passo que a intensidade dos pixels da região B são reduzidos com o mesmo valor constante. Assim, as alterações de contraste neste subconjunto da região codificam um bit da mensagem. Estas mudanças são geralmente pequenas e imperceptíveis, enquanto não mudar a luminosidade média da imagem.

Para recuperar a mensagem primeiro usa-se a chave para saber quais regiões da imagem foram usadas. Com isso, calcula-se a diferença de intensidade entre as regiões A e B. O resultado esperado desta diferença geralmente é zero, mas neste caso, com as alterações feitas no processo de inserção, o resultado da diferença será $2x$. Assim, um bit específico da mensagem é dado com 1 caso a diferença seja $2x$ e 0 caso contrário.

A desvantagem desta técnica é que apenas um bit é adicionado. Para contornar esse problema, pode-se dividir a imagem em sub-imagens e ir aplicando o algoritmo em

cada uma delas, incorporando a mensagem (Popa, 1998). Este método garante que a mensagem fique bem distribuída por toda imagem de cobertura e é ideal para mensagens pequenas, podendo sobreviver a processo de compressão com ou sem perda de dados.

2.5 Modulação de Amplitude

Proposto por Kutter et al (1997), este método usa a modulação de amplitude para incorporar bits de assinatura individual, modificando os valores dos pixels no canal de cor azul, que é o canal ao qual o olho humano é menos sensível. Estas modificações podem ser aditivas ou subtrativas, dependendo do valor do bit e é proporcional à luminância. O processo de decodificação não requer a existência da imagem original.

O método funciona da seguinte maneira, tendo s como um bit único a ser escondido em uma imagem $I = R, G, B$ e $p = (i, j)$ uma posição pseudo-aleatória em I, gerada através de uma chave K dividida em duas partes. O bit a ser escondido é incorporado na posição p no canal B modificando sua luminância (que é obtida através das componentes do sistema RGB usando a seguinte equação $L = 0,299R + 0,587G + 0,114B$) como:

$$B_{ij} = B_{ij} + (2s - 1)L_{ij}q, \tag{2.2}$$

onde, q é uma constante que determina a força da assinatura e seu valor depende da finalidade da incorporação da mensagem. O valor de q pode garantir proteção contra detecção ou proteção contra remoção.

Agora, para recuperar o bit da mensagem escondida é realizada uma estimativa do valor original do bit. Essa estimativa é baseada em uma combinação linear dos pixels vizinhos de p. De acordo com Kutter et al (1997), a vizinhança em forma de cruz gera bons resultados.

A estimativa $\hat{B}_{ij}$ é calculada da seguinte maneira:

$$\hat{B}_{ij} = \frac{1}{4c}\left(\sum_{k=-c}^{c} B_{i,j+k} - 2B_{ij}\right) \tag{2.3}$$

em que c é o tamanho da vizinhança em forma de cruz.

Assim o resultado da diferença entre B_{ij} e $\hat{B}_{ij}$ é o que determina o valor do

bit da mensagem secreta. Note que a função de estimativa não é inverso da função de incorporação. Assim a recuperação do bit não é certa. Para reduzir a probabilidade de erros, o bit pode ser inserido diversas vezes na imagem.

Esse método é resistente a manipulações de imagem como *blurring*[5], composição com outra imagem.

2.6 Método SSIS

Comunicação *Spread Spectrum* é o processo de propagação da largura de banda de uma sinal de banda estreito através de uma frequência de banda (Marvel et al, 1999). Para isso é necessário fazer uma modulação na forma da onda da banda estreita com a onda de uma banda larga. Depois de espalhada, a energia do sinal da banda estreita em qualquer frequência é baixa, e portanto, difícil de detectar (Marvel et al, 1999). O SSIS usa esta técnica para esconder uma mensagem em imagens.

O método *Spread Spectrum Image Steganography* (SSIS) fornece a capacidade de esconder e recuperar uma quantidade significativa de bits de informação dentro de imagens digitais. Além disso, é um esquema cego, porque a imagem original não é necessária para extrair a mensagem secreta. A única coisa que o destinatário deve ter, para revelar a mensagem secreta é uma chave (Popa, 1998).

Sendo similar ao método LSB, que insere a mensagem aleatoriamente na imagem no plano LSB, o método SSIS espalha a mensagem secreta sobre o espectro de frequência da imagem. Além disso, combina técnicas de comunicação de espectro de propagação de erros, codificação de controle, processamento de imagem.

Pelo SSIS, os dados da mensagem são incorporados no ruído, o qual é adicionado à imagem original.

A mensagem m é codificada com uma chave 1 opcional codificada através de um código de correção de erro (*low rate error-correcting code*). O remetente também fornece uma segunda chave, chave 2, para gerar uma sequência de espalhamento n.

O esquema de modulação é usado para espalhar o espectro de banda estreita da mensagem m com a sequência de espalhamento, isto de modo a compor o sinal incorporado

[5]processo em que a imagem é borrada para a remoção de pequenos detalhes.

s, que é a entrada para um intercalador e espalhador espacial usando a chave 3. O sinal obtido é adicionado a imagem original f para produzir a estego-imagem g, que, de alguma maneira, transmitido ao destinatário. A Figura 2.6 demonstra este processo.

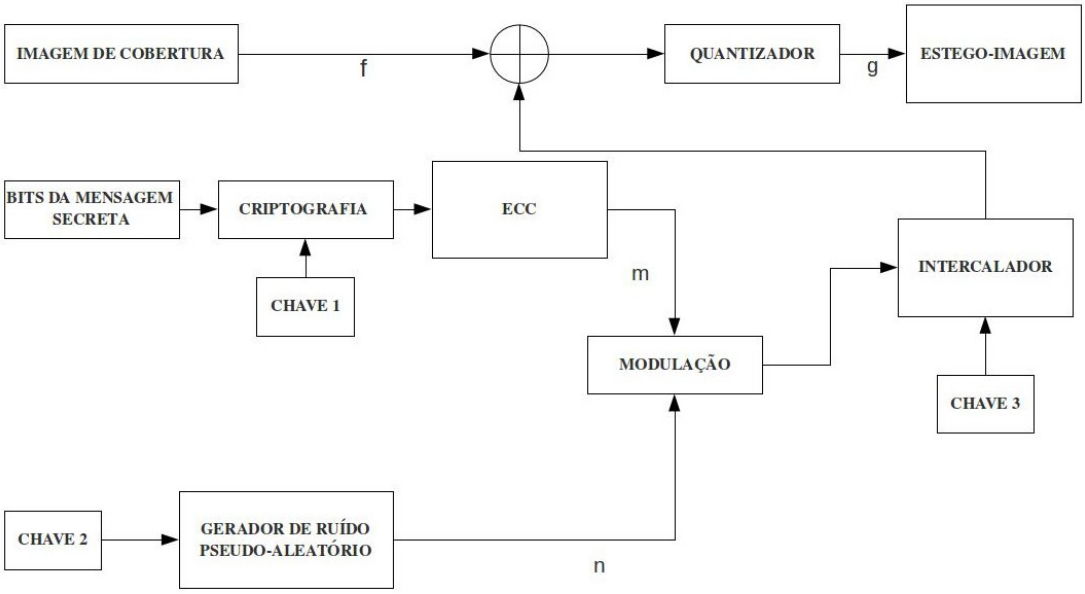

Figura 2.6: Esquema do processo do SSIS adaptado de (Marvel et al, 1999)

Quem recebe a mensagem tem as mesmas chaves e a estego-imagem e para ler a mensagem utiliza decodificador. Esse decodificador usa uma estimativa da imagem de cobertura, $\hat{f}$, a partir da imagem recebida $\hat{g}$. A diferença entre $\hat{f}$ e $\hat{g}$ é utilizada por um desintercalador de chaves (*keyed deinterleaver*) para construir uma estimativas do sinal adicionado $\hat{s}$ a partir da chave 3. Com a chave 2 a sequência de espalhamento n é reconstruída e, em seguida, usando demodulação uma estimativa da mensagem m é construída usando uma baixa taxa de controle de erro.

2.7 Técnicas que usam o domínio da frequência

Técnicas que usam o domínio da frequência da imagem para esteganografia são mais resistentes às modificações na imagem. Para esconder uma mensagem, estas técnicas agem no domínio da frequência, em que a imagem é transformada. A transformação da imagem é feita, geralmente, usando a (DCT). Outras transformações podem ser usadas como transformada discreta de Fourier e transformada Z.

A transformação pode ser aplicada em partes ou na imagem toda. O algoritmo consiste em em modificar alguns coeficientes de acordo com a mensagem a ser escondida. Normalmente, os coeficientes a serem modificados pertencem à uma gama média do espectro de frequências. O objetivo é encontrar um equilíbrio entre invisibilidade e robustez (Popa, 1998).

Alguns exemplos de técnicas de esteganografia que utilizam o domínio da frequência são descritas a seguir.

2.7.1 Técnicas que usam o DCT

Algoritmos que usam a transformada DCT geralmente são destinados a lidar com imagens no formato JPEG. A técnica proposta por Barni et al (1997) consiste em sobrepor a mensagem em alguns coeficientes DCT da imagem.

Para entender o processo tem-se, por exemplo, uma mensagem representada de forma discreta pela sequência: $X = x_1, x_2, ..., x_M$ em que, M é o tamanho da mensagem. Primeiramente são calculados os coeficientes DCT da imagem I. Os primeiros $L + M$ coeficientes são selecionados para gerar o vetor $T = t_1, t_2, ..., t_M$, em que L é a quantidade de coeficientes que são descartados para garantir a invisibilidade e robustez necessária. Depois a mensagem X é embutida nos M coeficientes restantes. Deste jeito um novo vetor, $T' = t'_1, t'_2, ..., t'_M$, é obtido, com a seguinte fórmula:

$$t'_i = \begin{cases} t_i & i = 1, ..., L \\ t_i + \alpha \cdot t_i \cdot x_i & i = L + 1, ..., L + M \end{cases} \tag{2.4}$$

em que, α é um valor, que é escolhido de acordo com nível de robustez desejado.

Esta técnica demonstrou ser resistente à manipulações, como filtros passa-baixa[6], redimensionamento, filtro mediana[7], *blurring* e ruídos, sendo um dos melhores métodos contra remoção da mensagem escondida (Popa, 1998)

Os aplicativos JSteg/JPHide e OutGuess usam a técnica de domínio da frequência DCT como base.

[6] filtros passa-baixa atenuam ou eliminam os componentes de alta frequência de uma imagem
[7] o filtro mediana tem como objetivo reduzir o ruído da imagem

2.7.2 Técnicas que usam a Transformação Wavelet

A Transformada discreta Wavelet (DWT) é uma função matemática que assim como DCT, também é usada no processo de compressão de imagens. O interesse nesta transformada vem da capacidade de representar sinais que possuem características diferentes para instantes e domínios espaciais diferentes (Hamid et al, 2012). Sua representação da imagem é em multiresolução.

Um dos exemplos é visto em Abdelwahab et al (2008), que aplica a DWT em blocos 4×4, da mensagem e da imagem de cobertura e faz comparações para determinar quais são as melhores combinações entre os blocos para inserir a mensagem.

Figura 2.7: Exemplo de aplicação da DTW (Luo, 2008)

2.7.3 YASS - *Yet Another Steganographic Scheme*

Este método também usa DCT para inserir mensagens em imagens JPEG, mas ao invés de inserir diretamente nos coeficientes DCT gerados para compressão, escolhe regiões aleatórias da imagem. Para isso, a imagem é dividida em blocos de pixels grandes de tamanho $B \times B$. Para cada bloco são selecionados aleatoriamente blocos de tamanho 8×8, onde o dado da mensagem será escondido. A mensagem é escondida em uma banda de baixa frequência AC, usando códigos de erros. Por fim, a imagem é compactada no formato JPEG (Solanki, 2007).

2.8 Algoritmos de distorção

Para este método são necessárias tanto a estego-imagem quanto a imagem original. Primeiramente, são inseridas distorções ao longo da imagem. Depois, as duas imagens são comparadas, se houver alguma diferença entre o pixel da estego-imagem e a imagem original, significa que o bit da mensagem escondida é 1 e caso contrário 0 (Johnson, 2000).

2.9 Inserção no cabeçalho

Existem formatos de imagens como TIFF, GIF, PNG e WMF que possuem um cabeçalho (*file header*), que contém informações sobre a imagem (resolução, profundidade de pixel, tipo de compressão etc). Assim, é possível inserir a mensagem no cabeçalho sem causar alterações na imagem. Contudo, esta técnica é frágil, já que uma vez que a mensagem é detectada, basta salvar a imagem no mesmo formato e a imagem será destruída (Hamid et al, 2012).

2.10 Seleção de cobertura para esteganografia

Quem deseja esconder uma mensagem pode escolher qual será a imagem de cobertura. A ideia desta abordagem é escolher a imagem de cobertura que melhor se adapta a mensagem, ou seja, não deixa alterações detectáveis (Kharrazi et al, 2006). Assim trabalha-se com os três possíveis cenários, ao esconder uma mensagem, que são: não possuir nenhuma informação sobre o processo de detecção, possuir informações parciais e possuir total informações a respeito do processo de detecção

Caso se encontre no primeiro cenário, é feita uma análise das características da imagem que podem ser alteradas pela mensagem e é escolhida a imagem com menor taxa de alteração. Agora se possuir informações parciais do processo de detecção, como por exemplo, possuir a entrada e a saída do método de esteganálise, mas não os detalhes em si, é possível usar estas informações para traçar um limiar para seleção da cobertura. Por fim, quando se sabe qual processo de detecção será usado, usa-se o mesmo pra identificar em qual imagem de cobertura a mensagem passará despercebida.

2.11 Aplicativos de Esteganografia

Na internet existem vários aplicativos de esteganografia em imagens, sendo que, algumas ferramentas usam diferentes técnicas e outras são específicas para um formato de imagem. Segue uma breve descrição de algumas destas ferramentas:

- *Hide and Seek*[8]: é um dos aplicativos mais antigos disponíveis, está na versão 5.0. Desenvolvida por Colin Maroney, suporta imagens GIF e tem um limite para o tamanho da imagem;

- *Ezstego e Stego Online*[9]: aplicativo Java que suporta imagens GIF e PICT, usa o método LSB e reorganiza a tabela de cores.

- *OutGuess*[10]: desenvolvido por Niel Provos, este aplicativo está disponível para UNIX. Sua versão atual (*OutGuess 0.2*) suporta os formatos PNM (*Portable Anymap Format*) e JPEG. Para o formato JPEG as propriedades estatísticas são preservadas no domínio da frequência;

- *S-tools*[11]: este aplicativo para Windows permite esconder mensagens em imagens no formato GIF e BMP e oferece a opção de criptografar sua mensagem. Também oferece a opção de verificar o tamanho máximo da mensagem que pode ser inserida, dependendo da imagem escolhida;

- *Steghide*[12]: é um aplicativo para ocultar dados em vários tipos de imagem preservando as propriedades estatísticas das cores das imagens. A versão atual é *Steghide 0.5.1*;

- *JK-PGS (Pretty Good Signature)*: desenvolvido por Martin Kutter e Frédéric Jordan, este aplicativo que usa a técnica de modulação de amplitude (Kutter et al, 1997) para inserir mensagens em imagens no formato JPEG;

[8]disponível em http://www.nic.funet.fi/pub/crypt/steganography/
[9]disponíveis em http://www.nic.funet.fi/pub/crypt/steganography/
[10]disponível em http://www.outguess.org/
[11]disponível em ftp://ftp.funet.fi/pub/crypt/mirrors/idea.sec.dsi.unimi.it/code/
[12]disponível em http://steghide.sourceforge.net

- *Digimarc*[13]: usado para marca d'água, este aplicativo insere mensagens em diversos formatos de imagem e também em outra mídias. Atualmente existe uma versão disponível para dispositivos móveis.

A maioria destes aplicativos são atualizados para corrigir falhas descobertas pelas técnicas de detecção de imagens. No próximo capítulo são descritos como funciona a detecção de mensagens em imagens e as principais técnicas desenvolvidas.

[13]disponível em http://www.digimarc.com/digimarc-for-images

3 Esteganálise

O processo de esteganografia não é perfeito. Algumas falhas podem ser detectadas através de uma análise detalhada da imagem. Para descobrir se uma imagem é, na verdade, um estego-objeto as técnicas de esteganálise procuram por alterações nas características da imagem e na maioria das vezes somente a confirmação que existe uma mensagem escondida é suficiente, pois como as mensagens possivelmente são inseridas aleatoriamente e podem ter sido criptografadas torna-se mais difícil recuperar o seu conteúdo (Pereira et al, 2009).

Além disso, confirmando a existência da mensagem, é possível destruí-la sem ao menos ler. Ou ainda, a mensagem pode ser substituída por outra, ou até mesmo anulada, bastando modificar aleatoriamente o plano LSB (Wayner, 2002).

Em geral, abordagens para detectar mensagens, ou ataques, tentam identificar distúrbios, características da imagem que foram alteradas pela mensagem escondida. Em vários casos, acontece da mensagem ser mais aleatória que o dados que ela substituiu.

Como há várias formas de inserir uma mensagem, geralmente se tem um algoritmo de ataque para um algoritmo específico de esteganografia. Não há, entretanto, garantias de detecção. Pode acontecer dele funcionar bem para a técnica determinada, mas a menor atualização do algoritmo esteganográfico pode resultar em falhas no ataque. Além disso, não há garantias que o ataque possa se atualizar a cada nova versão da técnica de esteganografia. De acordo com Wayner (2002) :

Não existe nenhuma bala mágica anti-esteganográfica.

Todavia, como já foi mencionado, a esteganografia também não está imune a falhas. Algumas técnicas de esteganografia podem ser facilmente detectadas.

As abordagens para detectar mensagens podem ser divididas em três classes, conforme o tipo de ataque (Julio et al, 2007):

- Ataques Aurais: retiram partes significativas da imagem como um meio de facilitar aos olhos humanos a busca por anomalias na imagem. Quando os métodos de esteganografia alteraram significativamente as cores da imagem, é fácil perceber

fazendo uma pequena manipulação dos bits . Um exemplo é verificar os bits menos significativos em busca de padrões. Em uma imagem em preto e branco o plano LSB não costuma ser aleatório. A inserção de uma mensagem pode resultar em imperfeições nas áreas saturadas da imagem. As Figuras 3.1 e 3.2 exemplificam as modificações no plano LSB causadas pelo processo de esteganografia.

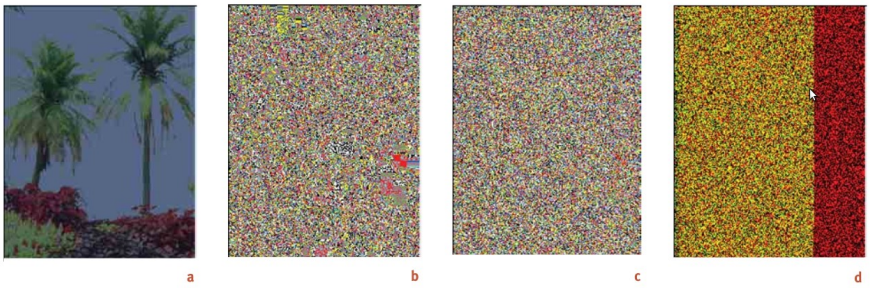

Figura 3.1: Modificações no LSB devido a Esteganografia (Wang, 2004)

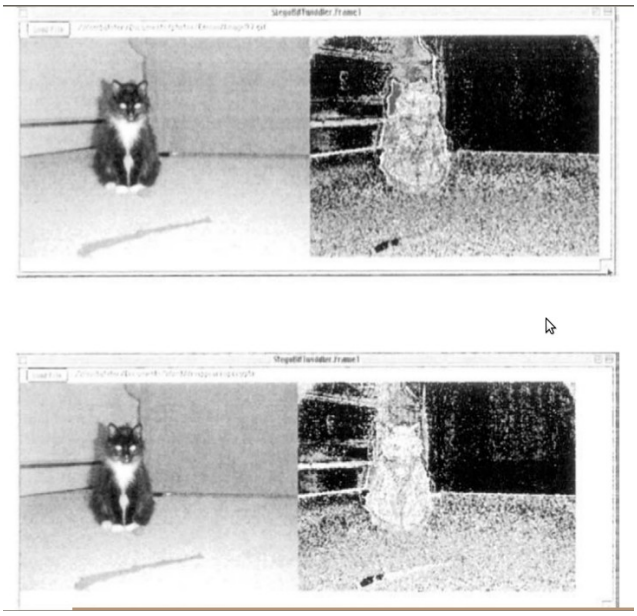

Figura 3.2: Exemplo de ataque aural (Popa, 1998)

Na Figura 3.1 (a) é uma estego-imagem, (b) é o plano LSB da imagem de cobertura utilizada, (c) o plano LSB da estego-imagem com os dados inseridos espalhados uniformemente e (d) o plano LSB da estego-imagem com os dados inseridos de forma sequencial e somente nos componentes vermelhos e em alguns componentes verdes.

Dependendo do formato da imagem, várias partes da imagem podem ter o mesmo

LSB. Como no caso do JPEG que em cada bloco codificado é armazenado o valor médio dos coeficientes DCT. No caso do formato GIF, as cores são representadas por índices para poupar espaço. Ambos os processos fazem com que o plano LSB não seja realmente aleatório (Wayner, 2002).

- Ataques Estruturais: Neste tipo de ataque é feita uma análise dos padrões estruturais da imagem, pois vários métodos de esteganografia alteram a estrutura da imagem. Por exemplo, em imagens indexadas (baseadas em paletas de cores), pode ser necessário usar diferentes versões de paletas. Com isso, há mudança nas características estruturais da imagem de cobertura. Logo as chances de detecção da presença de uma mensagem escondida aumentam;

  No caso de imagens no formato GIF e PNG, inserir a mensagem no LSB pode ser um erro, pois os índices da paleta de cores não estão, muitas vezes, perto o suficiente uns dos outros. Por exemplo o índice 01001001 pode ser azul escuro, enquanto o índice 01001000 pode ser um rosa escuro.

- Ataques Estatísticos: A estatística tem como uma das áreas de estudo determinar se algum fenômeno ocorre ao acaso. Diversas ferramentas estatísticas são usadas para esse propósito. Tais ferramentas podem servir também para identificar estego-imagens, porque, como já foi dito antes, a mensagem escondida poder ser mais aleatória que o dado que ela substituiu, ou seja, as mensagens inserem padrões que alteram as propriedades estatísticas da imagem (Rocha, 2006).

  Com isso, numa imagem, os padrões dos pixels e seus bits menos significativos frequentemente revelam a existência de uma mensagem secreta nos perfis estatísticos.

  Este tipo de ataque é o mais usado e costuma ter uma alta taxa de sucesso (Wayner, 2002). Várias análises estatísticas que podem ser feitas, como análise na paleta de cores, no padrão do LSB, nos coeficientes DCT, até uma análise baseada nas métricas de qualidade de imagens (IQM) foi desenvolvida. Algumas dessas técnicas serão detalhadas mais adiante.

Por fim, quem procura detectar uma mensagem pode se encontrar nas seguintes situações: possuir somente a imagem suspeita, pode possuir a imagem original e a imagem

suspeita e pode ter acesso a imagem e ao algoritmo de esteganografia, etc. Tudo isso vai se importante na hora determinar que tipo de ataque escolher.

Na Seção 3.1 as técnicas de esteganálise são descritas demonstrando os avanços que cada uma proporcionou para a detecção de mensagens escondidas em imagens.

3.1 Técnicas de Esteganálise

As técnicas de esteganálise são, em geral, análises estatísticas das características da imagem, que têm como resultado a probabilidade de haver uma mensagem inserida e em alguns casos uma estimativa do tamanho da mensagem. Tais técnicas exploram as falhas das técnicas de esteganografia.

Para realizar a detecção, as técnicas de esteganálise usam diversos recursos para analisar os dados das imagens. Alguns deles são modelos estatísticos de alta ordem que servem como uma meio de analisar os distúrbios causados pela esteganografia. Como no caso da abordagem apresentada em (Sullivan et al, 2005) que apresenta um modelo de cadeia de Markov para detectar o sinal da mensagem inserida pela técnica de esteganografia SSIS, vista anteriormente na Seção 2.6.

As técnicas de esteganálise podem seguir dois caminhos. Um deles é escolher um método específico de esteganografia e usar as falhas do método para detecção. O outro seria o que é chamado de esteganálise cega (*Blind Steganalysis*), onde não se tem conhecimento a respeito do método de esteganografia, mas sim o tipo de distorção que uma mensagem pode causar à imagem e a partir daí chega-se numa possível detecção.

A inserção de mensagens no plano LSB é a mais popular forma de esteganografia e existem diversas variações e adaptações (Julio et al, 2007a).

Assim, a maioria das técnicas de esteganálise visa detectar mensagens inseridas usando o LSB.

A Tabela 3.1 faz uma relação dos métodos de esteganálise e os processos ao qual são destinados. Estes métodos e suas limitações são descritos a seguir.

Tabela 3.1: Relação dos métodos de esteganálise e os processos aos quais são destinados

Métodos de Esteganálise	Tipos de análise	Método de Esteganografia
Teste do χ^2	análise dos coeficientes DCT da imagem	Esteganografia baseada em troca de pares de valores na escala de cinza do pixel, nas cores, ou em coeficientes DCT
Análise de cores do cubo RGB	análise da razão da quantidade de cores próximas	técnica LSB em imagens em cores reais
Análise RS	análise da correlação do plano LSB com os demais bits da imagem	várias técnicas de inserção no plano LSB
Taxa de inversão da energia do gradiente	análise da energia do gradiente da imagem	
Esteganálise Cega	análise baseada em diferentes propriedades estatísticas da imagem	várias técnicas de esteganografia

3.2 Teste do χ^2

Técnicas de esteganografia que inserem mensagens no LSB costumam alterar o histograma de frequência de cores (Provos, 2001). No caso de existir uma cor c_1 que possua maior frequência que alguma outra cor c_2, provavelmente c_1 terá seu LSB modificado mais vezes. alterando a frequência em que as cores são usadas na imagem. Assim a diferença entre c_1 e c_2 reduzirá.

Em uma imagem JPEG, os coeficientes da transformada DCT da imagem em questão, são analisados usando o teste do χ^2 .

O teste do χ^2 foi proposto por Karl Person (Rocha, 2006). Seu objetivo é comparar duas frequências de mesmo tamanho elemento a elemento, sendo uma delas a frequência observada y e outra a frequência esperada ou de comparação, y^*.

Quanto se tem somente a estego-imagem, é preciso computar a distribuição esperada, analisada a partir da mesma (Provos, 2001). Uma solução é computar a distribuição esperada, y_i , onde i denota o coeficiente analisado a partir da estego-imagem. Para isso, considere n_i a frequência para o coeficiente DCT i na imagem. Uma imagem com mensagem escondida tem frequência similar para coeficientes DCTs adjacentes. Como resultado, pode-se tomar a média aritmética

$$y_i^* = \frac{(n_{2i} + n_{2i+1})}{2} \tag{3.1}$$

para determinar a distribuição esperada. Esta é então comparada com a distribuição observada

$$y_i = n_{2i} \tag{3.2}$$

O valor χ^2 para a diferença entre as duas distribuições é dado como:

$$\chi^2 = \sum_{i=1}^{v+1} \frac{(y_i - {y^*}_i)}{{y^*}_i} \tag{3.3}$$

onde v é o grau de liberdade, isto é, o número de diferentes categorias (coeficientes) no histograma.

Quando o valor observado está muito longe do valor esperado, o resultado dessa diferença será grande, o que pode indicar que a frequência observada não está bem descrita pela frequência esperada (Rocha, 2006).

A probabilidade p de existir uma mensagem escondida é dada pelo complemento da função de distribuição acumulativa,

$$\int_0^{\chi^2} \frac{(t^{(v-2)/2} e^{-t/2})}{2^{(v/2)}\Gamma(v/2)}\, dt \tag{3.4}$$

onde Γ é a função Euler-Gama desenvolvida para estender a noção de fatorial para números não-naturais.

Pode-se calcular a probabilidade de incorporação de uma mensagem em diferentes partes de uma imagem. A seleção depende do tipo de abordagem de esteganografia que se está tentando detectar.

Para uma imagem que não contém qualquer informação escondida, se espera que a probabilidade de mascaramento seja igual a zero em todos os lugares (Provos, 2001). A Figura 3.3 mostra a probabilidade de uma imagem que não possui nenhuma mensagem escondida e a probabilidade de outra imagem que possui conteúdo escondido.

Diferentes processos de esteganografia causam diferentes distorções na imagem (Provos, 2001). Com isso, teste do χ^2 é aplicado em diferentes áreas da imagem dependendo do sistema que se quer identificar.

A desvantagem dessa abordagem é que, se não se sabe qual sistema de estega-

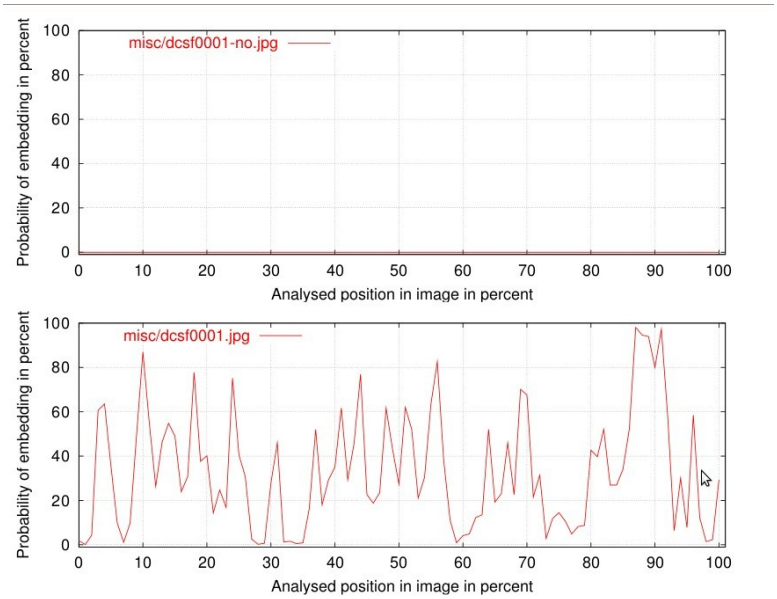

Figura 3.3: Comparação dos histogramas de frequência (Provos, 2001)

nografia foi utilizado, não tem como saber em qual local da imagem deverá ser aplicado o teste. E ainda, mensagens pequenas podem passar despercebidas. Existe também a possibilidade de inserir a mensagem tomando cuidado de preservar as propriedades do histograma de frequência de cores, como no aplicativo *OutGuess*. Além disso, Fridrich (2000) afirma que o Teste do χ^2 não é eficaz para imagens em *high-colors* (cada pixel da imagem possui dois *bytes*).

3.3 Análise de cores únicas no cubo RGB

Técnicas como a análise de cores únicas do cubo RGB (Fridrich, 2000) são exclusivas para imagens em cores reais (*true colors*). Isso porque elas estimam a probabilidade de existir uma mensagem escondida no LSB analisando o aumento dos pares de cores reais na paleta de cores. ale Nota-se que o número de cores únicas de uma imagem costuma ser menor que o seu número de pixels. A razão entre o número de cores únicas e o número de pixels varia de $1:2$ para imagens de alta qualidade a até $1:6$ para imagens JPEG de menor resolução. Assim a paleta de cores únicas é relativamente menor (Fridrich, 2000).

Com a inserção de uma mensagem no plano LSB o aspecto da paleta de cores únicas muda, formando vários pares de cores próximas (*close colors*).

Está técnica testa a presença de mensagens escondidas em imagens de cores reais

usando a seguinte ideia:

Seja U o número de cores únicas em uma imagem. Considerando somente as cores únicas tem-se P como o número de pares de cores próximas na paleta da imagem. Duas cores $(R1, G1, B1)$ e $(R2, G2, B2)$ são próximas se

$$(R1 - R2)^2 + (G1 - G2)^2 + (B1 - B2)^2 \leq 3. \tag{3.5}$$

O número de todos os pares de cores é dado pelo binômio

$$\binom{U}{2} \geq P \tag{3.6}$$

A razão R entre o número de pares de cores próximas e o número de todos os pares de cores dada pela equação 3.7

$$R = \frac{P}{\binom{U}{2}} \tag{3.7}$$

dá uma ideia do número relativo de cores próximas na imagem.

Depois que a mensagem é escondida, o número de cores próximas irá aumentar para U' e então se pode avaliar o número de pares de cores próximas P' e o número de todos os pares de cores. Para uma imagem que não contem uma mensagem escondida, a relação da quantidade de pares de cores próximas com o número de todos os pares de cores será menor do que em uma estego-imagem. No entanto, é difícil estabelecer uma razão R para todas as imagens devido a variação de cores únicas. Mas foi observado que se uma imagem já possui uma mensagem escondida, a inserção de uma nova mensagem não altera significativamente a razão R.

R é usado então como critério de decisão para classificar se a imagem contém ou não uma mensagem escondida. O processo de decisão, em uma imagem de tamanho $M \times N$, consiste nesses passos a seguir (Fridrich, 2000),

1. calcular a razão R entre o número de todos os pares possíveis de cores próximas P e o número de todas cores possíveis U, dada por:

$$R = \frac{P}{\binom{U}{2}} \tag{3.8}$$

2. inserir uma mensagem teste de tamanho $3\alpha MN$ pseudo-aleatoriamente no plano LSB, em I, onde α será discutido posteriormente.

3. calcular as quantidades correspondentes para a imagem I gerada no passo anterior como U e P e calcule a razão R para I com a mensagem de teste, dada por:

$$R' = \frac{P'}{\binom{U'}{2}} \tag{3.9}$$

Assim, se a imagem já possuía uma mensagem embutida, as razões terão valores muito próximos, $R \cong R'$. Caso contrário, a razão R' (após a inserção da mensagem teste) será significativamente maior que R, $(R' > R)$. E para facilitar o cálculo da separabilidade, pode-se definir a razão R'/R.

Para o caso de imagens com mensagens escondidas muito pequenas, as razões ficarão muito próximas. Com isso, a escolha do tamanho da mensagem de teste deve ser feita através do α correto, de modo a minimizar o número de falsas detecções. O tamanho da mensagem de teste é relacionado ao método de capacidade de mascaramento do LSB, um bit por canal de cor (3 bits por pixels). Os cálculos feitos por Fridrich (2000) mostram que o valor ótimo da mensagem teste seria com o $\alpha = 5\%$.

Uma das limitações dessa técnica é que o número de cores únicas da imagem de cobertura deve ser menor que 30% do número de pixels. Portanto, quanto menor o número de cores únicas, maior a confiabilidade desta técnica. Valores maiores que isso levam a resultados pouco confiáveis. Assim, esta técnica não é ideal para imagens com um número muito elevado de cores únicas. Como por exemplo, as digitalizações de alta qualidade, em que o número cores únicas pode chegar a mais de 50% da quantidade de pixels.

Além disso, de acordo com Zhi et al (2003), a análise do cubo RGB não é eficaz para imagens originais e naturais em escala de cinza e coloridas, porque em uma imagem original, cada pixel é diretamente representado por sua cor ou escala de cinza, assim o número de cores não pode ser reduzido pelo formato da imagem.

3.4 Análise RS

Proposto por Fridrich et al (2001), a Análise RS consiste em uma análise da capacidade de se inserir informações no plano LSB de uma imagem. Apesar do plano LSB parecer ser independente dos demais planos, de alguma forma existem relações entre os mesmos que não são lineares e podem ser estimadas pela capacidade de se esconder uma mensagem. Para isso é feito uma análise através da simulação artificial de um novo processo de mascaramento em uma imagem que precisa ser classificada como tendo ou não uma mensagem escondida. Esta simulação consiste na criação de funções que simulam o processo de inserção de informação e também na divisão da imagem analisada em grupos.

Seja uma imagem de cobertura I com $M \times N$ pixels e os valores dos pixel variem de acordo com o conjunto P. Por exemplo, para uma escala em tons de cinza de 8 bits, tem-se P {0,...,255}. A Análise RS começa por dividir a imagem em grupos de n pixels adjacentes $(x_1, ..., x_n)$. Um exemplo seria um grupo com os quatro pixels consecutivos. O próximo passo é definir uma função de discriminação f que atribui um número real $f(x_1, ..., x_n) \in \mathbb{R}$ para cada pixel do grupo G=$(x1, ..., xn)$. A função de discriminação quantifica a suavidade ou "regularidade" dos pixels, do grupo G. Quanto mais ruído um grupo G tiver, maior será a função de discriminação. Um exemplo seria a variação do grupo de pixels usada como função de discriminação.

$$f(x_1, x_2, ..., x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i| \tag{3.10}$$

Pode-se usar também estimativas estatísticas da imagem de cobertura para construir uma função de discriminação.

Assim, é definida uma função F inversível em P chamada *"flipping"*. *Flipping* é uma permutação em escala de cinza que consiste em dois ciclos. Então, F terá a função identidade F^2, ou seja, $F(F(x)) = x$ para todo $x \in P$. A permutação $F_1 : 0 \leftrightarrow 1, 2 \leftrightarrow 3, ..., 254 \leftrightarrow 255$ corresponde a inverter o LSB de cada nível de cor. Pode-se definir também uma função *shifting* $F_{-1} : -1 \leftrightarrow 0, 1 \leftrightarrow 2, ..., 253 \leftrightarrow 254, 255 \leftrightarrow 256$ ou

$$F_{-1}(x) = F_1(x + 1). \tag{3.11}$$

Para completar, define-se também uma função F_0 como identidade da permutação $F(x) = x$ para todo $x \in P$.

Uma máscara M pode ser usada para aplicar diferentes funções em diferentes pixels, denotando quais pixels deverão sofrer alteração, onde, a máscara contém n posições com os valores {0,1,-1}, sendo que, o valor -1 corresponde à aplicação da função F_1 , 1 corresponde a aplicação da função F_{-1} e 0 corresponde à aplicação da função F_0.

A função de discriminação e as funções F_0, F_1 e F_{-1} são usadas para definir três tipos de grupo de pixels, R_M , S_M e U_M:

- *Grupos Regulares* $G \in R_M \leftrightarrow f(F(G)) > f(G)$
- *Grupos Singulares* $G \in R_M \leftrightarrow f(F(G)) < f(G)$
- *Grupos Não-usáveis* $G \in R_M \leftrightarrow f(F(G)) = f(G)$

Sendo $F(G)$, o grupo G deslocado , com $(F_{M_{(1)}}(x_1), F_{M_{(2)}}(x_2), ..., F_{M_{(n)}}(x_n))$. Da mesma maneira , os grupos R_{-M} , S_{-M} e U_{-M} podem ser classificados sob a máscara $-M$.

O propósito da função flipping F é perturbar os valores do pixels, invertendo e assim simulando um pequeno ruído levando a um aumento da função de discriminação. Assim, o número de grupos regulares será maior do que o número de grupos singulares, permitindo uma adição de uma considerável quantidade de informação sem perdas, ou seja, o objetivo é simular uma mascaramento de mensagem (Fridrich et al, 2001).

Com isso, tem-se $R_M + S_M \leq 1$ e $R_{-M} + S_{-M} \leq 1$, para máscara negativa. A hipótese estatística desse método é que em uma imagem natural, o valor de R_M é igual a R_{-M} e o mesmo vale para S_M é igual a S_{-M}.

$$R_M \cong R_{-M} \text{ e } S_M \cong S_{-M} \tag{3.12}$$

Esta hipótese é justificada em Fridrich et al (2001) heuristicamente inspecionado a equação 3.11. a função de *flipping* F_1 é o mesmo que aplicar F_{-1} em uma imagem cujas cores foram deslocadas em um bit.

Para uma imagem típica, não existe, a priori, qualquer razão pela qual o número de grupos R e S deva mudar significativamente ao se mudar as cores em um bit. Esta

hipótese foi verificada experimentalmente para imagens tiradas com uma câmera digital, tanto para os formatos com perdas como para os formatos sem perdas . Ela também se mantém bem para imagens processadas com operações comuns de processamento de imagem e para a maioria das imagens digitalizadas. A relação 3.12, no entanto, é violada após randomização do plano LSB, por exemplo, devido a esteganografia usando o plano LSB.

A randomização do plano LSB força a diferença entre R_M e S_M para o valor zero a medida que o tamanho da mensagem embutido aumenta. Depois de alterado o LSB de 50% dos pixels, sendo a consequência do mascaramento da mensagem embutida, obtém-se $R_M \cong S_M$. Isto é o mesmo que dizer que a capacidade de mascaramento no plano LSB agora é zero. Surpreendentemente é o efeito contrário que acontece com R_{-M} e S_{-M}, sua diferença aumenta proporcionalmente ao tamanho da mensagem escondida.

A Figura 3.4 mostra o diagrama RS de uma imagem onde o eixo x é a percentagem de pixels cujos LSBs foram invertidos. O eixo y é o número relativo de grupos regulares e singulares sob as máscaras M = [0 1 1 0] e -M = [0 -1 -1 0].

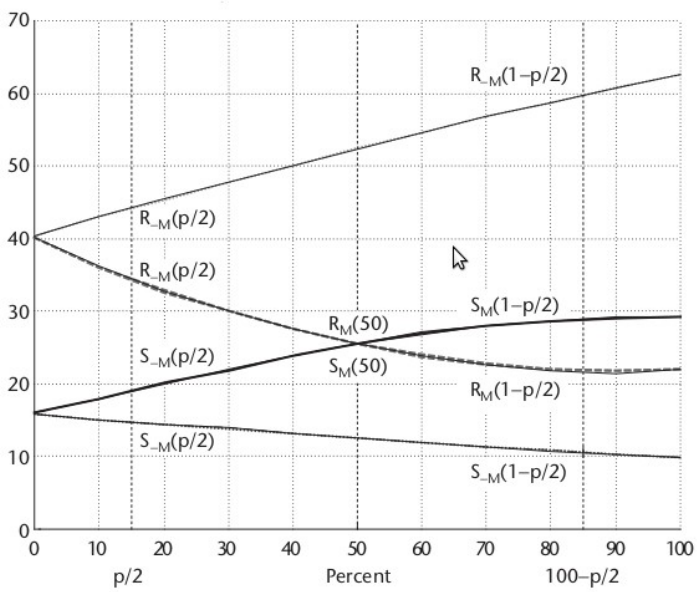

Figura 3.4: Diagrama RS de uma imagem (Fridrich et al, 2001).

O princípio da Análise RS é estimar quatro curvas do diagrama RS e calcular sua interseção usando extrapolação. A forma geral das quatro curvas varia de acordo com a imagem de cobertura a partir de curvas quase perfeitamente lineares. Os parâmetros das

curvas podem ser determinados a partir dos pontos marcados na Figura 3.4.

Em uma estego-imagem com uma mensagem de tamanho desconhecido p, incorporada aos LSBs aleatoriamente em pixels espalhados, as medidas inicias do número de grupos em R e S corresponde aos pontos $R_M(p/2)$, $S_M(p/2)$, $R_{-M}(p/2)$ e $S_{-M}(p/2)$. O fator é metade pois assume-se que a mensagem é uma sequência aleatória de um fluxo de bits. Em média, apenas metade será invertida. Se alterar o LSBs de todos os pixels da imagem e calcular o número dos grupos R e S, obtém-se os quatro pontos $R_M(1-p/2)$, $S_M(1-p/2)$, $R_{-M}(1-p/2)$ e $S_{-M}(1-p/2)$. Randomizando o plano LSB de uma estego-imagem obtém-se os pontos médios $R_M(1/2)$ e $S_M(1/2)$.

É possível encaixar linhas retas através dos pontos $R_{-M}(p/2)$, $R_{-M}(1-p/2)$ e $S_{-M}(p/2)$, $S{-}M(1{-}p/2)$. Os pontos $R_M(p/2)$, $R_M(1/2)$,$R_M(1{-}p/2)$ e $S_M(p/2)$ $S_M(1/2)$, $S_M(1-p/2)$ determinam duas parábolas.

Para evitar consumo de tempo da estimativa estatística dos pontos médios $R_M(1/2)$ e $S_M(1/2)$ e fazer a estimativa do comprimento da mensagem, aceitam-se duas condições adicionais:

- *1* O ponto de interseção das curvas R_M e R_{-M} tem a mesma coordenada x assim como o ponto de interseção das curvas S_M e S_{-M}
- *2* As curvas R_M e S_M se interceptam em $m = 50\%$, ou $R_{M(1/2)} = S_M(1/p)$.

Essas duas suposições tornam possível derivar uma fórmula para o tamanho da mensagem secreta p. Depois de reescalar o eixo x, de modo que $p/2$ se torna 0 e $100-p/2$ torna-se 1, a coordenada x do ponto de interseção é uma raiz da seguinte equação:

$2(d_1 + d_0)x^2 + (d_{-0} - d_{-1} - d_1 - 3d_0)x + d_0 - d_{-0} = 0$ onde,

$d_0 = R_M(p/2) - S_M(p/2)$

$d_1 = R_M(1-p/2) - S_M(1-p/2)$

$d_{-0} = R_{-M}(p/2) - S_{-M}(p/2)$

$d_{-1} = R_{-M}(1-p/2) - S_{-M}(1-p/2)$

O tamanho da mensagem é calculado a partir da raiz x que tiver o menor valor absoluto, $p = x/(x-1/2)$

Para imagens de cobertura que contêm muito ruído, a diferença entre o número

de pixels regulares e singulares já é pequena. Neste caso, a precisão da análise RS diminui. Esta técnica também tem resultados menos precisos para estego-imagens em que a mensagem é concentrada em áreas localizadas da imagem.

O método SLSB (Roque, 2009) pode enganar a analise RS. Isso acontece porque o nível de distorção causado no plano LSB por este método é pequeno o suficiente para passar despercebido pela análise.

3.5 Taxa de inversão da energia do gradiente.

Técnicas de esteganografia que usam o domínio espacial da imagem, como no plano LSB, alteram suavemente características entre os pixels adjacentes de imagem. Em Zhi et al (2003), os autores usam a análise da relação entre o tamanho de mensagem escondida e a energia gradiente como um método de esteganálise.

O gradiente de uma função $f(x, y)$ de duas variáveis x, y é definido por:

$$\nabla f = \begin{vmatrix} G_x \\ G_y \end{vmatrix} = \begin{vmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{vmatrix} \tag{3.13}$$

a magnitude é definida por:

$$|\nabla f| = [{G_x}^2 + {G_y}^2] \tag{3.14}$$

e a direção é dada por:

$$\alpha(x, y) = arctang \frac{G_x}{G_y} \tag{3.15}$$

O gradiente de uma imagem representa uma mudança direcional na intensidade ou cor. Para entender o processo, considere o sinal unidimensional $I(n)$. O gradiente $r(n)$, antes de se esconder a mensagem, é dado por:

$$r(n) = I(n) - I(n-1) \tag{3.16}$$

A energia do gradiente, EG, de $I(n)$ é definido como:

$$EG = \sum |I(n) - I(n-1)|^2 = \sum r(n)^2 \tag{3.17}$$

Depois da inserção da mensagem, no sinal original, $I(n)$ torna-se $I'(n)$ e o gradiente é reescrito como,

$$\begin{aligned} r(n) &= I(n) - I(n-1) \\ &= (I(n) + S(n)) - (I(n-1) + S(n-1)) \\ &= r(n) + S(n) - S(n-1), \end{aligned} \tag{3.18}$$

em que $S(n) = I'(n) - I(n)$, representa a quantidade alterada de $I(n)$ e está estreitamente relacionada com os dados da mensagem escondida e do próprio sinal.

Além disso, a função de distribuição de probabilidade de $S(n)$, $(\rho_{S(n)})$, é dada por:

$$\begin{cases} \rho_{S(n)}0 = 1/2 \\ \\ \rho_{S(n)} \pm 1 = 1/4 \end{cases} \tag{3.19}$$

Após a mensagem ser embutida, a nova energia do gradiente, EG.

$$\begin{aligned} EG' = \sum |r(n)|^2 &= \sum |r(n) + S(n) - S(n-1)|^2 \\ &= \sum |r(n) + \Delta(n)|^2, \end{aligned} \tag{3.20}$$

onde $\Delta(n) = S(n) - S(n-1)$.

De acordo com Zhi et al (2003), a variação da energia do gradiente causada pela inserção da mensagem no LSB, depende somente do tamanho da mensagem, e a variação da energia do gradiente é igual ao tamanho L da mensagem.

$$E(EG' - EG) = L \tag{3.21}$$

Para proceder o processo de detecção é necessário, antes, definir o processo de inversão dos bits do plano LSB da imagem. Para tal, define-se uma operação inversível F sobre um conjunto P de bits invertidos.

Caso a imagem de cobertura tenha $M \times N$ pixels e o tamanho da mensagem escondida seja $p < M \times N$, a operação F resulta em três propriedades:

1. Para $p = M \times N$, há $R = \dfrac{(M \times N)}{2}$ pixels com LSB invertido. Isto implica que a razão de mascaramento é de 50% e a energia do gradiente é dada por $EG = \left(\dfrac{M \times N}{2}\right)$.

2. A energia do gradiente da imagem original é dada por $EG(0)$. Ao fazer o inversão de todos os LSBs a partir da operação F, a energia do gradiente é $EG = (M \times N)$.

3. Para $p < M \times N$, há $p/2$ pixels com seu LSB invertido. Denota-se a imagem modificada como $I(p/2)$. A energia do gradiente correspondente é dada por $EG\left(\dfrac{(p/2)}{M \times N}\right) = EG(0) + p$. Caso a operação F seja aplicada sobre a imagem $I(p/2)$, a energia do gradiente resultante é $EG = \dfrac{(M \times N - p/2)}{M \times N}$.

A partir do algoritmo de inversão proposto, o processo de detecção é como segue:

1. calcular a energia do gradiente da imagem de teste $EG\left(\dfrac{p/2}{M \times N}\right)$;

2. aplicar a operação F sobre a imagem de teste e calcular $EG\left(\dfrac{M \times N - p/2}{M \times N}\right)$;

3. calcular $EG(M \times N/2) = \dfrac{\left[EG(p/2/M \times N) + EG\left(\dfrac{M \times N - p/2}{M \times N}\right)\right]}{2}$;

4. $EG(0)$ é dado, baseado em $EG\dfrac{(M \times N)}{2} = EG(0) + M \times N$;

5. o tamanho estimado para a mensagem escondida é dado pela expressão $p' = EG\left(\dfrac{p/2}{M \times N}\right) - EG(0)$.

Esta técnica pode falhar em casos onde os LSBs da imagem original forem uniformemente e aleatoriamente distribuídos, ou caso a mensagem secreta tenha esse mesmo tipo de distribuição, como observado nos testes realizados por Zhi et al (2003). Imagens com coeficientes de frequência muito altos e mensagens pequenas também dificultam a detecção.

3.6 Esteganálise Cega

A maior desvantagem das abordagens de esteganálise para técnicas de esteganografia específicas é que na maioria das vezes não se sabe qual técnica de esteganografia foi utilizada. Este é justamente o trunfo da esteganálise cega, já que, não é necessário saber qual método de esteganografia foi usado.

Na verdade, a esteganálise cega é um classificador que define se uma imagem é ou não uma estego-imagem. A esteganálise cega é considerada uma meta-detecção, visto que ela pode ser ajustada depois do treinamento com imagens de cobertura e estego-imagens, independente de onde a mensagem foi inserida (Luo, 2008).

Existem ainda duas formas de se trabalhar com esteganálise cega. A primeira classifica as imagens como estego-imagem ou não usando um conjunto de treinamento com imagens com e sem mensagem escondida e características sensíveis aos distúrbios provocados pela mensagem escondida, sem possuir de fato informações das mudanças geradas na estego-imagem. Já na segunda forma, as estego-imagens do grupo de treinamento são obtidas usando diversos métodos de esteganografia conhecidos. Ainda assim, é possível detectar técnicas de esteganografia novas ou desconhecidas.

O processo de esteganálise cega pode ser dividido em quatro fases. A primeira fase é o processamento das imagens. Nesta fase são realizadas algumas operações na imagem como conversão do sistema de cores, recorte, aplicação de transformadas como a DCT e DWT. Estas operações são necessárias para segunda fase.

A segunda fase é a seleção das características. São escolhidas as características que são sensíveis à inserção e modificação. Tais características formam um vetor que será usado nas fases de treinamento e classificação. Na terceira fase escolhe-se um classificador que será projetado baseado nas características selecionadas. Um conjunto com uma grande quantidade de imagens é usado para treinar o classificador e obter alguns parâmetros necessários para classificação. A quarta e última fase é a classificação que divide as imagens em duas classes: estego-imagem e imagem original. Um esquema deste processo é apresentado na Figura 3.6.

Quanto a seleção das características, alguns exemplos de abordagens vistos em Luo (2009) são:

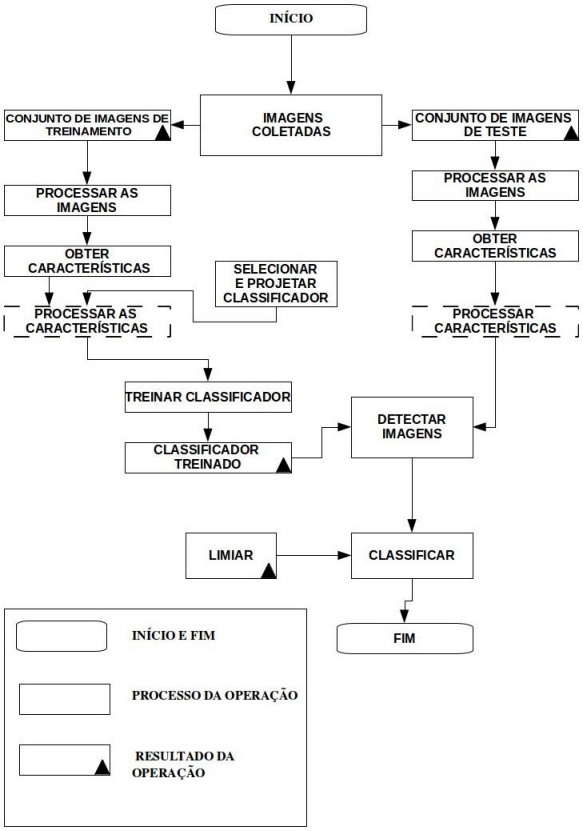

Figura 3.5: Esquema do processo de esteganálise cega adaptado de (Luo, 2008)

- O COF (*Center of Mass*) do histograma da imagem: O COF dá uma noção da distribuição de energia no histograma da imagem. Seu valor diminui quando uma mensagem é inserida;
- Distribuição dos coeficientes DWT: a inserção da mensagem torna a distribuição dos coeficientes DWT mais suave;
- Matriz Co-M(*Co-occurrence Matrix*): Co-M é uma representação das características estatísticas da imagem. Suas propriedades estatísticas da também são alteradas após a inserção da mensagem.

Duas abordagens desenvolvidas, uma que usa um modelo estatístico de alta ordem e outra que usa métricas de qualidade da imagem, são descritas nas seções 3.6.1 e 3.6.2, respectivamente.

A etapa de classificação implementa meios para computar relações entre conjuntos de imagens. O aprendizado de máquina visa melhorar o desempenho de algoritmos automaticamente através da experiência. Os métodos de aprendizagem de máquina têm

sido utilizados em diversas aplicações, como reconhecimento da fala (Ganapathiraju et al, 2004), detecção de fraudes em cartões de crédito (Maes et al, 1993), estratégias para a construção de jogos (Furnkranz, 2007), programas de mineração de dados que descobrem regras gerais em grandes bases de dados como, por exemplo, o sistema Weka (Hall et al, 2009).

No processo de aprendizado, várias imagens com e sem mensagens escondidas são apresentadas ao classificador de modo a conseguir os coeficientes mais exatos possíveis para realizar a classificação.

Diferentes tipos de classificadores podem ser usados, como o classificador baseado na análise de regressão multivariada (Avcibas et al, 2003), máquina de aprendizado (SVM) (Zuo et al, 2006) , redes neurais artificiais (RNA) (Wing et al, 2011), sistema imunológico computacional (CIS)(Jackson et al, 2003) e hiperestrutura geométrica (Mcbride et al, 2005).

Dentre as abordagens principais, pode-se citar a proposta por Zuo et al (2006), que usa uma cadeia de Markov 2D com um limiar de previsão de erro da imagem. Este método tem como base o fato de que um pixel da imagem pode ser previsto por sua vizinhança. Com isso, a partir da diferença do valor do pixel e do valor previsto pela vizinhança é construído um limiar para classificar as imagens . Esta abordagem usa máquina de aprendizado *Support Vector Machines (SVM)* como classificador.

Já em Shuang et al (2012), os autores usam um modelo estatístico de decomposição *wavelet*, em que as estatísticas da decomposição *wavelet* são testadas usando a análise de variância ANOVA[14] para verificar quais são mais sensíveis à mensagem inserida. Este método também usa máquina de aprendizado (SVM) como classificador.

A esteganálise cega é a abordagem mais desenvolvida atualmente. As razões para isso provavelmente são suas duas maiores vantagens: fato dela ser "cega" e adaptável Ainda assim, não é uma abordagem perfeita, visto que a esteganálise voltada para um método específico de esteganografia possui mais precisão quanto a estimativa da existência da mensagem, além de ser muito mais complexa e requerer mais recursos.

Também já existem algoritmos de esteganografia que tentam driblar essa abor-

[14]coleção de modelos estatísticos no qual a variância amostral é particionada em diversos componentes, que estão associados a um processo, produto ou serviço

dagem como o YASS (Solanki, 2007) que tenta preservar as características estatísticas da imagem ao se inserir a mensagem em regiões aleatórias da imagem.

Por outro lado já existe um método de esteganálise para o YASS. Em Li et al (2009) é feita uma análise da imagem em determinadas regiões partindo do fato de que, apesar de eficiente, o YASS não é totalmente aleatório.

3.6.1 Métricas de qualidade de imagens

Avcibas et al (2003) desenvolveram o que é considerado o primeiro sistema de esteganálise cega usando métricas de qualidade de imagem e regressão multivariada (Natarajan, 2012).

Métricas de qualidade são usadas, geralmente, na avaliação de codificação de ruídos, predição de performance de algoritmos de Visão Computacional, perda de qualidade devido a inadequabilidade de algum sensor, entre outras aplicações. Essas mesmas métricas podem ser utilizadas para construir um discriminador de imagens de cobertura de estego-imagens através da utilização de regressão multivariada.

Uma boa métrica de qualidade deve ser precisa, consistente e monótona para predizer a qualidade. Em relação a esteganálise, precisão pode ser interpretado como a possibilidade da métrica detectar a presença da mensagem oculta com o erro mínimo. Da mesma forma, monotonicidade significa que as métricas devem idealmente ser monótonas em seu relação a mensagem embutida. Finalmente, consistência diz respeito à capacidade de a medida de qualidade fornecer previsões consistentes para um grande conjunto de de técnicas de esteganografia e tipos de imagem (Avcibas et al, 2003).

Quanto a seleção das métricas de qualidade, são escolhidas várias medidas diferentes que respondem com diferentes sensibilidades a artefatos e distorções. As principais métricas de qualidade utilizadas são apresentadas(Avcibas et al, 2002):

- Métricas Baseadas na Diferença dos Pixels: esse tipo de medida calcula a distorção entre duas imagens com base nas suas diferenças pixel a pixel (*pixelwise*) ou, em certos momentos, de diferença da imagem (erro), como por exemplo, a média angular e o erro médio absoluto;

- Métricas de Distância Espectral e Distância de Fase do Bloco Espectral: neste tipo

método são considerados as distorções obtidas no espectro de Fourier das imagens e são relacionados a *blurring*. As mais utilizadas para esteganálise são a distância de fase de bloco espectral e distância espectral ponderada da mediana de bloco;

- Erro Médio Quadrático Normalizado do Sistema Visual Humano (SVH): a incorporação do modelo SVH em medidas objetivas leva a uma melhor correlação com avaliações subjetivas e também está relacionado à distorção com borrão;

- Medidas de Correlação: as funções de correlação são usadas para quantificar a similaridade entre duas imagens. Um exemplo seria a distância de Czenakowski (Avcibas et al, 2002) que é uma métrica útil para a comparação de vetores de elementos estritamente não-negativos.

Considere a imagem de cobertura como um sinal f e a mensagem a ser inserida como um sinal w, a estego-imagem seria então $g = f + w$. Sendo assim, a etapa de treinamento do classificador consiste em usar um conjunto de imagens de cobertura e estego-imagens de modo a conseguir os coeficientes de qualidade de imagem que sejam capazes de separar as duas classes de imagens.

Assim, os coeficientes obtidos na etapa de treinamento podem ser utilizados na etapa de teste. Dada uma imagem na etapa de teste, primeiro é obtido uma versão filtrada desta para realizar uma estimativa da imagem original de cobertura. Então, utilizando os coeficientes de predição é feita a sua regressão até um valor de saída ser obtido. Caso o valor de saída supere o limiar, então a decisão sobre a hipótese estatística é que a imagem possui uma mensagem escondida.

Avcibas et al (2003) também acreditam que é possível encontrar conjuntos de métricas de qualidade que sofram alteração em uma técnica de esteganografia específica. Os autores realizaram testes fazendo uma classificação extra, que identifica qual método de esteganografia foi usado. Neste caso, os autores usaram três conjuntos de imagens onde um conjunto foi submetido ao aplicativo *Digimarc*, outro conjunto para o JK-PGS (estes aplicativos são citados na Seção 2.11) e o terceiro conjunto para a técnica de esteganografia (Cox et al, 1997) que é baseada no método SSIS (visto na Seção 2.6).

3.6.2 Análise estatística de alta ordem

Proposta por Farid (2001), esta abordagem constrói modelos estatísticos de alta ordem para imagens naturais que procuram por desvios nos padrões, para detectar diferenças na estego-imagem e na imagem de cobertura.

Primeiramente é feita a decomposição baseada em filtros de quadratura em espelho *Quadrature Mirrors FIlters (QMF)*. Esta decomposição divide o domínio da frequência em múltiplas escalas e orientações. Para isso, aplica-se filtros passa-baixa e passa-alta[15] ao longo dos eixos da imagem, gerando subbandas vertical, horizontal, diagonal e passa-baixa. Sendo que, as escalas posteriores podem ser geradas recursivamente.

As subbandas vertical, horizontal e diagonal na i-ésima escala são representadas como, $V_i(x, y), H_i(x, y)$ e $D_i(x, y)$, respectivamente. A Figura 3.6 mostra uma decomposição no terceiro nível. A partir da decomposição da imagem, são calculadas para o modelo a média, variância, assimetria e curtose dos coeficientes em cada orientação e escala $i = 1, ..., n - 1$.

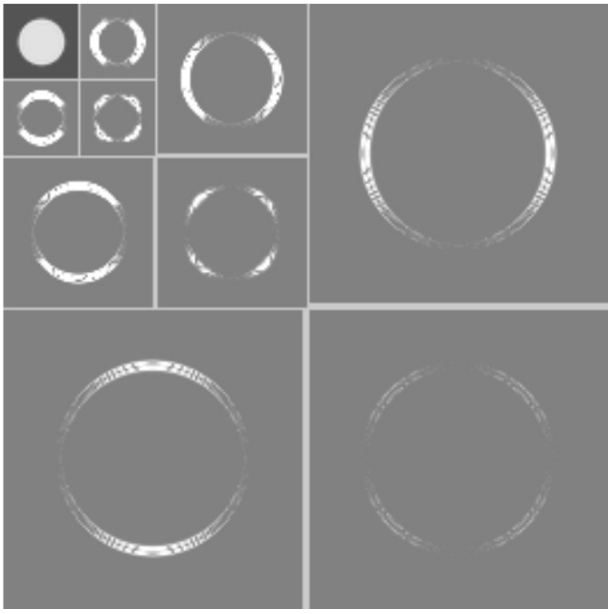

Figura 3.6: Decomposição da imagem "Disc"(Farid, 2001)

Depois, um segundo conjunto de estatísticas é calculado. Este modelo é baseado nos erros de um preditor linear ótimo dos coeficientes de magnitude. Assim, uma correlação entre a vizinhança espacial dos coeficientes das subbandas é realizada. A cor-

[15]filtros passa-alta atenuam ou eliminam os componentes de baixa frequência da imagem

relação para subbanda vertical, $V_i(x, y)$, com todos seus vizinhos é dada por :

$$\begin{aligned} V_i(x,y) =& w_1 V_i(x-1,y) + w_2 V_i(x+1,y) \\ &+ w_3 + V_i(x,y-1) + w_4 V_i(x,y+1) \\ &+ w_5 V_{i+1}(x/2, y/2) + w_6 D_i(x,y) \\ &+ w_7 V_i(x/2, y/2) \end{aligned} \tag{3.22}$$

em que w_k $(1 \leq k \leq 7)$ é o valor escalar dos coeficientes. Uma representação mais simples desta expressão seria $\vec{V} = Q\vec{w}$, onde o vetor coluna $\vec{w} = (w_1, ..., w_7)^T$, o vetor $\vec{V}$ contém os coeficientes de magnitude da subbanda vertical colocados em um vetor coluna e as colunas da matriz Q contém a vizinhança dos coeficientes de magnitude conforme a equação 3.22. Os coeficientes são determinados através da minimização do erro quadrático, dada pela equação:

$$\vec{E}(\vec{w}) = [\vec{V} - Q\vec{w}]^2 \tag{3.23}$$

Esta função de erro é minimizada através da diferenciação com respeito a $\vec{w}$:

$$\frac{dE(\vec{w})}{d(\vec{w})} = 2Q^T[\vec{V} - Q\vec{w}] \tag{3.24}$$

passando o resultado para zero, tem-se $\vec{w}$:

$$\vec{w} = (Q^T Q)^{-1} Q^T \vec{V} \tag{3.25}$$

Em seguida, o erro de log do preditor linear pode ser dado por:

$$\vec{E}_v = log_2(\vec{V}) - log_2(|Q\vec{w}|) \tag{3.26}$$

Com isso o erro adicional nos modelos estatísticos é coletado. O processo é repetido pra cada escala da subbanda vertical, como também é aplicado nas demais subbandas. Assim, encontra-se 12 estatísticas de erro, mais 12 calculadas no processo anterior. Essas 24 estatísticas são usadas para diferenciar as imagens.

4 Conclusão

Nota-se que a esteganografia em imagens digitais tem um vasto campo de aplicação. Os dados redundantes da imagem digital formam um campo fértil para quem quer esconder uma informação. São diversas as possibilidades, que se estendem a domínios diferentes, formatos diferentes. Atualmente, existe um bom número de métodos para esteganografia e vários aplicativos são disponibilizados na Internet.

Pelo estudo apresentado neste trabalho, pode-se constatar que a esteganografia não dispõe de mecanismos que permitam medir o quanto uma mensagem foi bem escondida, já que os métodos de esteganálise estão sempre se adaptando às técnicas usadas pela mesma. Mesmo que a mensagem não possa ser detectada visualmente não significa que não cause alterações na imagem. Estas alterações são as falhas dos métodos de esteganografia e a partir delas, é possível detectar a presença da mensagem.

A estegánalise surgiu como um meio de frustrar a esteganografia, detectando as mensagens escondidas a partir das alterações causadas por sua inserção na imagem.

Assim como a esteganografia, a esteganálise também pode falhar. Existem limitações, como o tamanho da mensagem que pode ser detectada, a geração de falsos positivos por ruídos naturais e a confiabilidade da estimativa resultante.

Além disso, a partir do momento em que os métodos de esteganálise superam os métodos de esteganografia, tais métodos procuram se adaptar buscando meios de remover suas falhas e tornar a mensagem cada vez mais invisível. Com isso, surgem novas técnicas de esteganálise para compensar as limitações das técnicas antigas. Portanto, ambos os lados se renovam constantemente.

Na Figura 4.1 é mostrado um esquema do desenvolvimento de técnicas de esteganografia e esteganálise ao longo do tempo usando algumas técnicas mencionadas neste trabalho.

Atualmente, o método de esteganálise que vem ganhando destaque é a esteganálise cega. Vários métodos já foram apresentados e uma lista deles pode ser vista em (Sharma et al, 2012). Mas apesar de apresentarem bons resultados, nenhum destes métodos é 100%

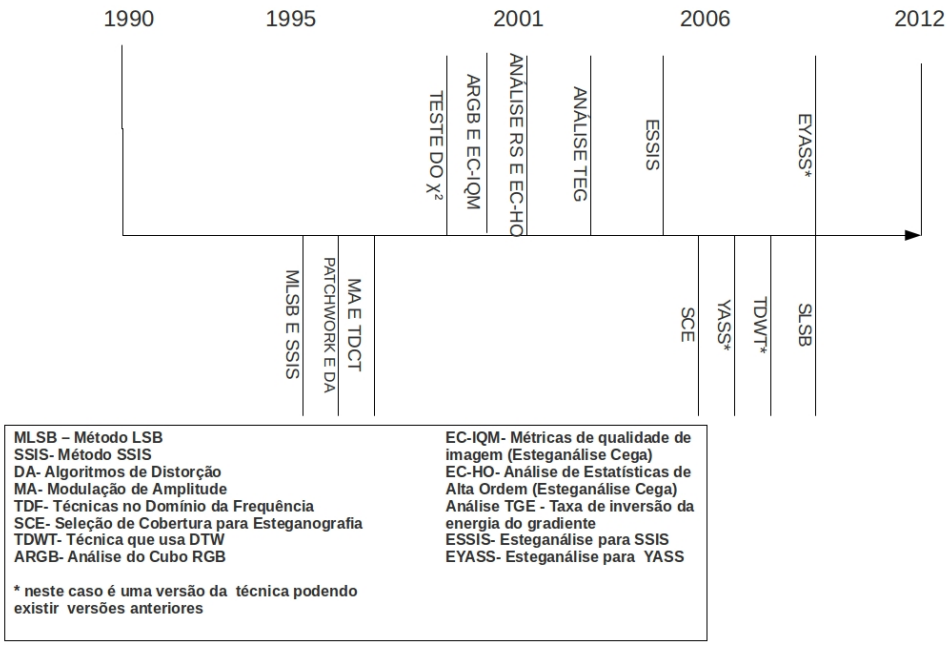

Figura 4.1: Linha do tempo das técnicas de esteganografia e esteganálise

eficaz.

Assim pode se concluir que não existe método de esteganografia, nem de esteganálise que seja perfeito. O que acontece na verdade é que existem métodos que são mais adequados que outros, dependendo dos aspectos como o objetivo da esteganografia, o formato da imagem, o tamanho da mensagem. E no caso da esteganálise, além das considerações anteriores, deve-se considerar o grau de precisão necessário, se a imagem original está disponível, ou se tem ou não conhecimento do método de esteganografia usado, se o objetivo vai além de detectar a imagem.

Neste sentido, Wayner (2002) afirma que:

> *O melhor que se pode fazer na esteganografia é mudar constantemente os parâmetros e os locais usados para esconder a mensagem. O melhor que se pode fazer na esteganálise é constantemente sondar padrões sutis deixados por engano.*

A Figura 4 reforça a afirmação feita por Wayner (2002) demonstrando o ciclo que gera a evolução nas duas áreas.

Figura 4.2: Relação entre esteganografia e esteganálise

Algumas sugestões para trabalhos futuros na área de esteganálise são, por exemplo, propor uma classificação mais geral das propriedades estatísticas entre as mais sensíveis e as menos sensíveis à esteganografia; quais técnicas de esteganografia causam uma maior variedade de distúrbios à imagem; quais formatos são mais difíceis de analisar ou que causam um maior número de falsos positivos etc.

No que se refere à esteganografia, uma sugestão de trabalhos futuros é uma análise do uso de Inteligência Computacional para otimização das técnicas esteganografia. Um exemplo deste tipo de abordagem já foi citado na Seção 2.3, visto em (Vieira et al, 2010). Outro exemplo é apresentado em (Silva et al, 2010), que usa heurísticas para que a mensagem inserida sobreviva à mudança de escala da imagem.

Outra sugestão é o estudo das técnicas de esteganografia e esteganálise para os demais objetos digitais que são arquivos de áudio, vídeo e texto.

Referências Bibliográficas

Abdelwahab, A.; Hasan, L. **A discrete wavelet transform based technique for image data hiding**. Proc. of 25th National Radio Science Conference, 2008.

Avcibas, I.; Sankur, B. ; Sayood, K. **Statistical evaluation of image quality measures**. In: Journal of Electronic Imaging, p. 206–223, 2002.

Avcibas, I.; Memon, N. ; Sankur, B. **Steganalysis using image quality metrics**. In: IEEE Transactions on Image Processing, 2003.

Back, T. **Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms**. p. 106–123. Oxford University Press, 1996.

Barni, M.; Bartolini, F.; Cappellini, V. ; Piva, A. **Robust watermarking of still images for copyright protection**. In: In Proc. 13th Inter. Conf. Digital Signal Processing, p. 499–502, 1997.

Cox, I. J.; Kilian, J.; Leighton, T. ; Shamoon, T. **Secure spread spectrum watermarking for multimedia**. In: IEEE Trans Image Processing, 1997.

Farid, H. **Detecting steganographic messages in digital images**. Dartmouth College, 2001.

Fridrich, J.; Du, R. ; Long, M. **Steganalysis of lsb encoding in color images**. In: 2000 IEEE International Conference, 2000.

Fridrich, J.; Goljan, M. ; Du, R. **Reliable detection of lsb steganography in color and grayscale images**. In: IEEE Multimedia, volume 8, p. 22–28, 2001.

Furnkranz, J. **Recent advances in machine learning and game playing**. In: ÖGAI Journal, 2007.

Ganapathiraju; A.Hamaker ; Picone, J. **Applications of support vector machines to speech recognition**. 2004.

Glover, F.; Laguna, M. **Tabu search**. In: Springer US, 1997.

Gonzalez, R.; Woods, R. **Processamento digital de imagens**. Editora Blucher, 2002.

Hall, M.; Eibe, F.; Geoffrey, H.; Bernhard, P.; Peter, R. ; H, W. I. **The weka data mining software: an update**. p. 10–18. ACM, 2009.

Hamid, N.; Yahya, A.; Ahmad, R. B. ; Al-Qershi, O. M. **Image steganography techniques: An overview**. In: International Journal of Computer Science and Security (IJCSS), 2012.

Jackson, J. T.; Gunsch, G. H.; Lamont, G. B. ; Jr., R. L. C. **Blind stenanography detection using a computational immune system: A work in progress**. In: IJDE, volume 1, 2003.

Johnson, N. F.; Katzenbeisser, S. **A survey of steganographic techniques**. In: Information Hiding Techniques for Steganography and Digital Watermarking, 2000.

Julio, E. P.; Brazil, W. G. ; Albuquerque, C. **Esteganografia: a arte das mensagens ocultas**. In: Fonte, 2007.

Julio, E. P.; Brazil, W. G. ; Albuquerque, C. **Esteganografia e suas aplicações**. In: VII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais, 2007a.

Kharrazi, M.; Taha, S. ; D., M. N. **Cover selection for steganographic embedding.** In: ICIP, p. 117–120. IEEE, 2006.

Kutter, M.; Jordan, F. ; Bossen, F. **Digital signature of color images using amplitude modulation**. In: Journal of Electronic Imaging, p. 326–332, 1997.

Li, B.; Huang, J. ; Fellow, Y. Q. S. **Steganalysis of yass**. In: IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, 2009.

Luo, X.-Y.; Dao-Shun, W.; Ping, W. ; Fen-Lin, L. **Review: A review on blind detection for image steganography**. volume 88, p. 2138–2157, Amsterdam, The Netherlands, The Netherlands, 2008. Elsevier North-Holland, Inc.

Luo, X.-Y.; Wang, D.-S.; Hu, W. ; Liu1, F. **Blind detection for image steganography: A system framework and implementation**. In: International Journal of Innovative Computing, Information and ControlVolume 5, Number 2, February 2009, 2009.

Maes, S.; Tuyls, K.; Vanschoenwinkel, B. ; Manderick, B. **Credit card fraud detection using bayesian and neural networks**. In: Maciunas RJ, editor. Interactive image-guided neurosurgery. American Association Neurological Surgeons, p. 261–270, 1993.

Marvel, L.; Boncelet, C. ; Retter, C. **Spread spectrum image steganography**. In: IEEE Transactions on Image Processing, volume 8, p. 1075–1083, 1999.

Mcbride; T., B.; Peterson; L., G.; Gustafson ; C., S. **A new blind method for detecting novel steganography**. 2005.

Morkel, T.; Eloff, J. H. P. ; Olivier, M. S. **An overview of image steganography**. In: Hein S Venter, Jan H P Eloff, L. L.; Eloff, M. M., editors, Proceedings of the Fifth Annual Information Security South Africa Conference (ISSA2005), Sandton, South Africa, June/July 2005. Published electronically.

Natarajan, V.; Anitha, R. **Blind image steganalysis based on contourlet transform**. In: International Journal on Cryptography and Information Security (IJCIS), 2012.

Pereira, E. D. V.; Weber, D. ; Goldani, C. **Deteção e acesso a mensagens ocultas em imagens digitais**. XX Congresso Nacional de Criminalística, 2009.

Petitcolas, F. A. P.; Anderson, R. J. ; Kuhn, M. G. **Information hiding-a survey**. IEEE, 1999.

Popa, R. **An analysis of steganographic techniques**. University of Timisoara, Timisoara, Romenia, 1998. Master's thesis, Department of Computer Science and Software Engineering of The Polytechnic.

Provos, N.; Honeyman, P. **Detecting steganographic content on the internet**. In: NDSS. The Internet Society, 2001.

Provos, N.; Honeyman, P. **Hide and seek: An introduction to steganography**. In: IEEE Security & Privacy Magazine, 2003.

Rocha, A. R.; Costa, H. A. X. ; Chaves, L. M. **Camaleão: Um software para segurança digital utilizando esteganografia.** Monografia (Ciência da Computação), Departamento de Ciência da Computação, Universidade Federal de Lavras, 2003.

Rocha, A. R. **Randomização progressiva para esteganálise**. UNICAMP, 2006.

Roque, J. J. **Slsb: Improving the steganographic algorithm lsb**. In: WOSIS 09, p. 57–66, 2009.

Sharma, D. M.; Bera, M. S. **A review on blind still image steganalysis techniques using features extraction and pattern classification method**. In: International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), 2012.

Shuang, H. Z.; Zhang, H.-B. **Spatial-frequency feature vector fusion based steganalysis**. In: IEEE, 2006.

Silva, T. R.; de Ávila, S. E. F. **Um algoritmo robusto ao aumento de escala em estego-imagens marcadas com a técnica lsb**. In: II ERIN, 2010, 2010.

Solanki, K.; Sarkar, A. ; Manjunath, B. S. **Yass: Yet another steganographic scheme that resists blind steganalysis**. In: Furon, T.; Cayre, F.; Doërr, G. ; Bas, P., editors, Computer, volume 4567, p. 1–15. Springer-Verlag, 2007.

Sullivan, K.; Madhow, U.; Ch, S. ; Manjunath, B. S. **Steganalysis of spread spectrum data hiding exploiting cover memory**. In: in Security, Steganography, and Watermarking of Multimedia Contents VII, p. 38–46, 2005.

Vieira, M. V. M.; Brazil, A. L.; Conci, A. ; Albuquerque, C. V. N. **Heurísticas para matriz de substituição do lsb utilizando algoritmos genéticos e path relinking**. In: Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais, SBSeg, 2010.

WANG, H.; WANG, S. **Cyber warfare:steganography vs. steganalysis**. In: COMMUNICATIONS OF THE ACM, 2004.

Wayner, P. **Disappearing cryptography: Information hiding: Steganography and watermarking (2nd edition)**. San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

Wing, N. W. Y.; Zhi-Min, H.; K., C. P. P. ; S., Y. D. **Blind steganalysis with high generalization capability for different image databases using l-gem**. In: ICMLC, p. 1690–1695, 2011.

Zhi, L.; Fen, S. A. ; Xian, Y. Y. **A lsb steganography detection algorithm**. In: The XIV IEEE 2003 International Symposium on Persona1,lndoor and Mobile Radio Communication Proceedings, 2003.

Zou, D.; Q., S. Y.; Wei, S. ; Guorong, X. **Steganalysis based on markov model of thresholded prediction-error image.** In: ICME, p. 1365–1368. IEEE, 2006.