UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE COMPUTAÇÃO

Proposta de Dissertação de Mestrado

ANÁLISE E IMPLEMENTAÇÃO DE TÉCNICAS DE DETECÇÃO DE MENSAGENS ESCONDIDAS EM IMAGENS DIGITAIS

Aluno: Anderson de Rezende Rocha[1]

Orientador: Prof. Dr. Siome Klein Goldenstein

Resumo

A *esteganografia*, a arte e a ciência das comunicações escondidas, inclui um vasto conjunto de métodos para comunicações secretas tais como tintas "invisíveis", micro-pontos, arranjo de caracteres (*character arrangement*), assinaturas digitais, canais escondidos (*covert channels*), entre outras. Por se tratar de um campo novo e pouco estudado, é muito importante conhecer seus meandros e suas reais potencialidades e perigos. Neste âmbito, o principal objetivo deste trabalho é estudar e desenvolver técnicas capazes de permitir a detecção de mensagens escondidas em imagens digitais. Este estudo trará à sociedade de forma geral uma ferramenta poderosa de combate a uma tecnologia de potencialidades ainda não tão bem demarcadas. A pesquisa é importante por colocar o país em movimento no sentido de estar contribuindo no combate ao terrorismo digital em todo o mundo.

1 Introdução

A busca por novos meios eficientes e eficazes de proteção digital é um campo de pesquisa fundamentado nos mais variados campos da ciência. Este campo de pesquisas se divide em duas ramificações. De um lado estão aqueles que buscam técnicas para se obter maior proteção digital. Do outro lado, estão aqueles que querem minar a proteção, isto é, querem ter acesso à informação.

Uma das áreas que tem recebido muita atenção recentemente é a *esteganografia*. Esta é a arte de mascarar informações como uma forma de evitar a sua detecção. *Esteganografia* deriva do grego, donde *estegano = esconder, mascarar* e *grafia = escrita*. Logo, *esteganografia* é a arte da escrita encoberta.

A *esteganografia* inclui um vasto conjunto de métodos para comunicações secretas desenvolvidos ao longo da história. Dentre tais métodos estão: tintas "invisíveis", micro-pontos, arranjo de caracteres (*character arranjement*), assinaturas digitais, canais escondidos (*covert channels*), comunicações por espalhamento de espectro (*spread spectrum communications*) entre outras.

O termo *esteganografia* ganhou muito destaque nas mídias informativas após o ataque terrorista ao *World Trade Center* em 11 de setembro de 2001. Muito se falou, após este lastimável acontecimento, que o grupo terrorista *Al Qaeda*, liderado pelo milionário saudita 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 [4].

[1]Projeto financiado pela *Fundação de Amparo à Pesquisa do Estado de São Paulo* (Fapesp). Processo número 04/02384-1.

Por fornecer às pessoas um alto grau de sigilo em suas operações, as modernas técnicas de *esteganografia* preocupam a sociedade como um todo e as autoridades políticas e policiais em específico. Uma vez que as comunicações estejam seguras, qualquer indivíduo de um país pode elaborar em segredo os mais mirabolantes planos que desejar. Estes planos podem muito bem tratar-se de um atentado terrorista e/ou ameaçar a segurança nacional de um país.

Assim, o estudo de técnicas e ferramentas que possam auxiliar na detecção de mensagens escondidas torna-se de grande valia para a sociedade. Neste contexto, aparece a *esteganálise digital* que é a arte de detectar mensagens escondidas nos mais diversos meios digitais. Além de fornecer técnicas robustas para detecção de mensagens escondidas, a *esteganálise* pode ser aplicada na criação de *watermarks* mais resistentes ao detectar e destruir marcações mais frágeis. A *esteganálise digital* pode também ser aplicada em *cyber-guerras*, computação forense, rastreamento de atividades criminais, coleção de evidências para investigações (particularmente em caso de terrorismo internacional) entre outras atividades [39].

Este trabalho tem por principais objetivos discutir e comparar as principais tecnologias de detecção de mensagens escondidas em imagens digitais. Além de levantar os pontos fortes e fracos e propor melhorias às técnicas existentes, busca-se desenvolver técnicas mais inteligentes. Espera-se desenvolver uma ferramenta de detecção eficiente e eficaz de caráter nacional que possa servir à sociedade como um todo.

2 Terminologia

Segundo o modelo geral de ocultamento de dados (*information hiding*), o *dado embutido* (*embedded data*) é a mensagem que se deseja enviar de maneira secreta. Freqüentemente, este dado é escondido em uma *mensagem inócua* (sem maior importância) conhecida como *mensagem de cobertura* (*cover-message*). As mensagens de cobertura podem variar de nome de acordo com o meio de cobertura sendo utilizado. Deste modo, pode-se definir uma *imagem de cobertura* (*cover-image*), *áudio de cobertura* (*cover-audio*) ou *texto de cobertura* (*cover-text*). Após o processo de inserção dos dados na mensagem de cobertura, obtém-se o chamado *estego-objeto* (*stego-object*), uma mensagem inócua contendo secretamente uma mensagem de maior importância [26].

Uma estego-chave (*stego-key*) é utilizada para controlar o processo de ocultamento de forma a restringir a detecção e/ou recuperação dos dados do material embutido.

A Figura 1 apresenta como o processo pode ser interpretado. Um indivíduo escolhe o dado a ser escondido e, a partir de uma chave, mascara estes dados em uma imagem de cobertura previamente selecionada. O resultado é a estego-imagem a ser enviada.

Um *ataque* com sucesso à *esteganografia* consiste em conseguir detectar a existência de uma mensagem escondida em algum meio observado. Por outro lado, os sistemas de marcação de *copyright* ou *watermarking* têm requisitos adicionais de robustez contra possíveis ataques. Deste modo, um *ataque* bem-sucedido consiste em conseguir detectar e remover a marcação digital [26].

Continuando, o sistema de *seriação digital* (*fingerprinting*), também conhecido como *labels*, consiste de uma série de números embutidos no material a ser protegido. Isto permite identificar, por exemplo, que um cliente quebrou um acordo de propriedade intelectual.

Finalmente, pode-se delimitar a grande-área de pesquisa conhecida como *ocultamento da informação* (*information hiding*) como apresentado hierarquicamente na Figura 2.

No segundo nível da hierarquia têm-se: *canais secretos*, *esteganografia*, *anonimato* e *marcação de copyright*.

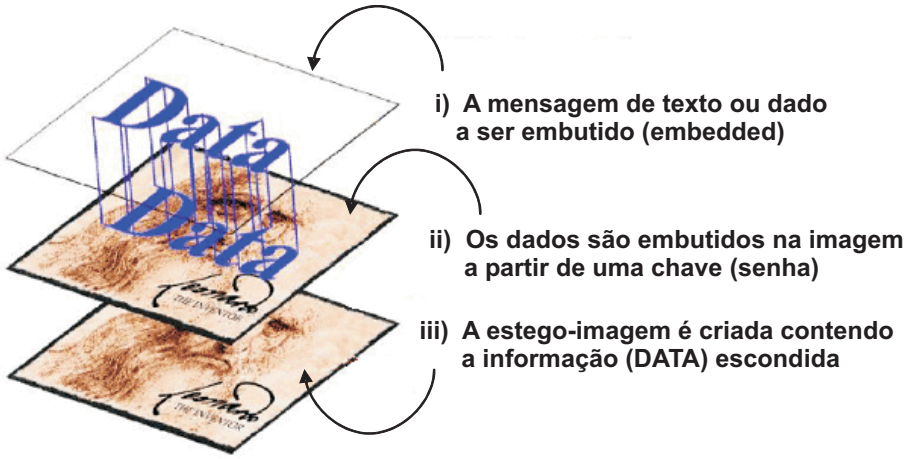

Figura 1: Exemplo de ocultamento de uma mensagem

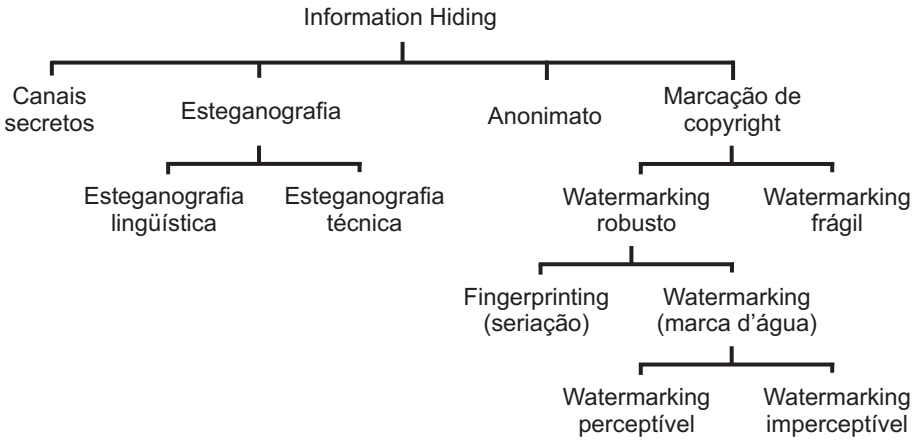

Figura 2: A hierarquia do *information hiding* [27]

Entende-se por *canais secretos*, a criação de uma comunicação entre duas partes em que o meio é secreto e seguro. Um exemplo seria as conversações militares em faixas de freqüências moduladas.

A arte da *esteganografia* constitui a segunda ramificação da hierarquia. Pode ser dividida em *lingüística* e *técnica*. Quando a mensagem é fisicamente escondida, tal como a maioria dos exemplos que serão apresentados na Seção 3, configura-se a chamada *esteganografia técnica*. Por outro lado, quando a mensagem é trabalhada e o seu ocultamento depende de propriedades lingüísticas, tal como a *esteganografia digital*, configura-se a chamada *esteganografia lingüística*.

Anonimato é um conjunto de técnicas para tentar navegar na *internet*, por exemplo, sem ser localizado. Isto poderia ser feito utilizando *sites* de desvio como o Anonymizer[2] e/ou *remailers* — *sites* capazes de enviar mensagens secretas não revelando seu remetente —.

Marcação de copyright é a tentativa de manter ou provar a propriedade intelectual sobre algum tipo de mídia, seja esta eletrônica ou impressa. Neste sentido, *sistemas de marcação robustos* (*Watermarking robusto*) são aqueles que, mesmo após tentativas de remoção, permanecem intactos. Por outro lado, *sistemas*

[2] `http://www.anonymizer.com`

de marcação frágeis (*Watermarking frágil*) são aqueles em que qualquer modificação na mídia acarreta perda na marcação. Estes sistemas são úteis para impedir a cópia ilegal. Ao se copiar um material original, o resultado é um material não marcado e, por conseguinte, pirata. *Sistemas de marcação imperceptível* (*Watermarking imperceptível*) são aqueles em que as logomarcas dos autores, por exemplo, encontram-se no material, mas não são diretamente visíveis. Em contrapartida, *marcação visível* (*Watermarking visível*) é aquela em que o autor deseja mostrar sua autoria a todos que observarem a sua criação. Um exemplo desta última forma é formado pelas imagens disponibilizadas na biblioteca do Vaticano[3]. Segundo [24], nesta biblioteca as imagens possuem um sistema de marcação digital visível como pode ser observado na Figura 3.

Figura 3: Exemplo de *marcação visível*. Biblioteca do Vaticano

3 Aspectos históricos

Através de toda a história, as pessoas têm tentado as mais inúmeras formas de esconder informações dentro de outros meios, buscando, de alguma forma, mais privacidade para seus meios de comunicação. Duas excelentes fontes podem ser encontradas em [19, 25].

Um dos primeiros registros sobre *esteganografia* aparece em algumas descrições de Heródoto, o pai da História, com vários casos sobre sua utilização. Um deles conta que um homem, de nome Harpagus, matou uma lebre e escondeu uma mensagem em suas entranhas. Em seguida, ele enviou a lebre através de seu mensageiro que se passou por um caçador [26].

Em outro caso, no século V AC, um grego de nome Histaieus, a fim de encorajar Aristágoras de Mileto e seus compatriotas a começar uma revolta contra Medes e os Persas, raspou a cabeça de um de seus escravos mais confiáveis e tatuou uma mensagem em sua cabeça. Assim que os cabelos do escravo cresceram, o

[3] `http://bav.vatican.va`

mesmo foi enviado à Grécia com instruções de raspar sua cabeça permitindo aos seus aliados receberem a mensagem [26].

Outra técnica interessante que aparece durante a História faz uso de inúmeras variações de tintas "invisíveis" (*invisible inks*). Tais tintas não são novidades e já apareciam em relatos de Plínio, o Velho, e Ovídio no século I DC. Ovídio, em sua *Arte do amor*, propusera o uso do leite para escrita de textos "invisíveis". Para decodificar a mensagem, o receptor deveria borrifar o papel com ferrugem ou carbono negro. Estas substâncias aderiam ao leite e a mensagem era revelada [19, 18].

Na segunda guerra mundial, com um aumento na qualidade das câmeras, lentes e filmes, tornou-se possível aos espiões nazistas, a criação de uma das formas mais interessantes e engenhosas de comunicação secreta. As mensagens nazistas eram fotografadas e, posteriormente, reduzidas ao tamanho de pontos finais (**.**) em uma sentença. Assim, uma nova mensagem totalmente inocente era escrita contendo o filme ultra-reduzido como final das sentenças. A mensagem gerada era enviada sem levantar maiores suspeitas. Esta engenhosidade ficou conhecida como *tecnologia do micro-ponto* [34].

Atualmente, a *esteganografia* não foi esquecida. Ela foi modificada em sinal de acompanhamento dos novos tempos. Na era da informação, não faz mais sentido escrever textos em tabletes de madeira ou mesmo borrifar pontos em uma revista através da utilização de tintas "invisíveis". Qualquer meio de *esteganografia* na atualidade, inevitavelmente, deve utilizar meios contemporâneos de tecnologia. Embora, em alguns casos, estes meios sejam apenas aperfeiçoamentos de técnicas clássicas.

4 O estado da arte da esteganografia

As abordagens mais comuns de inserção de mensagens em imagens incluem técnicas de:

- inserção no *bit* menos significativo;
- técnicas de filtragem e mascaramento;
- algoritmos e transformações.

Cada uma destas técnicas pode ser aplicada às imagens, com graus variados de sucesso. O método de inserção no *bit* menos significativo é provavelmente uma das melhores técnicas de *esteganografia* em imagem.

4.1 Inserção no bit menos significativo

Técnicas baseadas em LSB (*Least Significant Bit*) podem ser aplicadas a cada *byte* de uma imagem de 32-*bits*. Estas imagens possuem cada *pixel* codificado em quatro *bytes*. Um para o canal alfa (*alpha transparency*), outro para o canal vermelho (*red*), outro para o canal verde (*green*) e outro para o canal azul (*blue*). Seguramente, pode-se selecionar um *bit* (o menos significativo) em cada *byte* do *pixel* para representar o *bit* a ser escondido sem causar alterações perceptíveis na imagem [28, 26, 37].

Acompanhe o exemplo da Figura 4 para entender melhor. Suponha que se deseja esconder a letra **E** dentro da porção de imagem.

Na Figura 4, têm-se três *pixels* da imagem de cobertura. Como a letra **E** pode ser escrita em forma binária segundo seu código ASCII como **10000011**, é suficiente utilizar apenas os dois primeiros *pixels* da imagem. Assim, utilizando-se a técnica LSB, tem-se o resultado mostrado na Figura 5.

(00100111 11101001 11001000 11101010) [a, R, G, B]
(10100111 11001000 11101001 11101000) [a, R, G, B]
(11001000 00100111 11101001 00100111) [a, R, G, B]

Figura 4: Porção de uma imagem de cobertura

(00100111 11101000 11001000 11101010) [a, R, G, B]
(10100110 11001000 11101001 11101001) [a, R, G, B]
(11001000 00100111 11101001 00100111) [a, R, G, B]

Figura 5: Porção da estego-imagem gerada pela porção de imagem 4

Na Figura 5, os *bits* destacados pelo quadrado representam as modificações necessárias nos LSBs para esconder a letra **E**.

4.2 Técnicas de filtragem e mascaramento

Filtragem e mascaramento são técnicas mais robustas que a inserção LSB no sentido de gerarem estego-imagens imunes a técnicas de compressão e recorte. Ao contrário das modificações LSB, filtragem e mascaramento trabalham com modificações nos *bits mais significativos* das imagens. As imagens de cobertura devem ser em tons de cinza porque estas técnicas não são eficientes em imagens coloridas [28]. Isto deve-se ao fato de que modificações em *bits* mais significativos de imagens em cores geram alta quantidade de *ruído*[4] tornando as informações detectáveis.

4.3 Algoritmos e transformações

Utilizando técnicas como a *transformação discreta do cosseno*, *transformada discreta de Fourier* e *transformada Z*, entre outras, estes algoritmos tomam como aliado o principal inimigo da inserção LSB: a compressão. Por isso, configuram-se como as mais sofisticadas técnicas de mascaramento de informações conhecidas [28, 16].

4.4 Softwares de esteganografia disponíveis atualmente

Aplicações de *esteganografia* estão disponíveis na *internet* para executar em uma grande variedade de plataformas incluindo DOS, Windows, Mac OS, Unix e Linux.

Ezstego[5] e *Stego Online*[6] são duas ferramentas escritas na linguagem de programação *Java* e limitadas a imagens indexadas de oito *bits* e em formato GIF. Estas duas ferramentas foram desenvolvidas por Romana Machado.

Henry Hastur desenvolveu duas outras ferramentas: *Mandelsteg* e *Stealth*[7]. *Mandelsteg* gera imagens de fractais para esconder as mensagens. *Stealth* é um programa que recebe uma mensagem criptografada pelo

[4] Ruído aqui não é o ruído no sentido de *Processamento Digital de Imagens*. Este ruído explicita apenas a constatação de que o *bits* presentes não dizem nada (não possuem nenhum padrão) e estão aleatoriamente distribuídos.

[5] `http://www.stego.com`

[6] `http://www.stego.com`

[7] `ftp://idea.sec.dsi.unimi.it/pub/security/crypt/code/`

PGP, retira seu cabeçalho e a esconde em um arquivo. A retirada do cabeçalho de identificação é importante pois geralmente este contém informações do tipo de método criptográfico utilizado.

Duas outras ferramentas capazes de trabalhar em associação com a *criptografia* são *Ray Arachelian's White Noise Storm*[8] e o *S-Tools*[9].

Colin Maroney desenvolveu o *Hide and Seek*[10]. Esta ferramenta é capaz de mascarar uma lista de arquivos em uma imagem, mas não faz uso de *criptografia*. Niels Provos desenvolveu o *Outguess*[11] capaz de esconder mensagens com alto grau de robustez em vários formatos de arquivos de imagens. Testes estatísticos de primeira ordem não são capazes de detectar mensagens mascaradas com este *software* [31, 37].

Finalmente, duas outras ferramentas de destaque são *Jpeg-Jsteg*[12] capaz de fazer o mascaramento de informações utilizando os *pixels* mais significativos de uma imagem *jpg* e o *Camaleão*[13] [32] que utiliza cifragem de blocos e permutações cíclicas de modo a conseguir mascarar o conteúdo sem alterar significativamente as imagens.

5 Técnicas de esteganálise

Grande parte das técnicas de *esteganografia* possuem falhas e/ou inserem artefatos (padrões) detectáveis nos objetos de cobertura. Algumas vezes, basta um agressor (alguém interessado em descobrir indevidamente a mensagem) fazer um exame mais detalhado destes artefatos para descobrir que há mensagens escondidas. Outras vezes, o processo de mascaramento de informações foi robusto e as tentativas de detectar ou mesmo recuperar ilicitamente as mensagens podem ser bastante difíceis. Ao campo das pesquisas relacionado às tentativas de descobrir mensagens secretas dá-se o nome de *esteganálise*, uma alusão à *criptoanálise*, o campo de pesquisas relacionado à quebra de códigos [37, 38, 31].

Atualmente, as pesquisas em *esteganálise* estão mais concentradas em simplesmente identificar a presença de mensagens escondidas ao invés de extraí-las. Recuperar os dados escondidos, no momento, está além das capacidades da maioria dos testes porque muitos algoritmos de mascaramento utilizam geradores aleatórios criptográficos muito seguros para misturar a informação no processo de mascaramento. Na maioria das vezes, os *bits* são espalhados pelo objeto de cobertura. Os melhores algoritmos de *esteganálise* não são capazes de dizer onde está a informação, mas, provavelmente, podem dizer que os dados estão presentes.

A identificação da existência de uma mensagem escondida é suficiente para um agressor. As mensagens são, muitas vezes, frágeis e um agressor pode, sem muita dificuldade, destruir a mensagem mesmo sem tê-la recuperado. Algumas vezes, os dados podem ser destruídos simplesmente destruindo o objeto de cobertura. Outras vezes, basta aplicar um gerador de números aleatórios nos bits menos significativos destruindo qualquer mensagem (informação) ali presente. Ataques geométricos podem ser aplicados aos objetos de cobertura de modo que, se o agressor não pode recuperar a mensagem, então o destinatário também não poderá fazê-lo [37].

Todos estes ataques dependem da identificação de algumas características de um objeto de cobertura

[8] `ftp.csua.berkeley.edu/pub/cypherpunks/steganography/wns210.zip`
[9] `ftp://idea.sec.dsi.unimi.it/pub/security/crypt/code/s-tools4.zip`
[10] `ftp://csua.berkeley.edu/pub/cypherpunks/steganography/hdsk41b.zip`
[11] `http://www.outguess.org/`
[12] `ftp.funet.fi/pub/crypt/steganography`
[13] `http://www.ic.unicamp.br/~ra030014/ic/estego/`

(como imagens, vídeos, sons) que foram alteradas pelo processo de mascaramento.

Não há qualquer garantia de que um algoritmo *esteganográfico* possa resistir à *esteganálise*.

> *Pode-se desenvolver um software que seja capaz de enganar todos os computadores uma vez, ou mesmo pode-se enganar alguns computadores todas as vezes. No entanto, nunca se poderá desenvolver um software capaz de enganar todos os computadores todas as vezes.* (Anônimo)

5.1 Tipos de ataques

As principais abordagens de ataques à *esteganografia* em imagens são mostradas a seguir. Mais informações podem ser obtidas em [35, 38, 37, 14, 29, 30, 31].

5.1.1 Ataques aurais

Alguns ataques retiram as partes significativas da imagem como um meio de facilitar aos olhos humanos a busca por anomalias na imagem. Um teste comum é mostrar os *bits* menos signifitivos da imagem. Ruído completamente randômico freqüentemente revela a existência de uma mensagem escondida, uma vez que as imperfeições de câmeras, *scanners* e outros meios digitalizadores sempre deixam marcas de grande estrutura nos *bits* menos significativos[14].

Além disso, o cérebro do ser humano é capaz de descobrir as mais sutis diferenças. Esta é a razão pela qual muitas marcações de áudio (*audio watermarking*) de grandes gravadoras são frustradas graças aos ouvidos de músicos bem-treinados.

5.1.2 Ataques Estruturais

A estrutura do arquivo de dados freqüentemente muda assim que outra mensagem é inserida. Nesses casos, um sistema capaz de analisar padrões estruturais seria capaz de descobrir a mensagem escondida. Por exemplo, ao esconder mensagens em imagens indexadas (baseadas em paletas de cores), pode ser necessário usar versões diferentes de paletas. Este tipo de atitude muda as características estruturais da imagem de cobertura, logo as chances de detecção da presença de uma mensagem escondida aumentam.

5.1.3 Ataques estatísticos

Os padrões dos *pixels* e seus *bits* menos significativos freqüentemente revelam a existência de uma mensagem secreta nos perfis estatísticos [37]. Os novos dados não têm os mesmos perfis esperados.

Muitos dos estudos de *matemática estatística* objetivam classificar se um dado fenômeno ocorre ao acaso. Cientistas usam estas ferramentas para determinar se suas teorias explicam bem tal fenômeno. Estas técnicas estatísticas também podem ser usadas para determinar se uma dada imagem e/ou som têm alguma mensagem escondida. Isto é possível porque, na maioria das vezes, os dados escondidos são mais aleatórios que os dados que foram substituídos no processo de mascaramento ou inserem artefatos que alteram as propriedades estatísticas inerentes do objeto de cobertura.

Existem muitas abordagens estatísticas para esteganálise de imagens digitais tais como:

1. Teste do χ^2 (*Chi-Square Test*).

[14] Grande parte de câmeras digitais ou mesmo digitalizadores utilizam os LSBs das imagens para acrescentar seus próprios padrões. Isto elimina o ruído existente e acrescenta os padrões desejados pelos fabricantes.

2. Análise das inter-relações entre os planos de cores presente nas imagens analisadas (*RS Analysis*).
3. Análise de cores únicas no cubo RGB.
4. Taxa de inversão da energia do gradiente (*Gradient Energy Flipping Rate*).
5. Análise de estatísticas de alta ordem.
6. Métricas de qualidade de imagens (*Image Quality Metrics*).
7. Métricas de tons contínuos e análise de pares de amostragem (*Continuous Tone Metrics and Sample Pair Analysis*).

Teste do χ^2

Para determinar se uma imagem foi distorcida pelo processo de mascaramento, utiliza-se o teste do χ^2. No processo de mascaramento baseado na substituição LSB, caso exista alguma cor cr_1 na imagem que possua maior freqüência que alguma outra cor cr_2, provavelmente cr_1 terá seu LSB alterado mais vezes. Conseqüentemente, a diferença entre cr_1 e cr_2 aumentará. Pode-se dizer então que o processo de mascaramento reduz a diferença entre as cores [30]. Esta mesma observação é válida para imagens JPEGs. Entretanto, ao invés de medir as freqüências das cores, analisa-se as freqüências (histograma) dos coeficientes da transformada discreta do cosseno (*DCT – Discrete Cosine Transform*) da imagem em questão.

De forma geral, apenas a estego-imagem está presente para ser analisada e o teste χ^2 precisa ter uma distribuição esperada das freqüências dos coeficientes para ser calculado. 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 freqüência para o coeficiente DCT i na imagem. Assuma que uma imagem com mensagem escondida tenha freqüência similar para coeficientes DCTs adjacentes. Como resultado, pode-se tomar a média aritmética

$$y_i^* = \frac{n_{2i} + n_{2i+1}}{2}, \tag{1}$$

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

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

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

$$\chi^2 = \sum_{i=1}^{\nu+1} \frac{(y_1 - y_i^*)}{y_i^*}, \tag{3}$$

onde ν são os graus de liberdade, isto é, o número de diferentes categorias (coeficientes) no histograma.

A probabilidade p de mascaramento é dada pelo complemento da função de distribuição acumulativa,

$$p = 1 - \int_0^{\chi^2} \frac{t^{(\nu-2)/2} e^{-t/2}}{2^{\nu/2}\Gamma(\nu/2)} dt, \tag{4}$$

onde Γ é a função Euler-Gama. O cálculo desta probabilidade pode ser efetuado em diferentes partes da imagem e com diferentes janelas (amostragens).

De forma geral, para uma imagem que não contenha qualquer informação escondida, é esperada que a probabilidade de mascaramento seja próximo de zero em qualquer amostragem [30].

Análise RS

Este é um dos métodos de detecção mais robustos disponíveis. Suas análises podem ser válidas tanto para imagens em tons de cinza quanto coloridas. Não há distinção na profundidade de cores da imagem analisada, isto é, o método é válido tanto para imagens de 8 bpp (*bits* por *pixel*) quanto para imagens de 32 bpp. Algumas extensões do modelo inicial foram feitas e em algumas delas é possível também estimar o tamanho da mensagem escondida [13]. O método consiste em analisar a *capacidade de mascaramento sem perdas* do plano LSB[15].

A randomização do plano LSB, isto é, a aplicação de um gerador de números pseudo-aleatórios no plano LSB, diminui a capacidade de mascaramento sem perdas. No entanto, esta ação tem uma diferente influência na capacidade de mascaramento dos processos não limitados ao plano LSB [13]. É importante considerar que em um grande número de imagens, o plano LSB é essencialmente disperso, muito próximo de um estado pseudo-aleatório, e não possui qualquer artefato reconhecível. A análise da capacidade de mascaramento sem perdas reflete o fato de que mesmo que o plano LSB pareça ser independente dos demais planos, de alguma forma existe relações entre os mesmos. Esta relação é não linear e pode ser estimada pela capacidade de mascaramento sem perdas. Esta estimativa é feita 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 mascaramento e também na divisão da imagem analisada em grupos.

Para um maior entendimento, considere a imagem testada I com $M \times N$ *pixels*. Cada *pixel* tem valores dados por um conjunto P. Para uma imagem de 8 bpp, tem-se $P = \{0, \ldots, 255\}$. Então, divide-se I em grupos G de *pixels* disjuntos de n *pixels* adjacentes.

Como exemplo, pode-se escolher grupos de $n = 4$ *pixels* adjacentes. Feito isso, define-se uma função de discriminação f responsável por atribuir um número real $f(x_1, \ldots, x_n) \in \Re$ para cada grupo de *pixels* $G = (x_1, \ldots, x_n)$. O objetivo de f é capturar a regularidade (suavidade) do grupo de *pixels* G. Quanto mais distinto for o grupo de *pixels*, maior o valor da função de discriminação. Um exemplo pode ser

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i|. \tag{5}$$

Finalmente, define-se uma operação inversível F sobre P chamada *flipping*. Por *flipping* entende-se a permutação dos níveis de cores e consiste em dois ciclos. Assim, F tem a propriedade que $F^2 =$ Identidade ou $F(F(x)) = x$ para todo *pixel* $x \in P$. A função $F_1 : 0 \leftrightarrow 1, 2 \leftrightarrow 3, \ldots, 254 \leftrightarrow 255$ corresponde a inverter o LSB de cada nível de cor.

Adicionalmente, pode-se definir uma função $F_{-1} : -1 \leftrightarrow 0, 1 \leftrightarrow 2, \ldots, 255 \leftrightarrow 256$ de deslocamento (*shifting*). No caso de *inversão* e *deslocamento*,

$$F_{-1}(x) = F_1(x + 1) - 1 \text{ para todo } x \in P, \tag{6}$$

isto é, F_{-1} pode ser calculada a partir da função de inversão F_1. Para completar, define-se F_0 como sendo a função de identidade $F_0(x) = x$ para todo $x \in P$.

Caso seja desejado aplicar diferentes funções em diferentes *pixels*, deve-se usar uma máscara M que irá denotar quais os *pixels* deverão sofrer alterações. A máscara M é uma n-tupla com valores $\{-1, 0, 1\}$. O valor -1 denota a aplicação da função F_{-1}, 1 denota a aplicação da função F_1 e 0 denota a aplicação da

[15] Mascaramento sem perdas (*(lossless data embedding*) é o mascaramento em que todos os *bits* escondidos podem ser recuperados. Diferencia-se do mascaramento com perdas (*lossy*) em que é recuperada apenas uma estimativa dos *bits* da mensagem)

função F_0 sobre os *pixels* do grupo analisado. Uma máscara negativa $-M$ é definida como o complemento de M.

A função de discriminação f, bem como as funções F_1, F_{-1} e F_0 definidas mediante uma máscara M são então aplicadas sobre os grupos G para classificá-los em três categorias diferentes: R_M, S_M e U_M:

- Grupos regulares : $G \in R_M \Leftrightarrow f(F_M(G)) > f(G)$
- Grupos singulares : $G \in S_M \Leftrightarrow f(F_M(G)) < f(G)$
- Grupos não-usáveis : $G \in U_M \Leftrightarrow f(F_M(G)) = f(G)$

Da mesma forma, os grupos R_{-M}, S_{-M} e U_{-M} podem ser classificados sob a máscara $-M$. O objetivo da função F_M é perturbar os *pixels* de uma forma pouco significativa tal como aconteceria no processo de mascaramento de uma mensagem.

De forma geral, tem-se que

$$\frac{|R_M| + |S_M|}{T} \leq 1 \text{ e } \frac{|R_{-M}| + |S_{-M}|}{T} \leq 1, \tag{7}$$

onde T representa o número total de grupos G criados.

A hipótese estatística para o método é que, em imagens típicas, o valor esperado de R_M é aproximadamente igual ao de R_{-M} e o mesmo é verdade para S_M e S_{-M}.

A Equação 8 foi empiricamente comprovada por [13]. A randomização do plano LSB força a diferença entre R_M e S_M para zero à medida que o tamanho p da mensagem escondida cresce. Depois de alterar os LSBs de 50% dos *pixels* (é o que acontece quando se esconde uma mensagem aleatória em todos os *pixels*), obtém-se $R_M \cong S_M$, isto é o mesmo que dizer que a capacidade de mascaramento no plano LSB agora é zero. O fato surpreendente é que um efeito contrário acontece com R_{-M} e S_{-M}, sua diferença aumenta proporcionalmente ao tamanho da mensagem escondida.

$$R_M \cong R_{-M} \text{ e } S_M \cong S_{-M}. \tag{8}$$

A estimativa da presença ou ausência de uma mensagem escondida bem como do seu possível tamanho é feita através do cálculo por extrapolação da interseção dos pontos $R_M(50)$ e $S_M(50)$ que aparecem na Figura 6.

Análise de cores únicas no cubo RGB

Este método é baseado na análise estatística de cores do cubo RGB da imagem considerada. As técnicas de mascaramento LSB consideram que o plano menos significativo de *bits* é essencialmente randômico. Por conseguinte, a substituição dos *bits* deste plano por uma mensagem também randômica não irá artefatos detectáveis. No entanto, isso é essencialmente verdade apenas se o número de cores únicas[16] presente na imagem for comparável ao número de *pixels* da mesma. Entretanto, é observado que o número de cores únicas em imagens de cores verdadeiras (*true colors*) é tipicamente menor. A razão entre o número de cores únicas e o número de *pixels* variam de $1:2$ para imagens de alta qualidade a até $1:6$ para imagens *JPEGs* de menor resolução [12]. Esta observação é muito importante pois significa que muitas imagens de cores verdadeiras possuem uma paleta de cores relativamente pequena e, após o mascaramento LSB, a nova paleta de cores conterá muitos pares de cores relativamente muito próximas.

[16] Cada cor diferente presente na imagem.

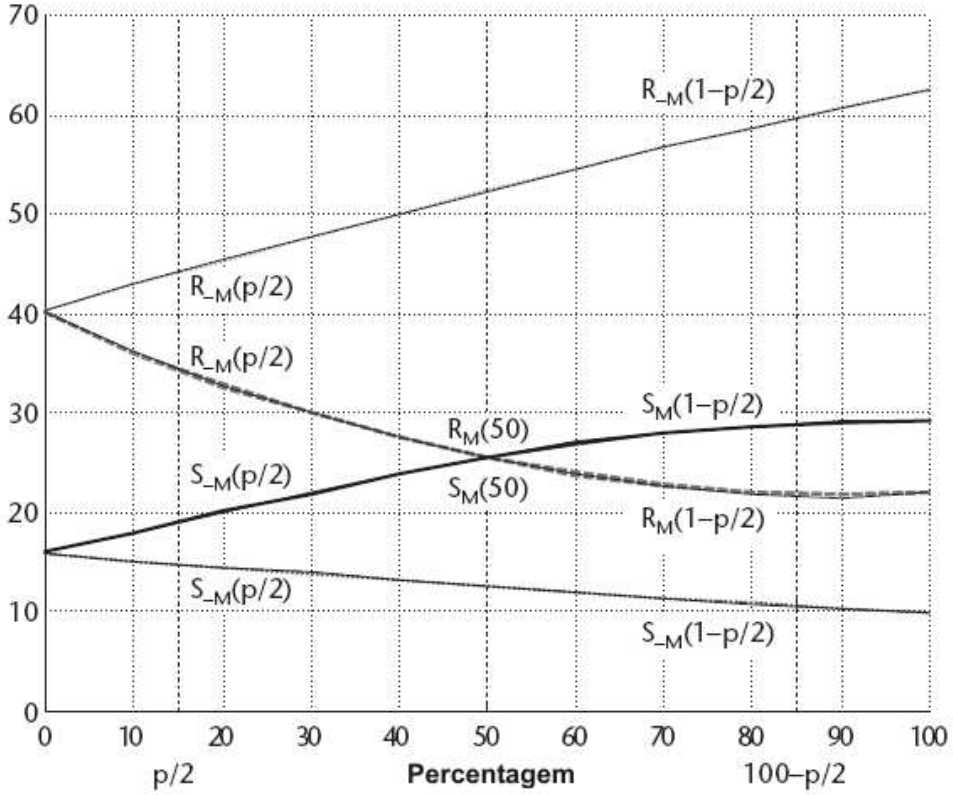

Figura 6: Diagrama RS de uma imagem. O eixo x é a percentagem de *pixels* cujos LSBs foram invertidos pela função F_M. O eixo y é o número relativo de grupos regulares e singulares sob as máscaras $M = [0, 1, 1, 0]$ e $-M = [0, -1, -1, 0]$.

Para fazer de tal artefato uma característica detectável, considere U o número de cores únicas em uma imagem. Olhando apenas em U, seja P o número de pares de cores próximas na paleta da imagem. Duas cores (R_1, G_1, B_1) e (R_2, G_2, B_2) são próximas se

$$(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2 \leq 3. \tag{9}$$

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

$$\binom{U}{2} \geq P. \tag{10}$$

A razão R entre o número de pares de cores próximas e o número de todos os pares de cores

$$R = \frac{P}{\binom{U}{2}}, \tag{11}$$

dá uma idéia sobre o número relativo de cores próximas na imagem. Após o mascaramento, o número de cores únicas na imagem irá aumentar. Para uma imagem que não contenha uma mensagem escondida, o número de pares de cores próximas em relação ao número de pares de todas as cores possíveis será menor que para uma imagem que contenha uma mensagem escondida. Embora esta afirmação possa ser verificada empiricamente, encontrar um limiar para R que separe imagens normais de imagens que contenham mensagens escondidas é uma tarefa quase impossível dado que o número de cores únicas em U pode variar muito. No entanto, uma observação interessante é que se uma imagem I já contém um conteúdo escondido,

o mascaramento uma mensagem adicional em I não altera significativamente o valor de R. Por outro lado, se I não contém uma mensagem escondida, o valor de R cresce significativamente após este procedimento. A razão R então é utilizada como critério de decisão para classificar uma imagem como tendo ou não uma mensagem escondida. O processo de decisão, sobre a imagem I de tamanho $M \times N$ testada, consiste em:

1. calcular a razão R entre o número de todos os pares de cores possíveis de cores próximas P e o número de todas cores possíveis U
$$R = \frac{P}{\binom{U}{2}}; \tag{12}$$

2. usando o mascaramento LSB, mascare uma mensagem de tamanho $3\alpha MN$ pseudo-aleatoriamente em I onde α será discutido posteriormente.

3. Denote 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
$$R' = \frac{P'}{\binom{U'}{2}}. \tag{13}$$

A hipótese estatística é que se I já contém uma mensagem escondida de tamanho significativo, as duas razões são praticamente as mesmas $R \cong R'$. Entretanto, se I não contém uma mensagem escondida, espera-se que $R' > R$. Para facilitar o cálculo da separabilidade pode-se definir a razão R'/R.

Caso o tamanho da mensagem escondida seja muito pequeno, as duas razões também serão muito próximas. Desta forma, a escolha correta de α deve ser feita de modo a minimizar o número de falsas detecções. O valor proposto por [12] e descoberto empiricamente é $\alpha = 5\%$.

Taxa de inversão da energia do gradiente

Consiste em analisar a variação da energia do gradiente, devido ao processo de mascaramento, dos planos de *bits* das imagens analisadas [33]. Para um maior entendimento, considere um sinal $I(n)$ unidimensional. O gradiente, $r(n)$, de um sinal antes do mascaramento é dado pela Equação 14.

$$r(n) = I(n) - I(n-1) \tag{14}$$

A *energia do gradiente*, EG, de $I(n)$ é dada pela Equação 15.

$$EG = \sum |I(n) - I(n-1)|^2 = \sum r(n)^2 \tag{15}$$

Depois do processo de mascaramento de um sinal, $S(n)$, no sinal original, $I(n)$ torna-se $I'(n)$ e o gradiente é reescrito como na Equação 16.

$$\begin{aligned} r(n) &= I(n) - I(n-1) \\ &= (I(n) + S(n)) - (I(n-1) + S(n-1)) \\ &= r(n) + S(n) - S(n-1) \end{aligned} \tag{16}$$

A função de distribuição de probabilidade de $S(n)$ é dada pela Equação 17.

$$\begin{cases} \rho_{S(n)\approx 0} &= \frac{1}{2} \\ \rho_{S(n)\approx \pm 1} &= \frac{1}{4} \end{cases} \tag{17}$$

Após o processo de mascaramento, a nova *energia do gradiente*, EG', é dada pela Equação 18.

$$\begin{aligned} GE' &= \sum |r(n)|^2 = \sum |r(n) + S(n) - S(n-1)|^2 \\ &= \sum |r(n) + \Delta(n)|^2, \text{onde } \Delta(n) = S(n) - S(n-1). \end{aligned} \tag{18}$$

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 como já definido para a *Análise RS*.

Caso a imagem de cobertura tenha $M \times N$ *pixels* e o tamanho da mensagem escondida seja $p \leq M\text{x}N$, a operação F resulta em três propriedades.

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

2. A *energia do gradiente* da imagem original é dada por $G(0)$. 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 $\frac{p}{2}$ *pixels* com seu LSB invertido. Denotamos a imagem modificada como $I(\frac{p}{2})$. A *energia do gradiente* correspondente é dada por $EG = \frac{p/2}{M \times N} = EG(0) + p$. Caso a operação F seja aplicada sobre a imagem $I(\frac{p}{2})$ a *energia do gradiente* resultante é $EG = \frac{M \times N - p/2}{M \times N}$.

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

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

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

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

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

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

Análise de estatísticas de alta ordem

Esta abordagem de detecção consiste em construir modelos estatísticos de alta ordem para imagens naturais e procurar por desvios nestes padrões [9, 20, 21, 10].

As imagens naturais possuem regularidades que podem ser detectadas com estatísticas de alta ordem através de uma decomposição *wavelet*, por exemplo [21]. O processo de mascaramento de uma mensagem altera significativamente estas estatísticas tornando o processo de mascaramento detectável.

Após a construção dos modelos de alta ordem é necessário utilizar classificadores capazes de dizer se uma dada imagem possui ou não uma mensagem escondida. Dois classificadores muito conhecidos são os SVMs (*Support Vector Machines*) [6] e os *Discriminantes de Fisher* [11].

O processo de decomposição das imagens usando funções base que são localizadas no domínio espacial de orientação e escala é extremamente útil em aplicações como compressão e codificação de imagens, remoção de ruído entre outras. Isto se deve ao fato destas decomposições exibirem regularidades estatísticas que podem ser exploradas.

A decomposição empregada neste tipo de detecção é baseada nos *filtros de quadratura em espelho* (QMFs – Quadrature Mirrors Filters) [36]. Esta decomposição divide a imagem no domínio da freqüência

em múltiplas escalas e orientações. Esta decomposição é feita através da aplicação de filtros de passa-baixas e passa-altas sobre a imagem gerando quatro sub-bandas: *vertical, horizontal, diagonal* e de *passa-baixas*. Escalas subseqüentes são criadas aplicando-se o processo novamente sobre a sub-banda de *passa-baixas*. As sub-bandas vertical, horizontal e diagonal em uma dada escala $\{i = 1 \ldots n\}$ são denotadas por $V_i(x, y)$, $H_i(x, y)$, $D_i(x, y)$, respectivamente. A Figura 7 demonstra o processo.

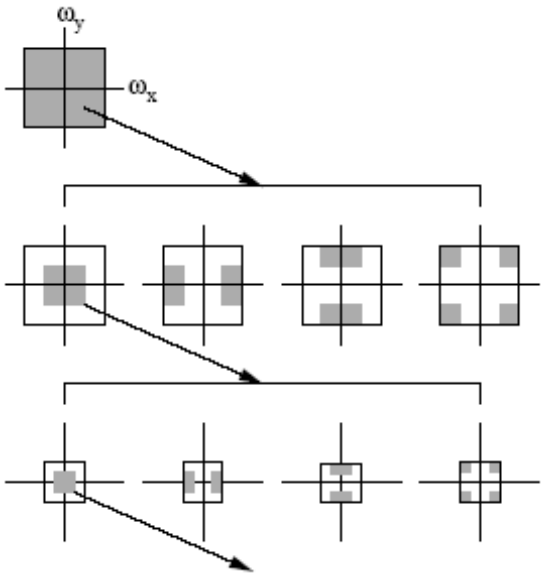

Figura 7: Decomposição da imagem no domínio da freqüência em múltiplas escalas e orientações

Dada esta decomposição da imagem, o modelo estatístico é composto de *média*, *variância*, *assimetria* e *curtose* dos coeficientes da sub-banda em cada orientação e escala $\{i = 1 \ldots n\}$. Estas estatísticas caracterizam os *coeficientes básicos de distribuição*.

Para um maior grau de exatidão no processo de classificação, um segundo conjunto de estatísticas é necessário . Este segundo conjunto é baseado nos erros de um *preditor linear ótimo de coeficentes de magnitude*. Os coeficientes das sub-bandas são correlacionados com seus vizinhos espaciais, de escala, e de orientação [5].

Para um maior entendimento, considere a sub-banda vertical, $V_i(x, y)$, na escala i. Um preditor linear para a magnitude destes coeficientes em um subconjunto de todos os possíveis vizinhos é dado por

$$\begin{aligned} V_i(x,y) &= w_1 V_i(x-1,y) + w_2 V_i(x+1,y) + w_3 V_i(x,y-1) \\ &\quad + w_4 V_i(x,y+1) + w_5 V_{i+1}(\frac{x}{2},\frac{y}{2}) + w_6 D_i(x,y) \\ &\quad + w_7 D_{i+1}(\frac{x}{2},\frac{y}{2}), \end{aligned} \tag{19}$$

onde w_k denota os valores escalares de peso dos coeficientes. Os coeficientes do erro são calculados através da minimização da *função de erro quadrática*

$$E(w) = [V - Qw]^2, \tag{20}$$

onde $w = (w_1, \ldots, w_7)^T$, V contém os coeficientes de magnitude de $V_i(x, y)$ dispostos em um vetor coluna e Q os coeficientes de magnitude dos vizinhos como especificado na Equação 19. A função de erro

é minimizada por diferenciar com respeito a w

$$\frac{dE(w)}{dw} = 2Q^T[V - Qw]. \tag{21}$$

Após algumas simplificações, o cálculo de w_k pode ser feito diretamente através do do *erro do log* no preditor linear

$$E = \log_2(V) - \log_2(|Qw|). \tag{22}$$

Todo o processo é recursivamente aplicado para cada sub-banda em cada escala e orientação. Ao final, tem-se um total de $12(n-1)$ estatísticas de erro mais $12(n-1)$ estatísticas básicas, totalizando $24(n-1)$ estatísticas para o vetor característico que deverá ser analisado pelo classificador escolhido.

O processo de classificação pode ou não necessitar de uma etapa de aprendizado. 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 proceder a classificação. Esta pode ser feita utilizando-se *SVMs lineares*, *SVMs não lineares* ou mesmo o *Discriminante de Fisher*. No processo sem aprendizado, utiliza-se *SVMs com hyper-esferas*. O número de *hyper-esferas* indica a exatidão do processo [21].

Métricas de qualidade de imagens

Estas métricas são utilizadas, de forma geral, na avaliação de codificação de artefatos, predição de performance de algoritmos de *Visão Computacional*, perda de qualidade devido a inadequabilidade de algum sensor, entre outras aplicações. Nesta abordagem, essas mesmas métricas são utilizadas para construir um discriminador de imagens de cobertura (naturais) de estego-imagens (com conteúdo escondido) através da utilização de *regressão multivariada* [2, 1, 3]. Também é feito um treinamento sobre uma estimativa da imagem original.

Dado que o mascaramento de uma mensagem pode ser interpretado como um sinal w adicionado ao sinal da imagem de cobertura f, tem-se que a estego-imagem gerada após o processo de mascaramento é dada por $g = f + w$. O discriminador é treinado sobre um conjunto de imagens de cobertura e de estego-imagens de modo a conseguir os coeficientes de qualidade de imagem que sejam capazes de separar as duas classes de imagens. Dado que no processo de *esteganálise* a imagem de cobertura quase sempre não está disponível para análise, é feita uma estimativa da imagem de cobertura através de filtragem baseada no filtro de baixa passagem Gaussiano [15].

As principais métricas de qualidade utilizadas são *média angular*, *distância de fase de bloco espectral*, *distância espectral ponderada da mediana de bloco*, *erro médio quadrático normalizado do Sistema Visual Humano (SVH)*, *medida de correlação de Czekznowski* e *erro médio absoluto*.

Após os cálculos de cada coeficiente para todas as imagens do conjunto de treinamento, é feita a regressão normalizada aos valores -1 e 1 para cada coeficiente. No modelo de regressão, cada decisão é expressa por y_i num conjunto de n imagens de observação.

Uma função linear dos coeficientes de qualidade de imagem é dada pela Equação 23.

$$\begin{cases} y_1 &= \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_q x_{1q} + \epsilon_1 \\ y_2 &= \beta_2 x_{21} + \beta_2 x_{22} + \ldots + \beta_q x_{2q} + \epsilon_2 \\ &\vdots \\ y_N &= \beta_N x_{N1} + \beta_2 x_{12} + \ldots + \beta_q x_{Nq} + \epsilon_N. \end{cases} \tag{23}$$

Nesta equação, x_{ij} indica os coeficientes de qualidade onde o primeiro índice i indica a i-imagem, $i = 1, \ldots, N$, e o segundo indica a métrica de qualidade considerada, $j = 1, \ldots, q$ e q é o número total de

métricas sob consideração. Os coeficientes de regressão são expressos por β_k e o erro aleatório[17] adicionado é expresso por ϵ.

Uma vez que estes coeficientes sejam obtidos na etapa de treinamento eles podem ser utilizados na etapa de teste. Dada uma imagem na etapa de teste, primeiro é obtido uma versão filtrada desta numa tentativa de estimar a imagem original de cobertura. Então, utilizando os coeficientes de predição é feita então a sua regressão até um valor de saída ser obtido. Caso o valor de saída supere o limiar 0 então a decisão sobre a hipótese estatística é que a imagem possui uma mensagem escondida. Caso contrário, a decisão é de que a imagem não contém mensagem escondida.

Métricas de tons contínuos e pares de amostragem

Esta abordagem consiste em analisar as relações de identidade estatística existentes sobre alguns conjuntos de *pixels* considerados. As identidades observadas são muito sensíveis ao mascaramento LSB e as mudanças nestas identidades podem indicar a presença de conteúdo escondido [8, 7].

Para um maior entendimento desta abordagem, considere um particionamento de uma imagem dentro de pares de *pixels* (u, v) horizontalmente adjacentes. Seja $\mathcal{P}$ o conjunto de todos estes pares. Defina os subconjuntos X e Y de $\mathcal{P}$:

- X é o conjunto de pares $(u, v) \in \mathcal{P}$ tais que v é par e $u < v$, ou v é ímpar e $u > v$.
- Y é o conjunto de pares $(u, v) \in \mathcal{P}$ tal que v é par e $u > v$, ou v é ímpar e $u < v$.

A importância de X e Y para a *esteganálise* vem da Equação 24.

$$|X| = |Y|. \tag{24}$$

A relação exposta pela Equação 24 é verdadeira para imagens naturais (sem mascaramento). Isto se deve ao fato de as imagens naturais serem isotrópicas em relação à *função de intensidade do gradiente* [8]. Em outras palavras, o gradiente tem igual probabilidade de ser o positivo e negativo. Defina Z como o subconjunto de pares $(u, v) \in \mathcal{P}$ tal que $u = v$. Além disso, considere a partição do subconjunto Y em dois subconjuntos: W e V, com W sendo o conjunto de pares em $\mathcal{P}$ da forma $(2k, 2k+1)$ ou $(2k+1, 2k)$ e $V = Y - W$. Neste caso, k demonstra a seleção de um par de *pixels* cuja diferença $|u - v| = 1$. Os conjuntos X, V, W e Z são chamados conjuntos primários, logo $\mathcal{P} = X \cup W \cup V \cup Z$.

Quando uma mensagem é escondida no plano LSB, o mascaramento modifica o valor de alguns *pixels* por inverter os LSBs. Assim, a cardinalidade dos *pixels* pertencentes a $\mathcal{P}$ irá mudar. Há quatro casos possíveis:

1. ambos os valores u e v não são modificados;
2. apenas u é modificado;
3. apenas v é modificado;
4. ambos os valores u e v são modificados.

Caso esteja ocorrendo a situação 1 (2,3,4) diz-se que os padrões de modificação devido ao mascaramento LSB é 00 (10, 01, 11, respectivamente). O processo de mascaramento leva a mudanças na pertinência de alguns pares de *pixels* entre os conjuntos primários. O processo é mostrado na Figura 8. Dados

[17] Este erro denota a variação natural que pode haver entre as amostras consideradas.

dois conjuntos quaisquer C_{Origem} e $C_{Destino}$, uma seta desenhada de C_{Origem} para o conjunto $C_{Destino}$ representa a transição de um par de *pixels* pertencente a C_{Origem} que passou para $C_{Destino}$ sob o padrão de modificação expresso pelo valor colocado sobre a aresta em questão. Para cada padrão de modificação $\pi \in \{00, 10, 01, 11\}$ e qualquer subconjunto $A \subset \mathcal{P}$, denote $\rho(\pi, A)$ a probabilidade de um par de *pixels* de A ser alterado pelo padrão π.

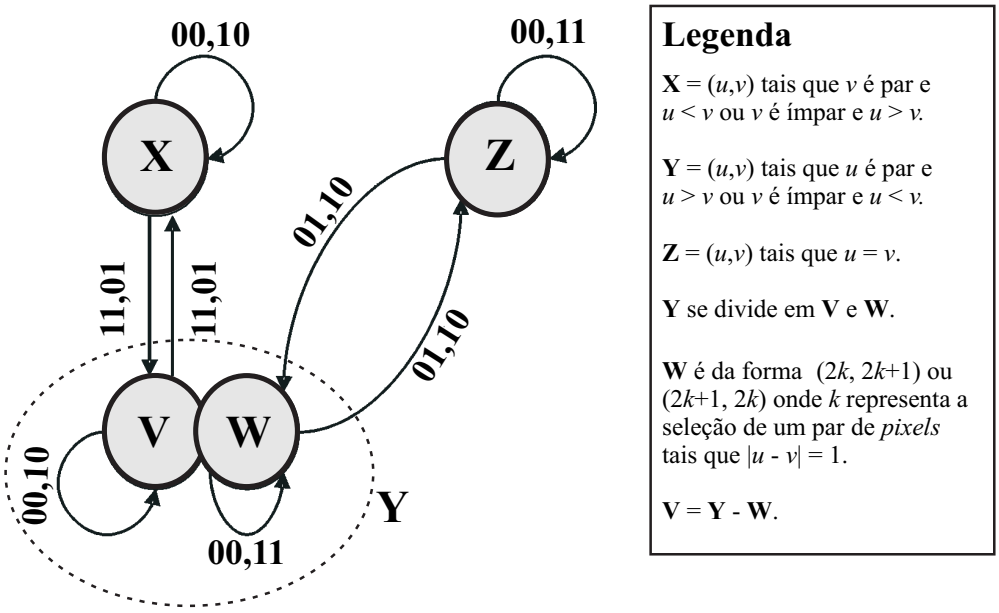

Figura 8: Diagrama de estados para as transições entre os conjuntos X, V, W, Z devido à inversão LSB

A partir das definições feitas, para cada padrão $\pi \in \{00, 10, 01, 11\}$ e cada conjunto primário $A \subset \{W, V, W, Z\}$,

$$\rho(\pi, A) = \rho(\pi, \mathcal{P}). \tag{25}$$

A Equação 25 implica que os *bits* da mensagem estão randomicamente distribuídos no plano LSB da imagem independentemente de qualquer característica da imagem. Desta forma, as relações a seguir são válidas. Seja p a razão de mascaramento dos *pixels* modificados devido ao mascaramento LSB pelo número total de *pixels*.

1. $\rho(00, \mathcal{P})$ = $(1 - p/2)^2$;
2. $\rho(01, \mathcal{P})$ = $\rho(10, \mathcal{P})$ = $p/2(1 - p/2)^2$;
3. $\rho(11, \mathcal{P})$ = $(p/2)^2$.

A partir do que foi exposto e do diagrama de estados da Figura 8, é possível expressar as cardinalidades dos conjuntos primários antes e depois do processo de mascaramento como funções de p. Para cada $A \in \{X, Y, V, W, Z\}$, seja A' o conjunto definido da mesma forma que A mas considerando os *pixels* após o processo de mascaramento. Desta forma, obtemos as relações expressas na Equação 26.

$$\begin{aligned} |X'| &= |X|(1 - p/2) + |V|p/2 \\ |V'| &= |V|(1 - p/2) + |X|p/2 \\ |W'| &= |W|(1 - p + p^2/2) + |Z|p(1 - p/2) \end{aligned} \tag{26}$$

A partir das relações expressas,

$$|X'| - |V'| = (|X| - |V|)(1 - p). \tag{27}$$

Dado que, estatisticamente, $|X| = |Y|$ tem-se que $|X| = |X| + |W|$ e assim

$$|X'| - |V'| = |W|(1 - p). \tag{28}$$

Observe na Figura 26 que o processo de mascaramento não altera o conjunto $W \subset Z$. Seja $\gamma = |W| + |Z| = |W'| + |Z'|$. Substituindo $|Z|$ por $\gamma - |W|$,

$$|W'| = |W|(1 - p)^2 + \gamma p(1 - p/2), \tag{29}$$

eliminando $|W|$,

$$|W'| = (|X'| - |V'|)(1 - p)^2 + \gamma p(1 - p/2), \tag{30}$$

dado que $|X'| + |V'| + |W'| + |Z'| = |\mathcal{P}|$,

$$0.5\gamma p^2 + (2|X'| - |\mathcal{P}|)p + |Y'| - |X'| = 0. \tag{31}$$

A Equação 31 resulta p, a estimativa do tamanho da mensagem escondida na imagem.

5.2 Cenários de ataque

Nas abordagens de ataque à *esteganografia* pode-se a combinação de diversos cenários [17].

- **Apenas estego (*stego-only*)**. Apenas o meio estego está disponível para análise.
- **Cobertura conhecida (*known-cover*)**. Estão disponíveis para análise tanto o meio estego quanto o meio original utilizado para cobertura.
- **Mensagem conhecida (*known-message*)**. Quando, de alguma forma, o *esteganalista* tem em seu poder uma mensagem escondida conhecida e a utiliza para comparação em busca de padrões nos meios de cobertura.
- **Meio estego escolhido (*chosen-stego*)**. Quando o *esteganalista* gera um meio estego a partir de um algoritmo esteganográfico conhecido e uma mensagem conhecida.
- **Mensagem escolhida (*chosen-message*)**. Quando o meio estego e o algoritmo de codificação utilizado são conhecidos.

6 Importância da pesquisa proposta

O interesse pela *esteganografia* deve-se, também, à busca por técnicas de *copyright* eficientes e eficazes. A partir do momento em que áudio, vídeo e outras formas de comunicação de mensagens tornaram-se disponíveis em formatos digitais, a facilidade com que qualquer um destes possa ser perfeitamente copiado aumentou significativamente. Isto está levando a uma imensa quantidade de reproduções destas formas de comunicação de mensagens não autorizadas pelo mundo todo [26]. Como contra-medidas, técnicas avançadas de "marcas-d'água" (*watermarking*) ou mesmo técnicas de seriação (*fingerprinting*), estruturadas na *esteganografia*, buscam restringir a pirataria indiscriminada. A *esteganánálise* vem ao encontro destas

tecnologias proporcionando técnicas que auxiliam na detecção de marcas mais frágeis além de fornecer os meios necessários à verificação de autenticidade de marcas resistentes [16].

A *esteganografia*, bem como a *esteganálise*, são muito novas. O primeiro debate aberto dessas tecnologias foi feito em 1996 no *First International Workshop in Information Hiding* [19]. A *esteganálise* é importante por ser um dos meios mais eficazes de combate às chamadas **dez bestas do tecnocalipse** [23, 22]:

1. fraudes e lavagem de dinheiro;
2. comunicações criminais;
3. roubo de propriedade intelectual;
4. *hacking*;
5. desvio de dinheiro;
6. jogos e pornografia;
7. assédio e extorsão;
8. disseminação de vírus;
9. pedofilia;
10. tráfico de pessoas.

Qualquer indivíduo, seja ele qualificado ou não, pode associar as técnicas acima à *esteganografia* tornando sua atividade altamente perigosa à sociedade e à segurança nacional.

7 Objetivos e metodologia

7.1 Objetivos

Este projeto visa atender às seguintes tarefas:

- estudar as diversas técnicas de esteganálise levando em conta seu desempenho e aproveitamento real como meio de detecção de mensagens escondidas em imagens digitais tais como textos maliciosos, conteúdo pornográfico e/ou pedófilo entre outros;
- identificar as vantagens e desvantagens de tais técnicas;
- buscar, na literatura, e/ou propor possíveis soluções para minimizar estas desvantagens;
- desenvolver uma ferramenta (software) onde será possível acompanhar o processo de algumas técnicas de detecção. Pretende-se implementar algumas destas técnicas além de propor formas híbridas de modo a diminuir as fraquezas conhecidas em outras técnicas do tipo. Deste modo, objetiva-se criar uma ferramenta robusta que possa servir efetivamente à sociedade e às autoridades como um meio eficaz na detecção de atos ilícitos escondidos em imagens digitais;
- disponibilizar todo o material bibliográfico utilizado para o desenvolvimento da pesquisa. Desta forma, há a divulgação dos estudos do andamento da pesquisa. Para isso, será construída uma página (site) na rede mundial de computadores (*internet*);

7.2 Metodologia

Este projeto atenderá aos seus objetivos utilizando-se os seguintes métodos:

- **Levantamento bibliográfico**. Será feito um levantamento de artigos científicos clássicos e atuais relacionados à *esteganografia* e *esteganálise*.
- **Estudo sobre os impactos da *esteganografia***. Será realizado um estudo sobre os impactos da *esteganografia* no mundo e seu real perigo para a sociedade tanto civil. Isto objetiva levantar as principais necessidades de se estudar ferramentas capazes de combatê-la.
- **Estudo das principais técnicas de detecção**. Um estudo das técnicas de detecção de mensagens mascaradas em imagens atualmente. Estas são o estado da arte da *esteganálise*.
- **Implementação**. Dá-se início à implementação mantendo-se a preocupação de construir códigos-fonte manuteníveis será constante. O paradigma de programação a ser utilizado é a orientação a objetos e a linguagem de programação será JAVA devido a alguns aspectos intrínsecos considerados importantes, por exemplo, a portabilidade.
- **Validação**. Terminada a implementação, passa-se para a etapa de testes em laboratório com o uso de exemplos práticos.
- **Documentação**. Será desenvolvida uma documentação das técnicas desenvolvidas para posterior divulgação em artigos científicos e na *internet*.
- **Artigos, dissertação e relatórios**. Ao término do projeto, artigos serão elaborados para divulgação através da submissão à eventos e periódicos científicos relacionados ao tema. Além disso, um produto de *software* e uma dissertação de mestrado são esperados como resultado final do projeto.

8 Cronograma e plano de trabalho

Abaixo se encontra o cronograma de atividades a ser seguido.

	MARÇO DE 2004 A FEVEREIRO DE 2005											
ETAPA	MAR	ABR	MAI	JUN	JUL	AGO	SET[1]	OUT	NOV	DEZ	JAN	FEV
1												
2												
3												
4												
5												
6												
7												
8												

[1] Data de efetivação da bolsa FAPESP

MARÇO DE 2005 A FEVEREIRO DE 2006												
ETAPA	MAR	ABR	MAI	JUN	JUL	AGO[1]	SET	OUT	NOV	DEZ	JAN	FEV[2]
1												
2												
9												
10												
11												
12												
13												
14												
15												

[1] Relatório técnico a ser entregue à FAPESP (10/08/05).

[2] Relatório técnico a ser entregue à FAPESP (10/02/06).

1. **Coleta** de material bibliográfico.
2. **Desenvolvimento** do site da pesquisa.
3. **Estudo** da *esteganografia* clássica e sua história.
4. **Impactos** da *esteganografia digital* na sociedade como um todo. Quais serão os efeitos de curto, médio e longo prazos da *esteganografia digital*. A sociedade estaria ameaçada?
5. **Estudo do estado da arte** de técnicas *esteganográfico-digitais*.
6. **Comparação das vantagens e desvantagens** de tais técnicas. Nem todo sistema baseado em técnicas *esteganográfico-digital* é perfeito. Listar e apontar as principais limitações destes sistemas para se poder desenvolver técnicas robustas de detecção.
7. **Estudo das principais técnicas de detecção** de mensagens escondidas em imagens
8. **Comparação das técnicas** estudadas e seleção das mais robustas.
9. **Implementação das técnicas** previamente selecionadas.
10. **Proposição de melhorias** às técnicas estudadas bem como implementação de uma técnica própria.
11. **Testes de verificação** da implementação.
12. **Documentação**.
13. **Escrita do relatório final** a ser entregue à FAPESP.
14. **Dissertação**.
15. **Escrita de artigos**. Escrita de artigos relacionados à pesquisa e submissão a eventos e/ou periódicos científicos relacionados ao tema.

Referências

[1] AVCIBAS, I., MEMON, N., AND SANKUR, B. Steganalysis based on image quality metrics. In *IEEE 4th Workshop on Multimedia Signal Processing* (Oct. 2001), pp. 517–522.

[2] AVCIBAS, I., MEMON, N., AND SANKUR, B. Steganalysis based on image quality metrics. Master's thesis, Department of Computer and Information Science Polytechnic University, Brooklyn, USA, 2002.

[3] AVCIBAS, I., MEMON, N., AND SANKUR, B. Steganalysis using image quality metrics. In *IEEE Transactions On Image Processing* (Feb. 2003), vol. 12, pp. 221–229.

[4] BAGNALL, R. J. Reversing the steganography myth in terrorist operations: The asymmetrical threat of simple intelligence dissemination techniques using common tools. The SANS Institute, 19 de agosto de 2002. Disponível em `http://www.sans.org/rr/papers/index.php?id=556`, 2002.

[5] BUCCIGROSSI, R. W., AND SIMONCELLI, E. P. Image compression via joint statistical characterization in the wavelet domain. In *IEEE Transactions on Image Processing* (1998), vol. 8, pp. 1688–1701.

[6] BURDS, C. J. C. A tutorial on support vector machines for pattern recognition. In *Data Mining and Knowledge Discovery* (1998), vol. 2, pp. 121–167.

[7] DUMITRESCU, S., AND WU, X. Steganalysis of lsb embedding in multimedia signals. In *Proceedings of the International Conference on Multimedia and Expo* (Aug. 2002), vol. 3, pp. 581–584.

[8] DUMITRESCU, S., WU, X., AND MEMON, N. On steganalysis of random lsb embedding in continuous-tone images. In *Proceedings of the International Conference on Image Processing* (June 2002), vol. 3, pp. 641–644.

[9] FARID, H. Detecting steganographic messages in digital images. In *Technical Report TR2001-412* (Hanover, USA, 2001).

[10] FARID, H. Detecting hidden messages using higher-order statistical models. In *Proceedings of the International Conference on Image Processing* (2002), vol. 2, pp. 905–908.

[11] FISHER, R. The use of multiple measures in taxonomic problems. In *Annals of Eugenics* (1936), vol. 7, pp. 179–188.

[12] FRIDRICH, J., DU, R., AND LONG, M. Steganalysis of lsb enconding in color images. In *ICME – International Conference Multimedia Expo* (Aug. 2000), vol. 3, pp. 1279–1282.

[13] FRIDRICH, J., GOLJAN, M., AND DU, R. Detecting lsb steganography in color and grayscale images. In *IEEE Proceeding on Multimedia and Security* (Jan. 2001), IEEE Multimedia.

[14] FRIDRICH, J., GOLJAN, M., AND DU, R. Reliable detection of lsb steganography in color and grayscale images. *Center for Intelligent Systems, SUNY Binghamton* (2002).

[15] GONZALEZ, R. C., AND WOODS, R. E. *Digital Image Processing*. Prentice-Hall, Boston, 2002. ISBN 0-20118-075-8.

[16] JOHNSON, N., AND JAJODIA, S. Exploring steganography: seeing the unseen. In *IEEE Internet Computing* (Feb. 1998).

[17] JOHNSON, N., AND JAJODIA, S. Steganalysis of images created using current steganography software. In *Lecture Notes in Computer Science* (Apr. 1998), vol. 1525, pp. 273–289.

[18] KAHN, D. *The CODEBREAKERS: The Comprehensive History of Secret Communication from Ancient Times to the Internet*. Scribner, Boston, 1996. ISBN 0-68483-130-9.

[19] KUHN, M. G. The history of steganography. In *Proceedings of the First International Information-Hiding Workshop* (May 1996), Springer–Verlag, Berlim.

[20] LYU, S., AND FARID, H. Detecting hidden messages using higher-order statistics and support vector machines. In *Lectures Notes on Computer Science, Proceedings of the 5th International Workshop on Information Hiding* (2002), pp. 340–354. Noordwijkerhout, The Netherlands.

[21] LYU, S., AND FARID, H. Steganalysis using color wavelet statistics and one-class support vector machines. Master's thesis, Department of Computer Science, Dartmouth College, Hanover, USA, 2002.

[22] NHTCU. NHTCU – National High Tech Crime Unit. Disponível em `www.nhtcu.org`, 2003.

[23] USPS. USPS – US Postal Inspection Service. Disponível em `www.usps.com/postalinspectors/ar01intr.pdf`, 2003.

[24] MINTZER, F. C., BOYLE, L. E., AND CASES, A. N. Toward on-line, worldwide access to vatican library materials. In *IBM Journal of Research and Development* (1996), vol. 40.

[25] NORMAN, B. *Secret warfare, the battle of Codes and Ciphers*. Acropolis Books Inc, 1980.

[26] PETITCOLAS, F. A., ANDERSON, R. J., AND KUHN, M. G. Information hiding - a survey. In *Proceedings of IEEE* (July 1999). Special issue on Protection on multimedia content.

[27] PFITZMANN, B. Information hiding terminology. In *Proceedings of the first international information-hiding workshop* (May 1996), Springer–Verlag, Berlim.

[28] POPA, R. An analysis of steganography techniques. Master's thesis, Department of Computer Science and Software Engineering of The "Polytechnic" University of Timisoara, Timisoara, Romênia, 1998.

[29] PROVOS, N. Defending against statistical steganalysis. In *Proceedings of the 10th USENIX Security Symposium* (2001). Washington, DC, USA.

[30] PROVOS, N., AND HONEYMAN, P. Detecting steganographic content on the internet. In *Technical Report CITI 01-1a, University of Michigan* (2001). Michigan, USA.

[31] PROVOS, N., AND HONEYMAN, P. Hide and seek: An instroduction to steganography. In *IEEE Security & Privacy Magazine* (May 2003), vol. 1, pp. 32–44.

[32] ROCHA, A. R. Camaleão: um software para segurança digital utilizando esteganografia. In *Monografia de final de curso* (2003), Departamento de Ciência da Computação, Universidade Federal de Lavras. Disponível em `www.comp.ufla.br/curso/ano2003/`.

[33] SHI, L., FEN, S. A., AND XIAN, Y. Y. A lsb steganography detection algorithm. In *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications* (2003), vol. 3, pp. 2780–2783.

[34] SINGH, S. *O livro dos códigos*. Record, Rio de Janeiro, 2001. ISBN 8-50105-598-0.

[35] TZSCHOPPE, R., BäUML, R., HUBER, J. B., AND KAUP, A. Steganographic system based on higher-order statistics. *Center for Intelligent Systems, SUNY Binghamton* (2003).

[36] VAIDYANATHAN, P. P. Quadrature mirror filter banks, m-band extensions and perfect reconstruction techniques. In *IEEE ASSP Magazine* (1987), pp. 4–20.

[37] WAYNER, P. *Disappearing cryptography*. Morgan Kaufmann Publishers, San Francisco, 2002. ISBN 1-55860-769-2.

[38] WESTFELD, A., AND PFITZMANN, A. Attacks on steganographic systems. In *Lecture Notes In Computer Science, Proceedings of the 3rd International Workshop on Information Hiding* (1999), pp. 61–76.

[39] ZHANG, T., AND PING, X. Reliable detection of lsb steganography based on the difference image histogram. In *Proceedings on Acoustics, Speech, and Signal Processing* (2003), vol. 3, pp. 543–548.