Métodos de Esteganografia Aplicados a Imagens Médicas

Rafael de Assunção Sampaio
rsampaio@ime.usp.br

Orientador: Prof. Dr. Marcel P. Jackowski

13 de novembro de 2012

Motivações

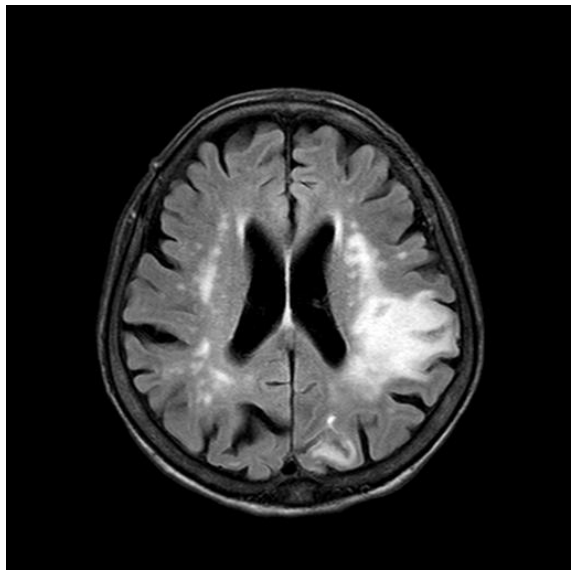
- Disciplina de Sistemas Operacionais
- Interdisciplinaridade
- Esteganografia? O quê?
- MedSquare – http://ccsl.ime.usp.br/medsquare/

Objetivos

- Estudar métodos de esteganografia aplicáveis às imagens médicas;
- Propor uma mudança na representação das imagens médicas, incorporando-se um algoritmo de esteganografia ao padrão DICOM, para proteger informações relevantes.

Imagens Médicas

- As imagens médicas são obtidas por meio de processos especializados, como radiografia, ressonância magnética, tomografia computadorizada. Tais imagens são essenciais para avaliação de tumores, condição óssea, estado de gestação etc.



Ressonância magnética de encéfalo

DICOM

- DICOM (*Digital Imaging and Communications In Medicine*) é um padrão bem estabelecido de *armazenamento*, *impressão* e *transmissão* da informação presente nas imagens médicas.

- Os arquivos possuem duas componentes:
  - cabeçalho, em que se encontram diversas informações confidenciais do paciente, do local onde o exame foi realizado, dos aparelhos e responsáveis, entre outras;
  - matriz, em tons de cinza, que representa a imagem obtida pelo aparelho.

Exemplo de cabeçalho DICOM

- *Tags* referentes ao exame:

0008-0070	Manufacturer	LO	Philips Medical Systems
0008-0080	InstitutionName	LO	INSTITUTO DE RADIOLOGIA
0008-0090	ReferringPhysicianName	PN	.^NAO LOCALIZADO^^^
0008-0100	CodeValue	SH	
0008-0102	CodingSchemeDesignator	SH	DCM
0008-0104	CodeMeaning	LO	
0008-1010	StationName	SH	ACHIEVA
0008-1030	StudyDescription	LO	R.M. ENCEFALO

- *Tags* referentes ao paciente:

0010-0010	PatientName	PN	Morris^Rachel^B^^^
0010-0020	PatientID	LO	02005A46
0010-0030	PatientBirthDate	DA	19270127
0010-0040	PatientSex	CS	F
0010-1000	OtherPatientIDs	LO	204175157280003
0010-1030	PatientWeight	DS	49
0010-21C0	PregnancyStatus	US	4

Pontos fracos DICOM

- O padrão não dispõe de métodos de segurança, armazenando os dados em *tags* irrestritamente editáveis;
- As imagens podem inadvertidamente ser compartilhadas, expondo tais informações, sem que se possa verificar a autenticidade e o uso indevido.

O que é Esteganografia?

- A palavra *esteganografia* tem origem nas palavras gregas *steganos*, que significa "encoberto", e *graphia*, que significa "escrita";
- A Esteganografia é formada por um conjunto de técnicas que permitem ocultar informação, por exemplo, em imagens, de tal sorte que ninguém desconfie de sua *existência*.

Exemplo

A imagem à direta foi ocultada na imagem à esquerda utilizando esteganografia[1]

[1] Fonte: http://en.wikipedia.org/wiki/Steganography

Ocorrências de Esteganografia

- Heródoto, um géografo e historiador grego do século V a.C., narra sobre um escravo que teria sido enviado por Histiaeus a Mileto com uma mensagem secreta tatuada na cabeça.
- Embate enxadrístico entre Viktor Korchnoi e Anatoly Karpov pelo Campeonato Mundial de Xadrez em 1978. Durante uma das partidas, um dos assistentes de Karpov traz um iugorte. A delegação de Korchnoi imediatamente protestou, alegando que, dependendo da cor do iogurte, isso poderia significar oferecer ou rejeitar um empate. Karpov passaria a poder consumir apenas um iugorte – sempre da mesma cor – num mesmo momento das partidas.
- No *The New York Times* de 11 de novembro de 2006, lê-se que Dhiren Barot, membro da Al Qaeda, realizou uma filmagem entre a *Broadway* e *South Street*, ocultando esse material numa cópia de *Die Hard: With a Vengeance*, com Bruce Willis. Esse vídeo teria sido utilizado nos ataques de 11 de setembro de 2001.

Inserção no Bit Menos Significativo (LSB Insertion)

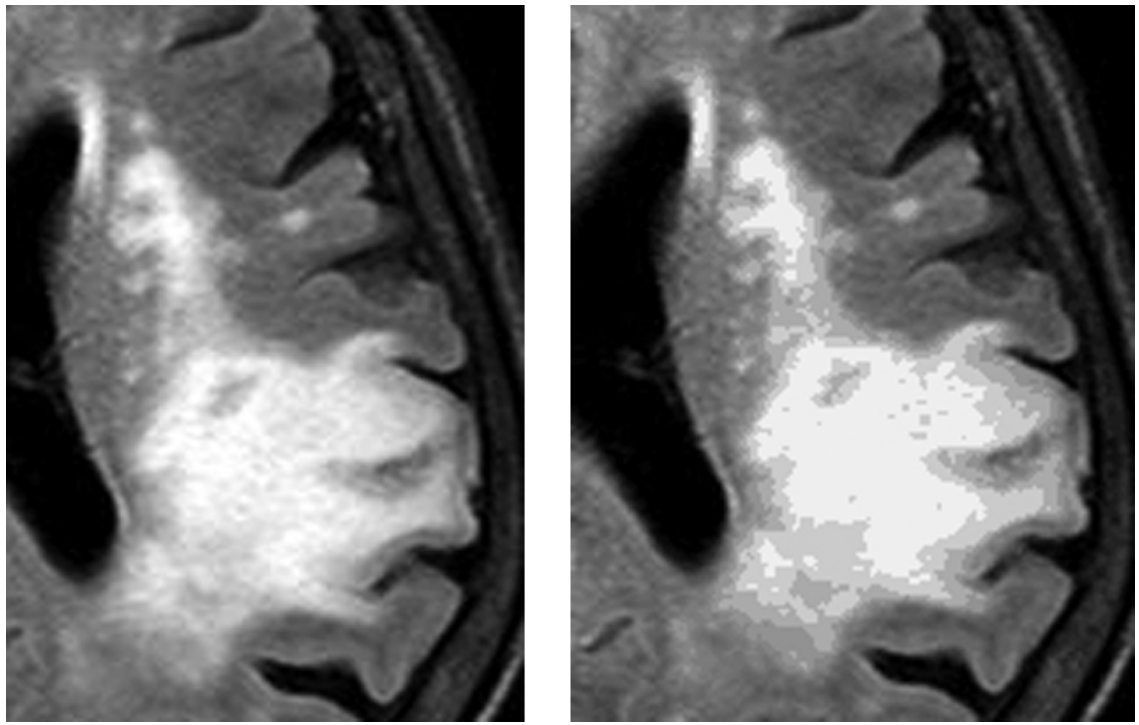
- Considera-se a representação binária da informação S a ser ocultada, o bit menos significativo de cada pixel da imagem será sobrescrito por S_i, para $1 \leq i \leq |S|$. No caso de imagens coloridas, por exemplo, 24-bit, a modificação é mínima e imperceptível ao olho humano. As imagens 8-bit, em tons de cinza, porém, sofrem alterações relevantes e são facilmente detectadas;

- Exemplo: a letra A é representada por 65 (código ASCII), cuja representação binária é 1000001. Para esta sequência de pixels de uma imagem 8-bit: 10000000, 10100100, 10110101, 10110101, 11110011, 10110111, 11100111, a transformação seria 1000000**1**, 1010010**0**, 1011010**0**, 1011010**0**, 1111001**0**, 1011011**0**, 1110011**1**.

Pontos fracos do LSB

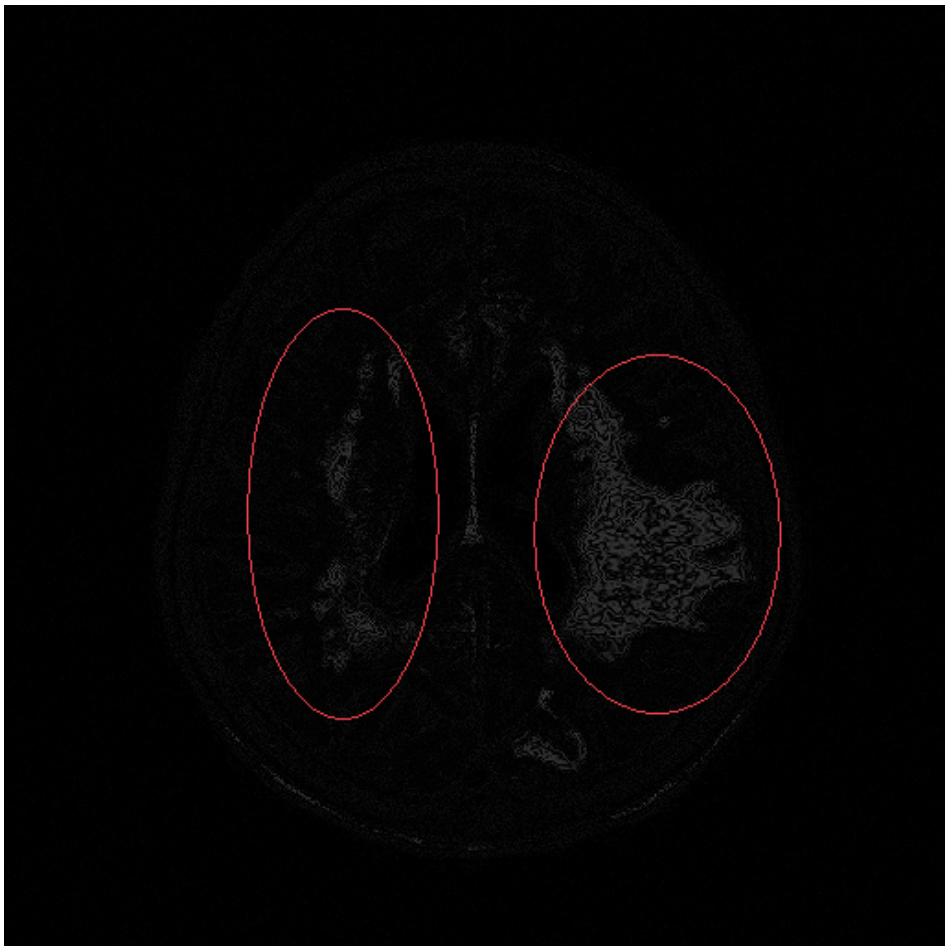
- Pouco robusto: a maioria das transformações geométricas e filtros, bem como esquemas de compressão (por ex., JPEG), modificam os pixels menos significativos e a informação esteganografada acaba destruída;
- Fácil recuperação da informação que se pretende ocultar.

LSB em imagens médicas



À esquerda, a imagem original. À direita, a imagem após LSB.

LSB em imagens médicas



Diferença entre as imagens

Divisão em blocos e alteração média

- A imagem é dividida em blocos de dimensão "suficientemente grande", isto é, tais blocos são capazes de armazenar a informação S e não provocam diferença aparente antes e depois da esteganografia;
- Seja B um bloco de dimensão $m \times n$ da imagem, o bit que corresponde à média deste bloco será alterado para o k-ésimo bit de S, ou seja, $\frac{1}{m \times n} \sum_i \sum_j B_{ij} = S_k$;
- Embora esse método ofereça bons resultados, alguns problemas ocorrem: é aparente nas regiões uniformes da imagem o efeito do bloco, ainda que a dimensão tenha sido bem escolhida; é necessário um pré-processamento da imagem para definir o intervalo de tons de cinza, o que pode causar alteração do brilho da imagem ou eliminar pequenos detalhes.

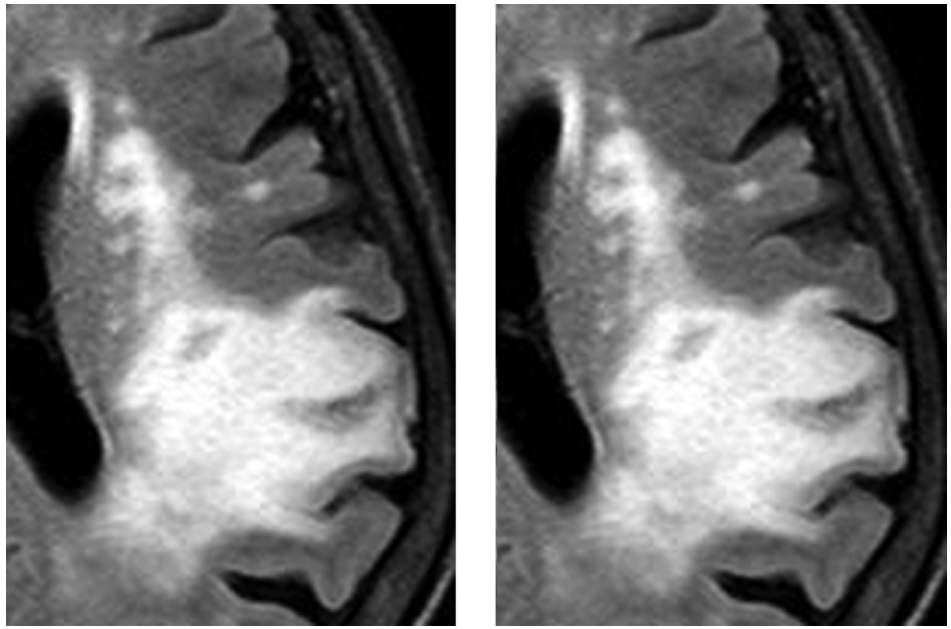
Método de alteração da média modificado

- Para resolver as incorreções do método anterior, é proposto em [2] que seja realizado um processo dinâmico de tomada de decisão, a partir de uma imagem cujos pixels são embaralhados e, posteriormente, tomados em blocos;
- Esse algoritmo reduz o valor de cada pixel de cada bloco até que a alteração necessária seja atingida, mas evita que a alteração total exceda um limitante previamente estipulado;
- A tomada de decisão baseia-se no espectro das médias geradas e pode ser definida pela equação: $M_d(i) = argmin|M_{S_i}(j) - M_i|$, em que M_i é o valor médio do bloco B, $M_{S_i}(j)$ é o j-ésimo centro do espectro médio para S_i e $M_d(i)$ será a nova média.

Algoritmo

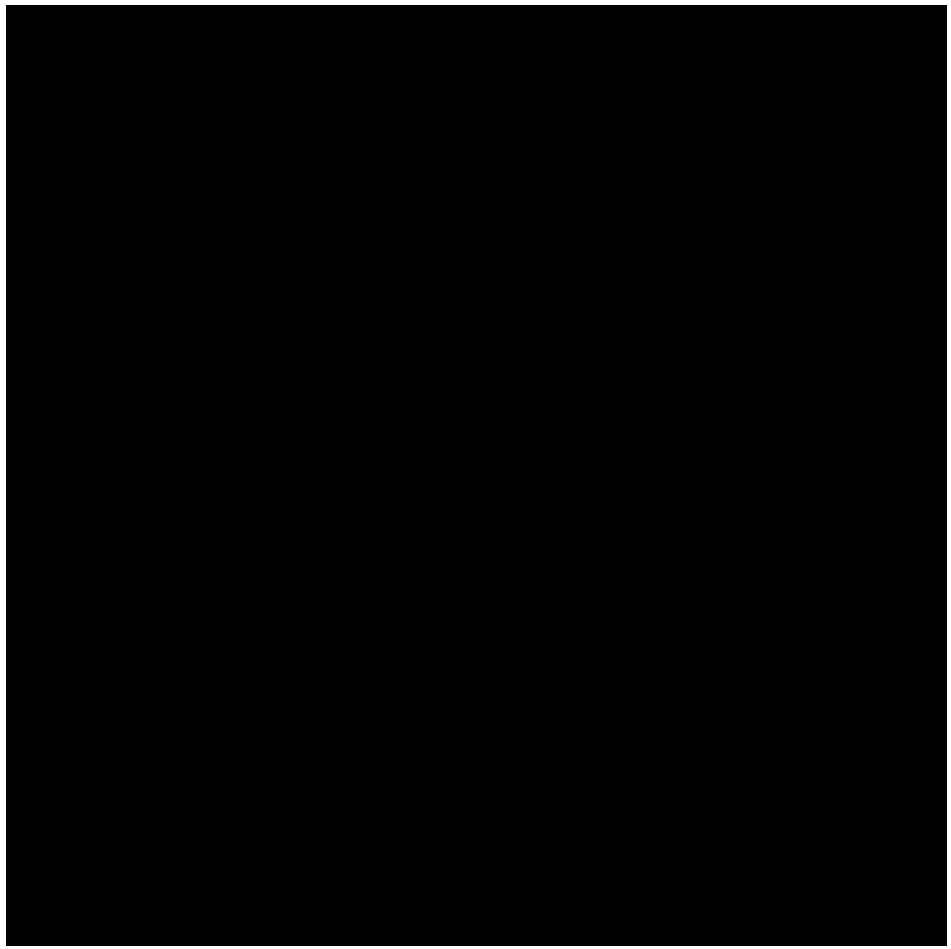
1. Codifico a mensagem em representação binária;
2. Embaralhamento (*shuffle*) da imagem original utilizando a *chave*;
3. Seja $i = 1$;
4. A partir de i, seleciono N pixels, de $(i-1) \times N + 1$ a $i \times N$, como $CB(i)$;
5. Determino $M(i)$;
6. Determino $M_d(i)$ – tomada de decisão;
7. Altero os pixels necessários, considerando o *threshold* estipulado;
8. Escrevo $CB(i)$ na imagem embaralhada;
9. Enquanto houver bits a serem esteganografados, repito o processo;
10. Desembaralho a imagem para obter a imagem final.

Experimento



À esquerda, a imagem original. À direita, a imagem após aplicação do MAMM.

Experimento



Diferença entre as imagens

Métricas e Resultados

- Alteração visual significativa, visando à utilização clínica;
- Métricas de similaridade descritas em [3]: Soma dos Quadrados das Diferenças (SSD), Soma das Diferenças Absolutas (SAD) e Máxima Diferença Absoluta (MAD);
- O método de alteração da média modificado produz apenas alterações de baixa degradação na imagem original, o que viabiliza sua utilização clínica.

Desafios

- Bibliografia escassa;
- Autor do principal artigo não retornou.

Aprendizados

- Web design, Matlab, ImageMagick, LaTeX / Beamer;
- Ler e preencher lacunas de artigos científicos;
- Produzir qualquer trabalho predispõe continuação: manter-se disponível e colaborar!

Pesquisa

- Aumento do *payload* com a utilização de vários cortes;
- Códigos corretores de erros;
- OCR para ocultação de informação na superfície da imagem;
- A combinação de Esteganografia e Criptografia é boa?

Referências

1 J. Fridrich. "Steganography in Digital Media: Principles, Algorithms, and Applications". Cambridge University Press. December 21, 2009.

2 P. Mortazavian, M. Jahangiri and E. Fatemizadeh. "A Low-degradation Steganography Model For Data Hiding In Medical Images". Proceedings 4th IASTED. September 6-8, 2004.

3 J. N. Ulysses and A. Conci. "Measuring Similarity in Medical Registration". IWSSIP 2010.

Obrigado!