

The background of the slide is a collage of various scientific and digital elements. It includes chemical structures, a chromatogram plot, and text from a data file. The text visible in the background includes 'gram Plot', 'std ei,100(4)280(12) db1', '751', 'Retention Time: 14:34', 'RIC: 7125768', 'Gesamtreaktion 13.41:36', 'NH2', 'AanH', '+ 2 NaOH + 2 H2O2', '+ N2 + 4 H2O', 'COO- Na+', 'atogram Plot', 'ent: std ei,100(4)2', 'No: 751', 'Reten', 'ed: 200 to 800', and '* = Saturated scans'.

Análise Forense de Documentos Digitais

Prof. Dr. Anderson Rocha

anderson.rocha@ic.unicamp.br

<http://www.ic.unicamp.br/~rocha>

Reasoning for Complex Data (RECOD) Lab.
Institute of Computing, Unicamp

Av. Albert Einstein, 1251 – Cidade Universitária
CEP 13083-970 • Campinas/SP – Brasil

*** Slides preparados baseados em apresentação de
E. A. Silva e F. O. Costa, MO447 (2010s2)**

Detecção de Cópia-colagem em Imagens Digitais

Organização

Organização

- ▶ Motivação
- ▶ Introdução
- ▶ Técnicas de Manipulação
- ▶ Cópia-colagem
- ▶ Detecção de Cópia-colagem
- ▶ Estado da Arte

Motivação

Motivação

- ▶ É simples enganar observadores e espectadores utilizando imagens manipuladas digitalmente
- ▶ Pode-se formar falsas opiniões e criar situações que não condizem com a realidade

Introdução

Introdução

- ▶ Facilidade na manipulação de imagens digitais por meio de ferramentas de *software*:
 - Adobe Photoshop
 - GIMP
 - etc.

Introdução

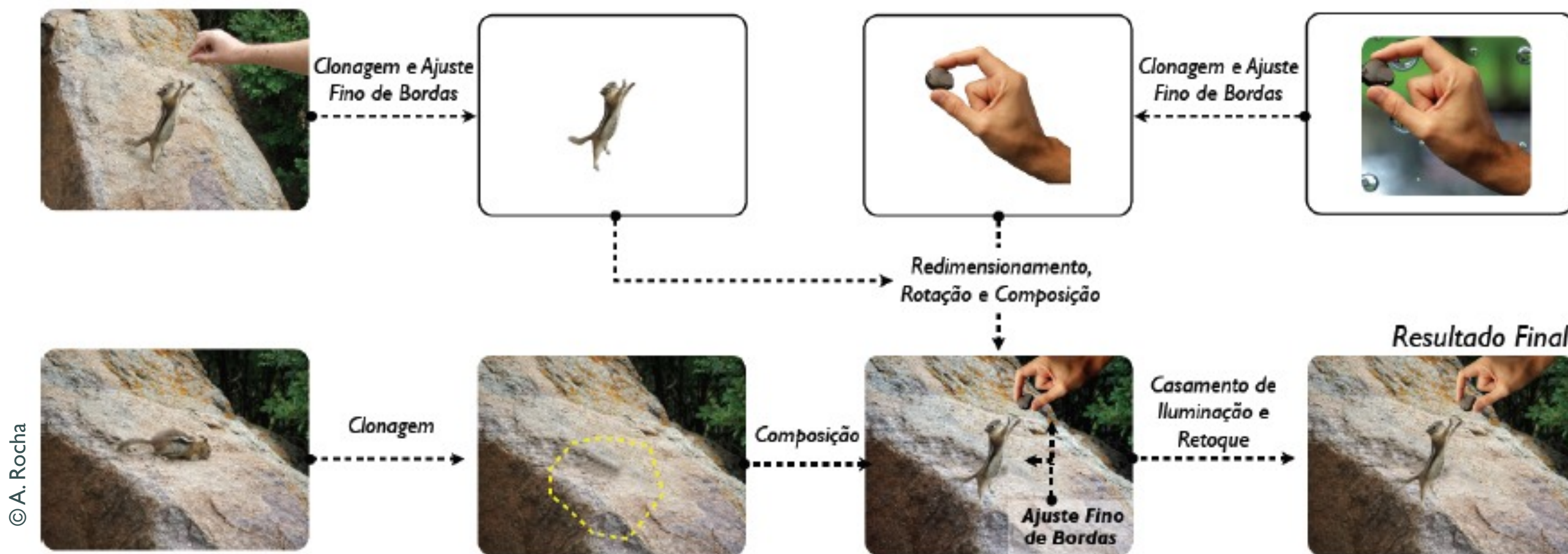
- ▶ Facilidade na manipulação de imagens digitais por meio de ferramentas de *software*:
 - Adobe Photoshop
 - GIMP
 - etc.
- ▶ Tipos de manipulação:
 - Melhoria da Imagem
 - Manipulação Intencional

Técnicas de Manipulação

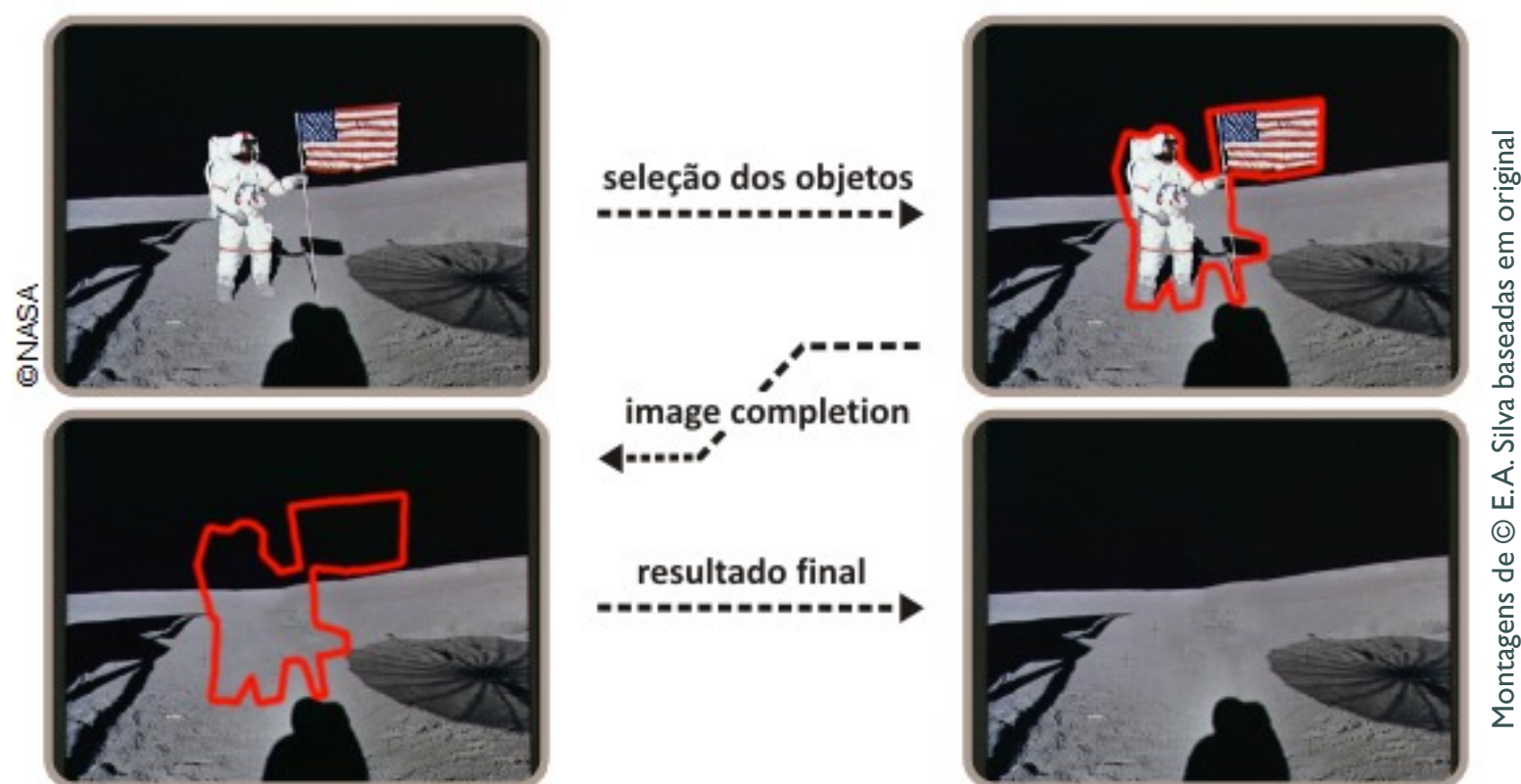
Técnicas de Manipulação

- ▶ Composição
- ▶ Ajuste Fino de Bordas
- ▶ Casamento de Padrões de Iluminação
- ▶ Realce de Nitidez
- ▶ Geração em Computador
- ▶ Cópia-colagem
- ▶ Retoque e Conciliação
- ▶ Técnicas modernas de manipulação

Técnicas de Manipulação



Técnicas de Manipulação



Técnicas de Manipulação

- ▶ Composição
- ▶ Ajuste Fino de Bordas
- ▶ Casamento de Padrões de Iluminação
- ▶ Realce de Nitidez
- ▶ Geração em Computador
- ▶ Cópia-colagem (Clonagem)
- ▶ Retoque e Conciliação
- ▶ Técnicas modernas de manipulação

Cópia-colagem

Cópia-colagem

- ▶ Consiste na cópia de segmentos da imagem e colagem destes em outras regiões da mesma imagem
- ▶ Objetiva mascarar detalhes da imagem
- ▶ Segmentos são, geralmente, texturas como:
 - Folhagem
 - Areia
 - Céu azul e nuvens
 - etc.

Cópia-colagem

© J.A. Finnis



Montagem de © E.A. Silva baseada em original

Cópia-colagem

© M. G. Jackson



Montagem de © E.A. Silva baseada em original



Cópia-colagem



Cópia-colagem

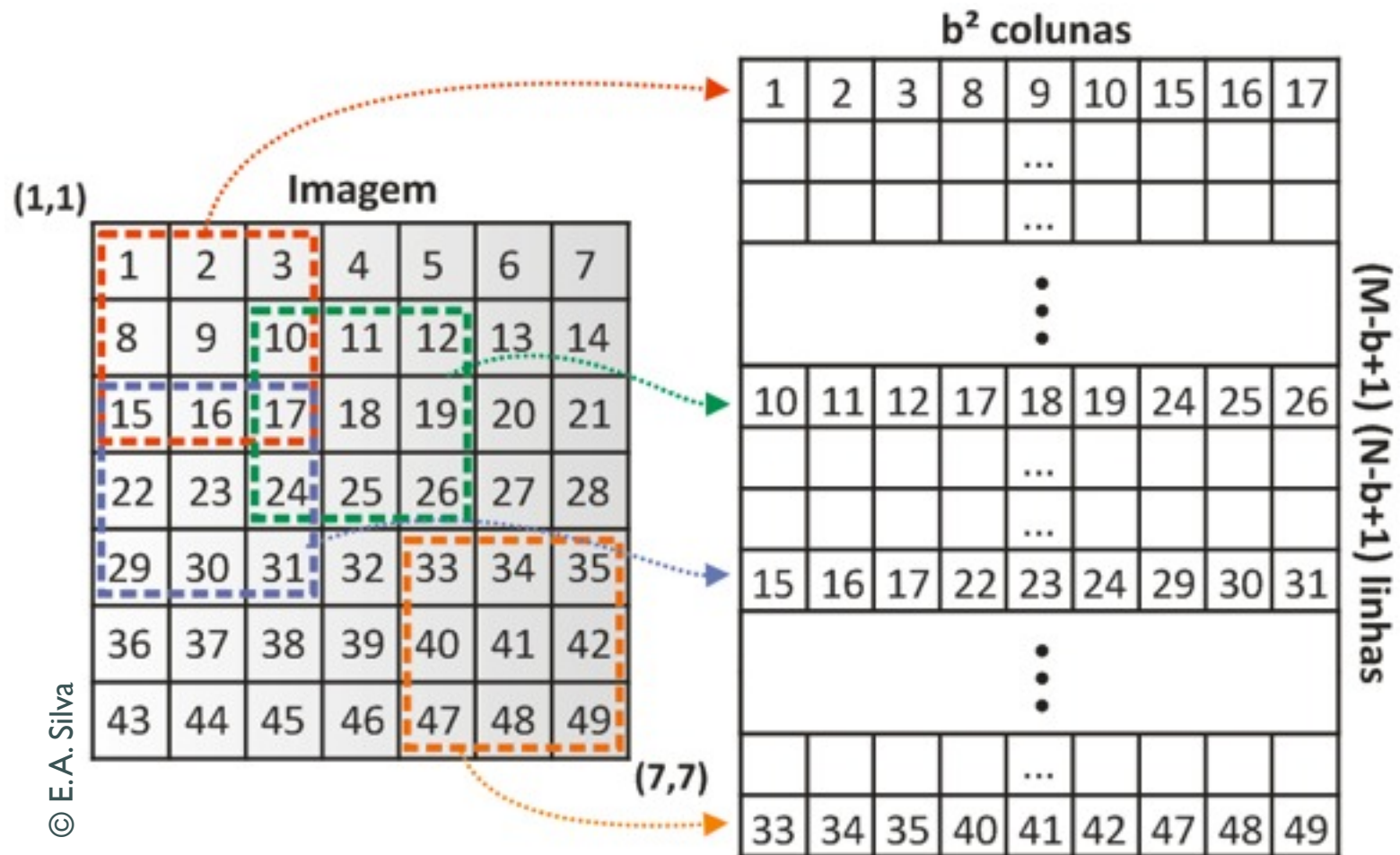
- ▶ Na prática, são usadas várias estratégias para enganar nossa visão e tornar a imagem convincente:
 - Suavização de bordas
 - Operações geométricas
 - Retoque e conciliação
 - Casamento de iluminação
 - Compressão JPEG
 - Ruídos aditivos

Detecção de Cópia-colagem

Detecção de Cópia-colagem

- ▶ 1. Comparar cada par de pixels da imagem
 - Complexidade exponencial no número de pixels
 - Devemos considerar a semelhança entre segmentos
- ▶ 2. Comparar blocos de pixels
 - Diminui a quantidade de computações efetuadas
 - Flexibilidade para encontrar regiões similares
 - Método mais utilizado, com muitas variações e adaptações

Detecção de Cópia-colagem



Estado da Arte

Estado da Arte

- ▶ Detecção de Manipulações de Cópia-colagem em Imagens Digitais [Fridrich et al. 2003]
- ▶ Expondo Manipulações Digitais por meio de Detecção de Regiões Duplicadas em Imagens [Popescu e Farid 2004]
- ▶ Detectando Duplicação de Regiões de Imagens Utilizando Características SIFT [Pan e Lyu 2010]

Estado da Arte

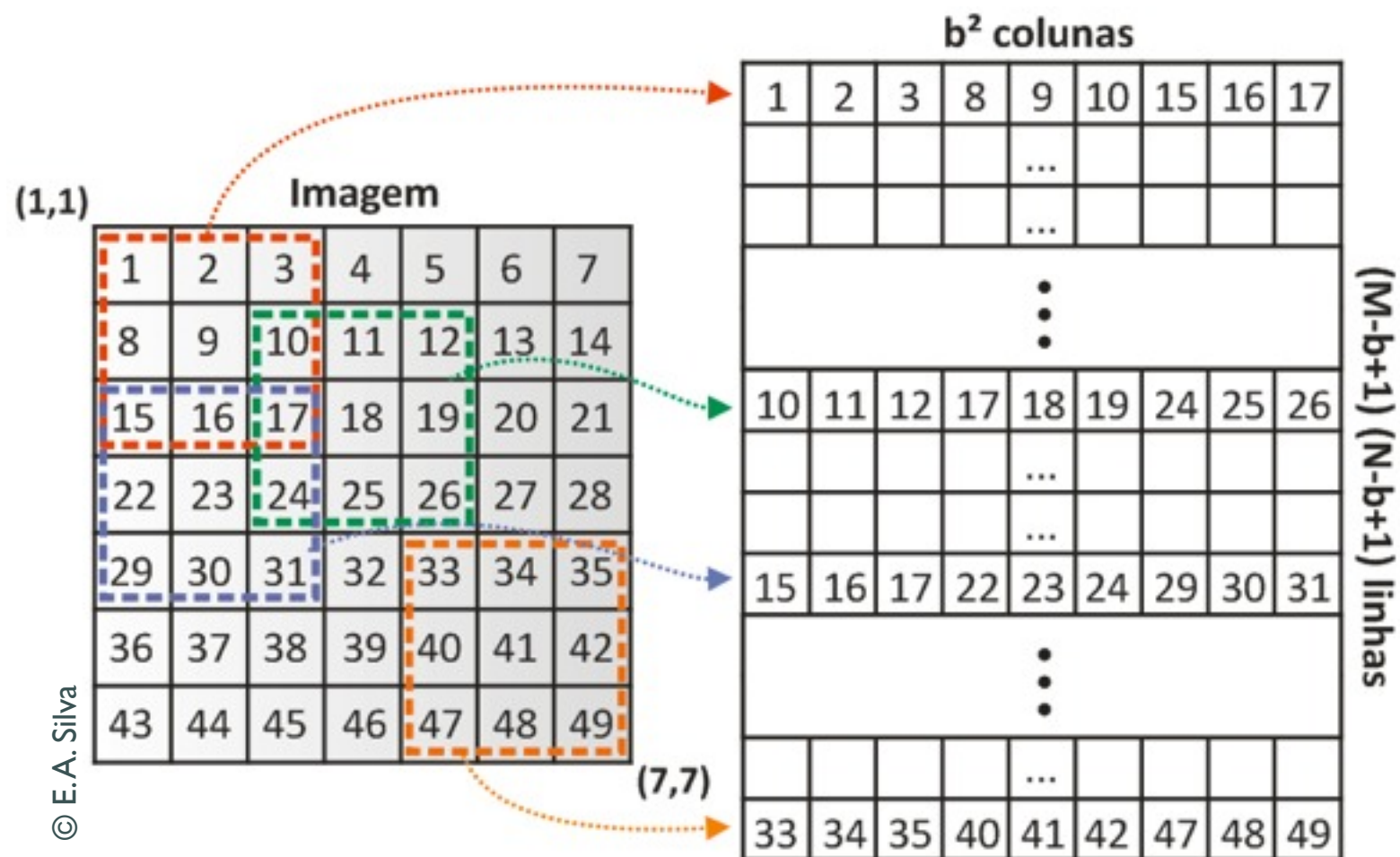
- ▶ Detecção de Manipulações de Cópia-colagem em Imagens Digitais [Fridrich et al. 2003]
- ▶ Expondo Manipulações Digitais por meio de Detecção de Regiões Duplicadas em Imagens [Popescu e Farid 2004]
- ▶ Detectando Duplicação de Regiões de Imagens Utilizando Características SIFT [Pan e Lyu 2010]

Detecção de Manipulações de Cópia-colagem em Imagens Digitais

[Fridrich et al. 2003]

- ▶ Propõem 3 métodos para detecção:
 - Busca Exaustiva
 - Autocorrelação
 - Casamento de Blocos
 - ▶ Casamento Exato
 - ▶ Casamento Aproximado (Robusto)

Casamento Exato



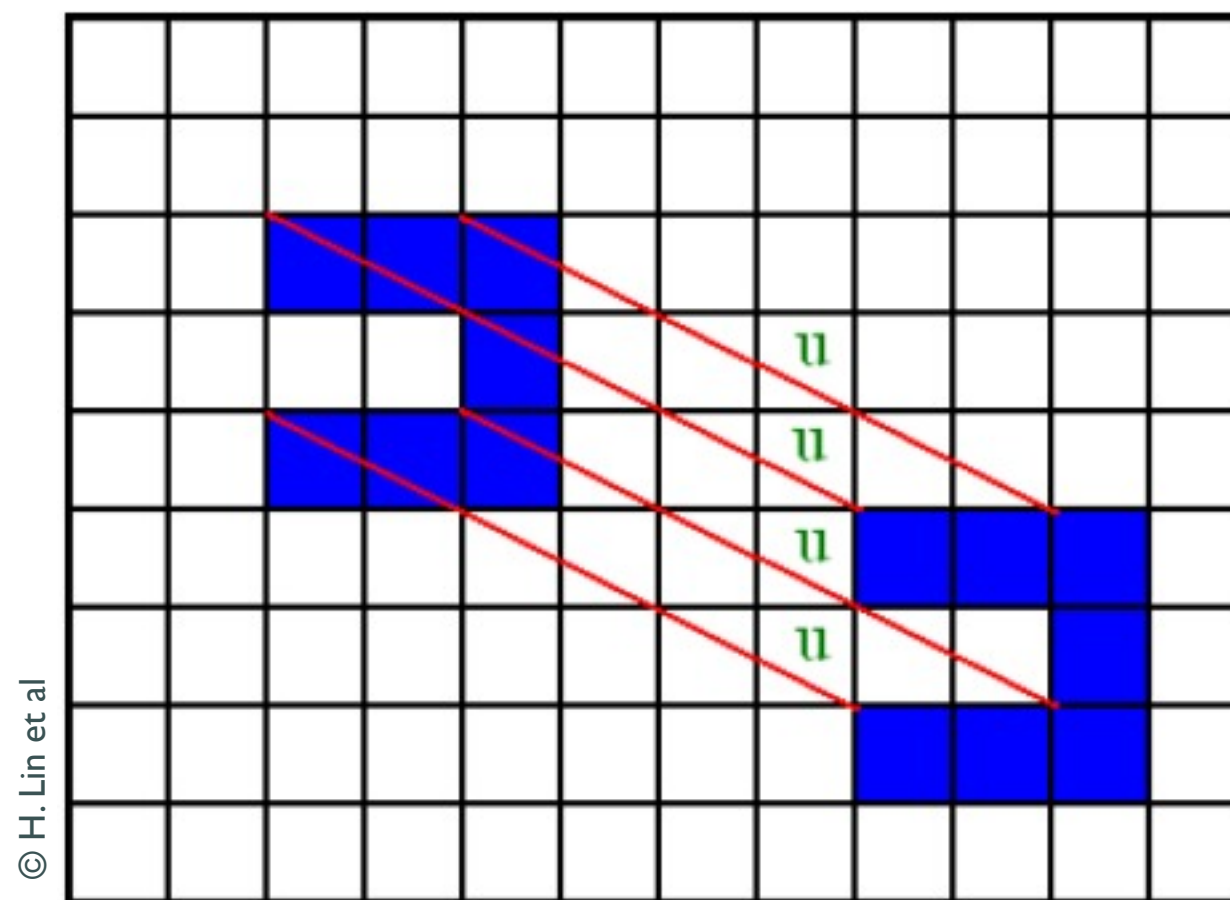
Casamento Aproximado

- ▶ Detecção de regiões semelhantes na imagem
- ▶ Uso da Transformada Discreta do Cosseno (DCT)
- ▶ DCT é aplicada em cada bloco previamente à ordenação lexicográfica
- ▶ Objetiva diminuir a variabilidade ocasionada por operações de pós-processamento

Casamento Aproximado

- ▶ A posição de cada bloco é dada pela coordenada do *pixel* do canto superior esquerdo
- ▶ Para cada bloco idêntico, o algoritmo calcula o vetor de deslocamento em valores absolutos
- ▶ Utilização de um contador para esses vetores
- ▶ Um limiar T determina o tamanho mínimo da região clonada que o algoritmo pode identificar

Casamento Aproximado



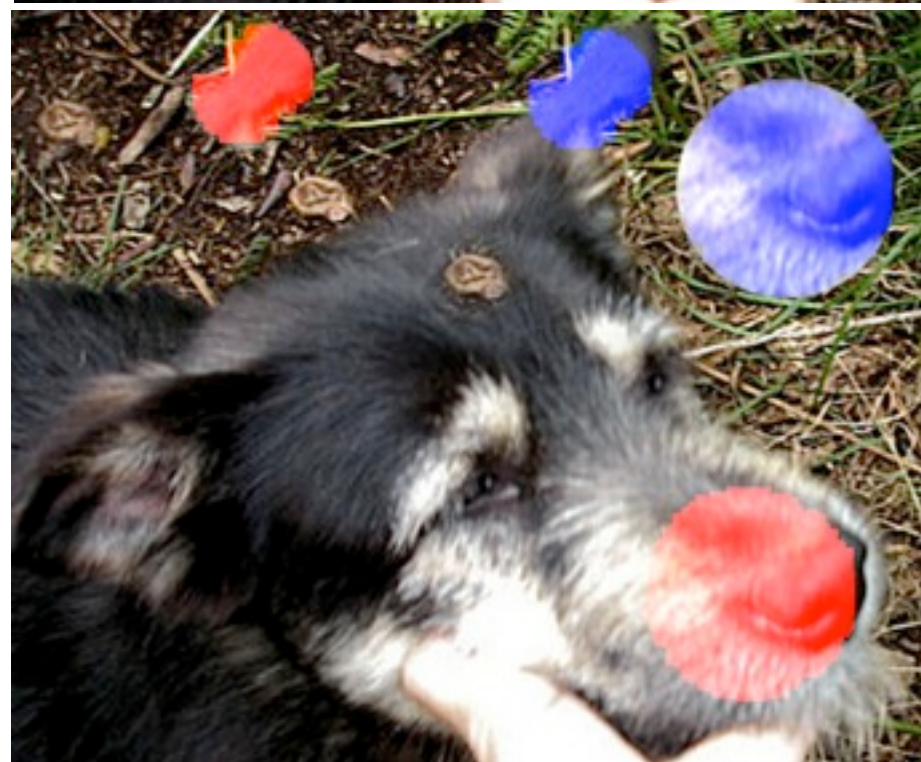
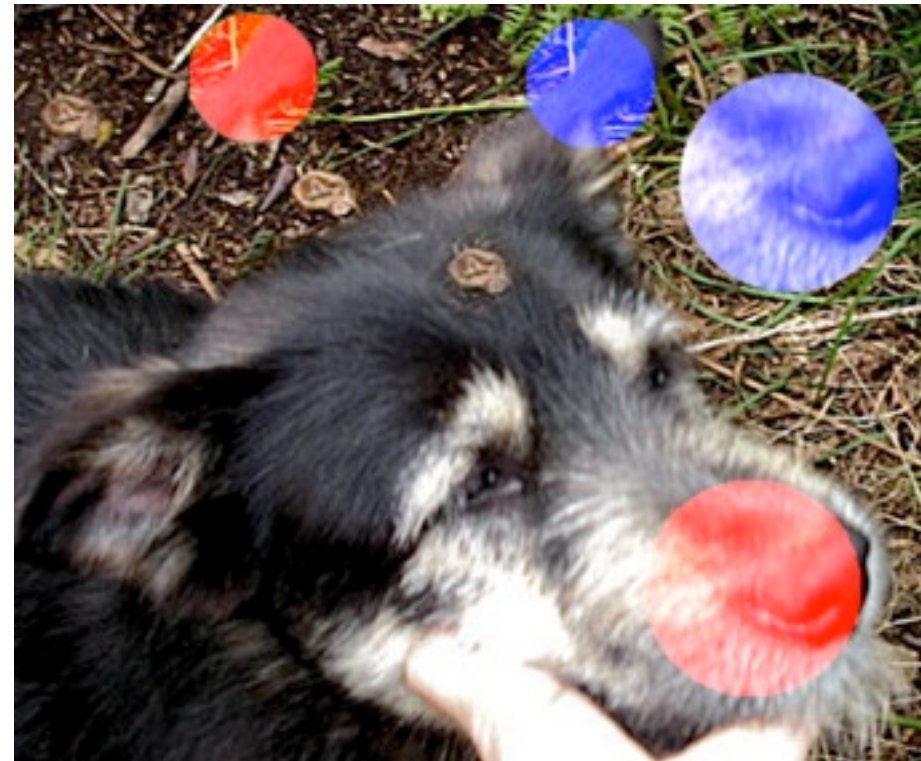
Casamento Aproximado

- ▶ Blocos de tamanho fixo 16 x 16
 - Blocos maiores possuem maior variabilidade nos coeficientes DCT
- ▶ Um fator Q é usado na quantização dos coeficientes DCT
- ▶ Matriz de quantização usada:

$$Q_{16} = \begin{pmatrix} Q'_8 & 2.5q_{18}I \\ 2.5q_{81}I & 2.5q_{88}I \end{pmatrix}, \text{ where } Q'_8 = \begin{pmatrix} 2q_{00} & 2.5q_{12} & \dots & 2.5q_{18} \\ 2.5q_{21} & 2.5q_{22} & \dots & 2.5q_{28} \\ \dots & \dots & \dots & \dots \\ 2.5q_{81} & 2.5q_{82} & \dots & 2.5q_{88} \end{pmatrix}$$

© Fridrich et al.

Experimentos e Resultados



Experimentos e Resultados



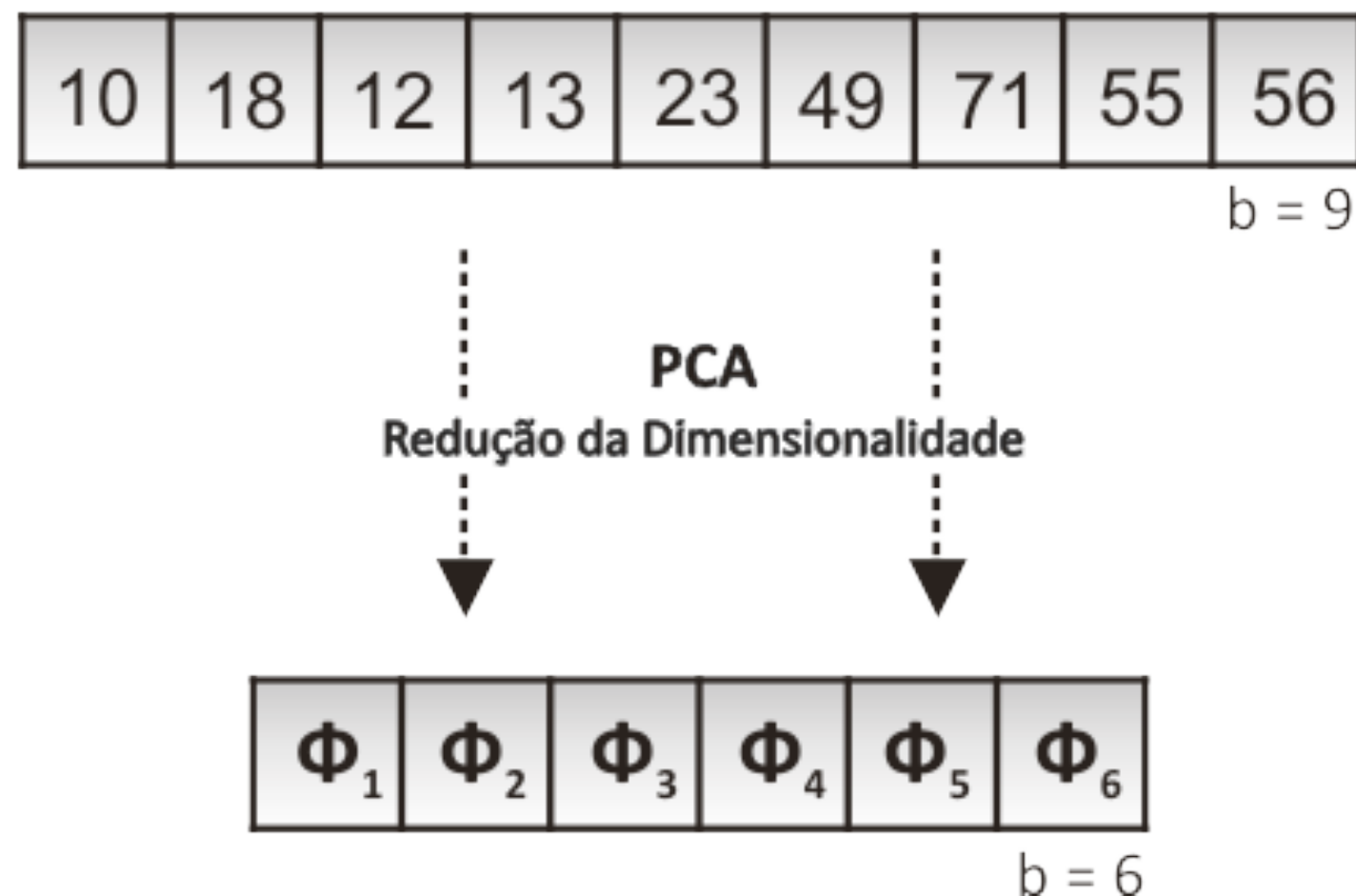
Estado da Arte

- ▶ Detecção de Manipulações de Cópia-colagem em Imagens Digitais [Fridrich et al. 2003]
- ▶ Expondo Manipulações Digitais por meio de Detecção de Regiões Duplicadas em Imagens [Popescu e Farid 2004]
- ▶ Detectando Duplicação de Regiões de Imagens Utilizando Características SIFT [Pan e Lyu 2010]

Expondo Manipulações Digitais por meio de Detecção de Regiões Duplicadas em Imagens [Popescu e Farid 2004]

- ▶ Propõem a utilização de Análise de Componentes Principais para:
 - Reduzir a dimensionalidade dos blocos
 - Eliminar as variações menores dos blocos causadas pelo pós-processamento
- ▶ Após a aplicação de PCA, cada bloco é quantizado utilizando um inteiro Q (número de bins)

Descrição do Método



© E.A. Silva

Descrição do Método

- ▶ Em seguida, cada par de linhas dentro de uma certa distância d (na matriz) é analisado
- ▶ O vetor de deslocamento (*offset*) entre os pares é calculado

$$\begin{aligned} (x_i - x_j, y_i - y_j) & \text{ if } x_i - x_j > 0 \\ (x_j - x_i, y_i - y_j) & \text{ if } x_i - x_j < 0 \\ (0, |y_i - y_j|) & \text{ if } x_i = x_j \end{aligned}$$

© Popescu e Farid

Descrição do Método

- ▶ Os vetores com maior ocorrência determinam as regiões duplicadas
- ▶ Com isso, uma região poderá ser constituída de vários blocos pequenos, todos com o mesmo *offset*
- ▶ Ainda, vetores com magnitude menor que um limiar especificado são descartados

Descrição do Método

- ▶ Para imagens coloridas, há duas soluções:
 - Executar o método para cada canal de cor e juntar o resultado final
 - Aplicar PCA em blocos de tamanho $3b$

Experimentos e Resultados

► Parâmetros especificados:

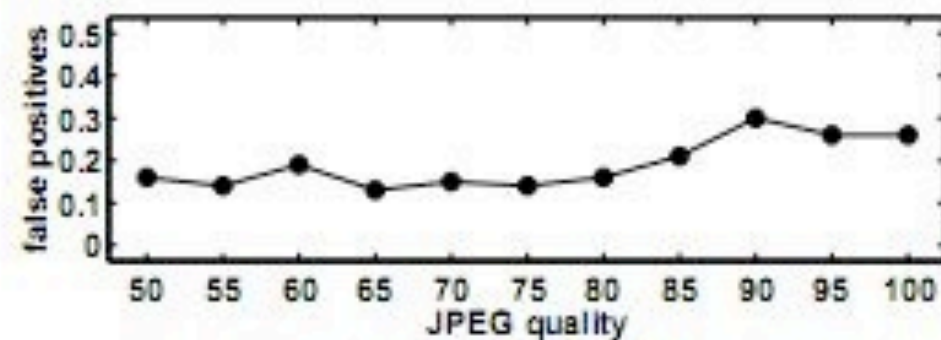
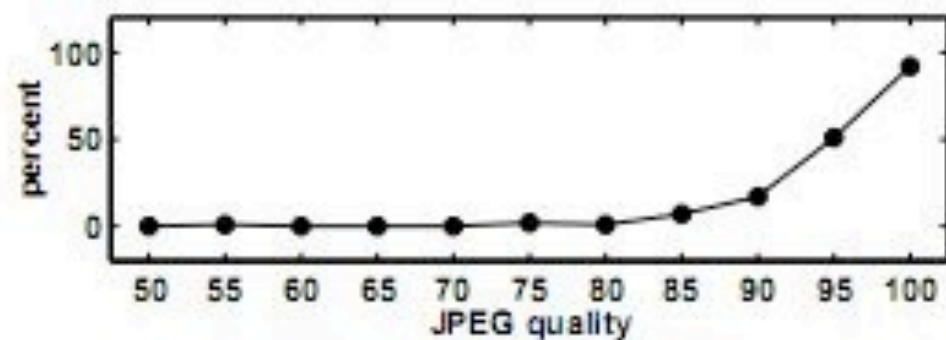
- $b = 64$ (tamanho do bloco)
- $\epsilon = 0.01$ (fração da variância ignorada)
- $Q = 256$ (fator de quantização)
- $Nn = 100$ (número de linhas vizinhas na matriz)
- $Nf = 128$ (frequência mínima de *offset*)
- $Nd = 16$ (magnitudo mínima do *offset*)

$$1 - \epsilon = \frac{\sum_{i=1}^{N_t} \lambda_i}{\sum_{i=1}^b \lambda_i}$$

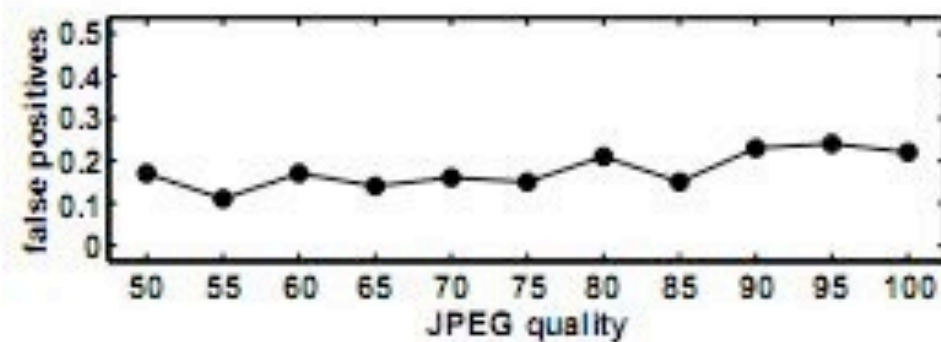
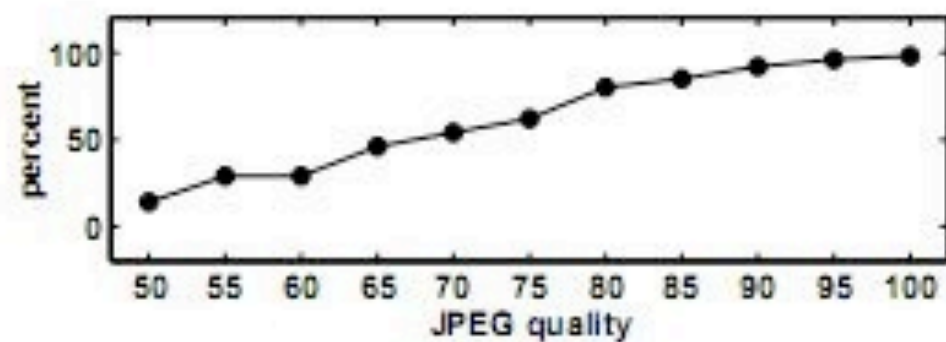
Experimentos e Resultados

- ▶ Testes com 100 imagens coloridas de tamanho 512 x 512 *pixels*
- ▶ Imagens comprimidas em JPEG (50% a 100%)
- ▶ Ruídos aditivos Gaussianos (24dB a 40dB)
- ▶ Para blocos de tamanho 64, a redução obtida com PCA foi de 50%

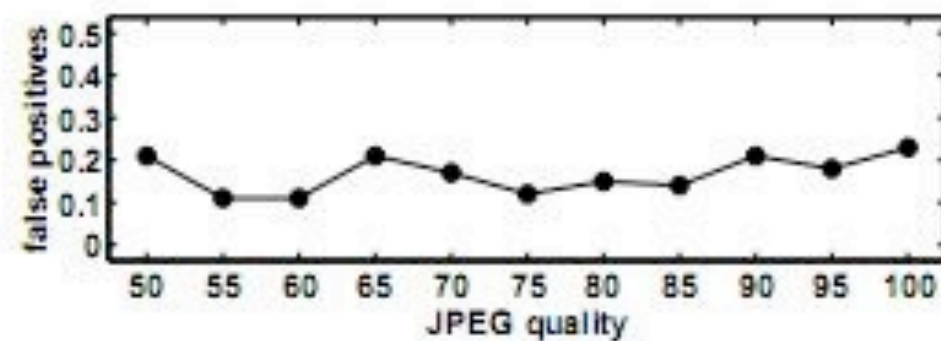
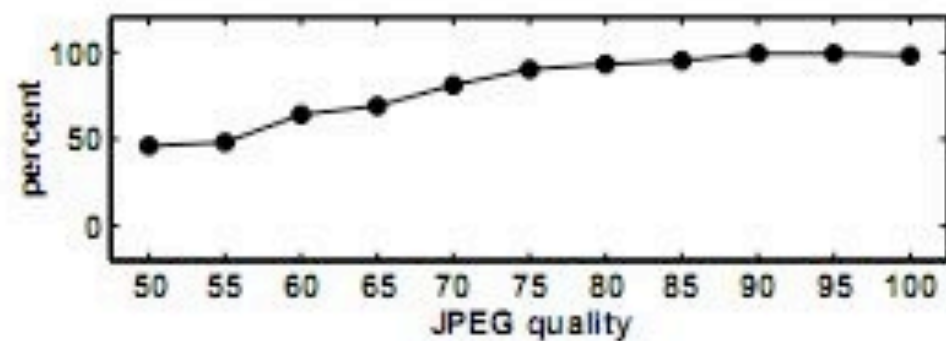
32×32



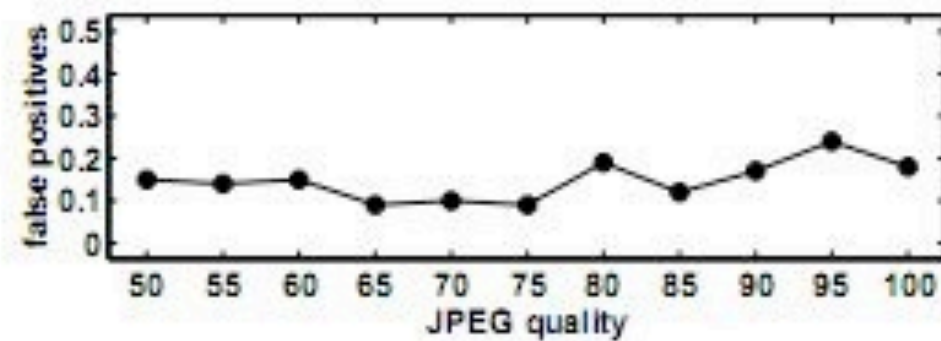
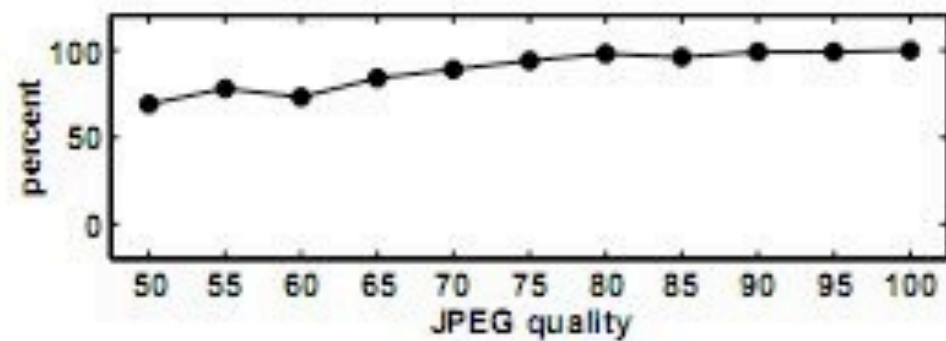
64×64



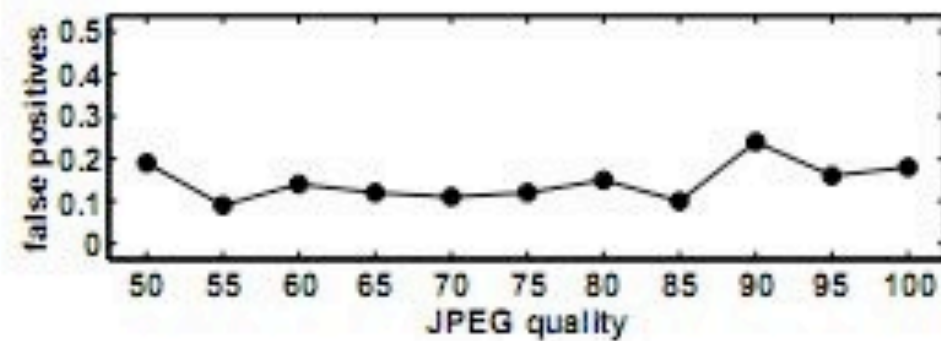
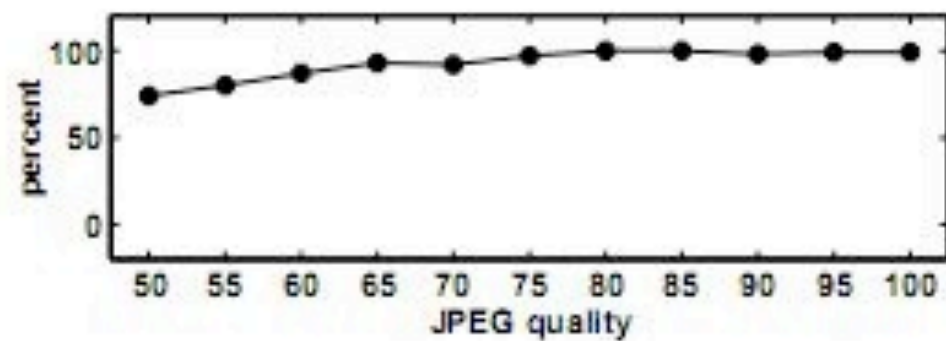
96×96



128×128



160×160





50



60

70



80

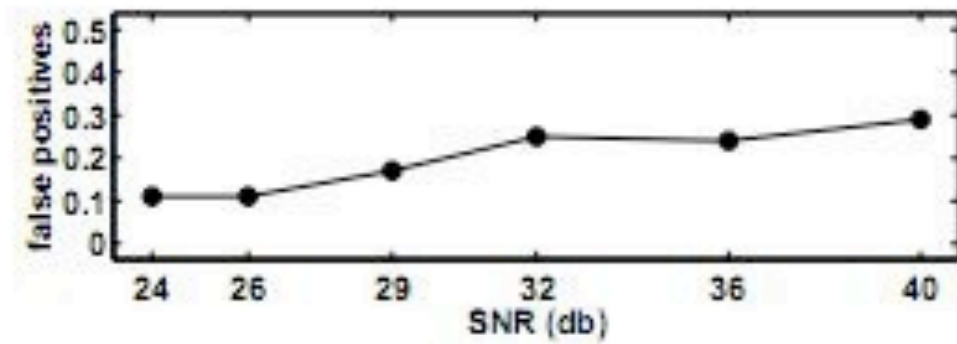
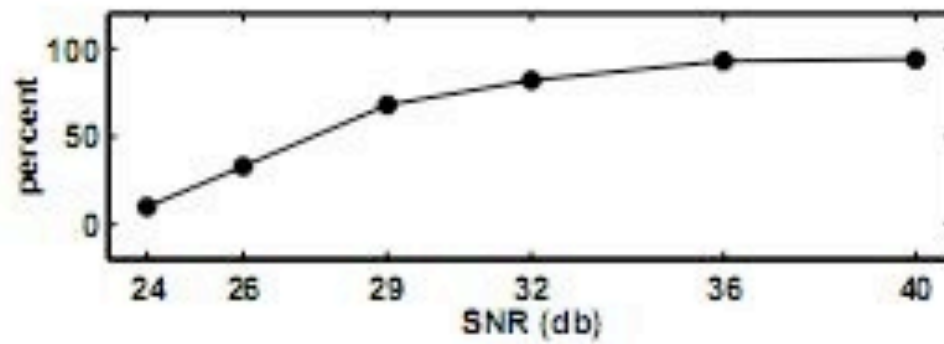
90

100

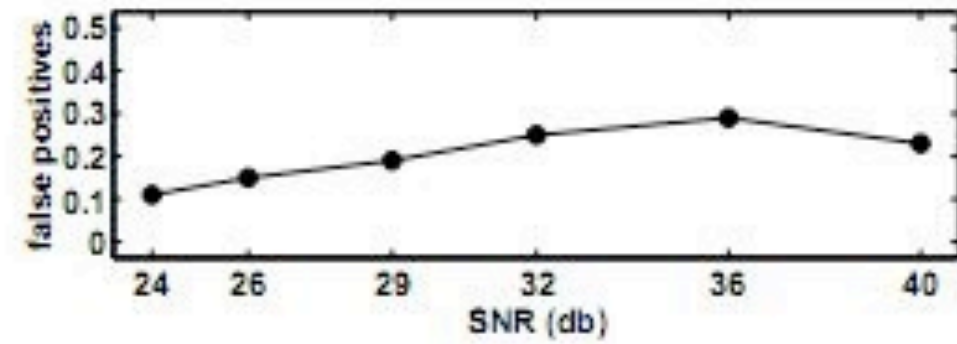
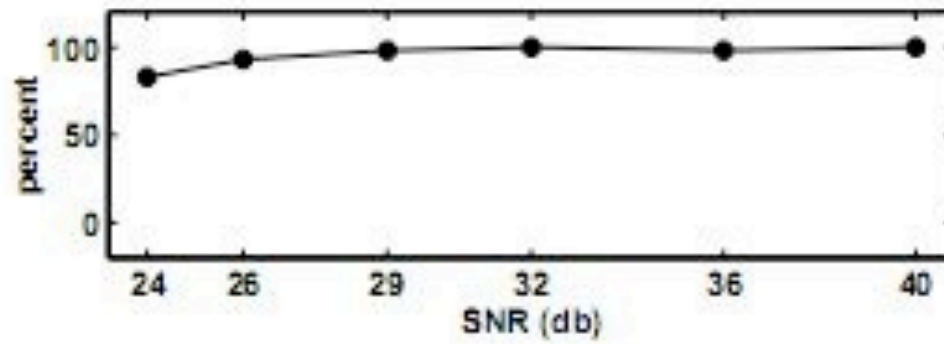


© Popescu e Farid

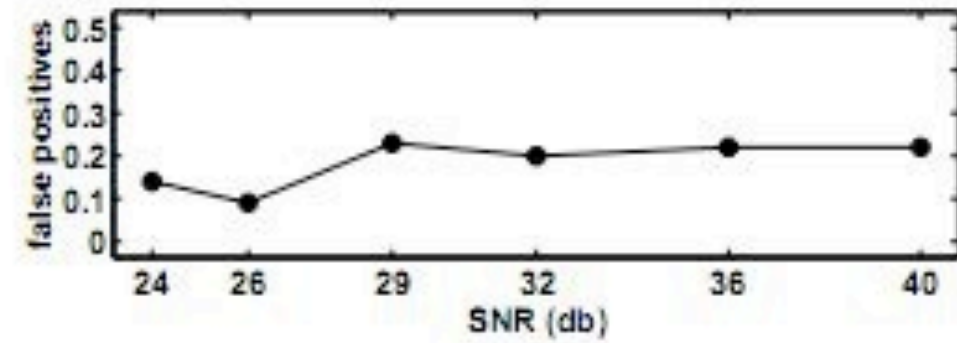
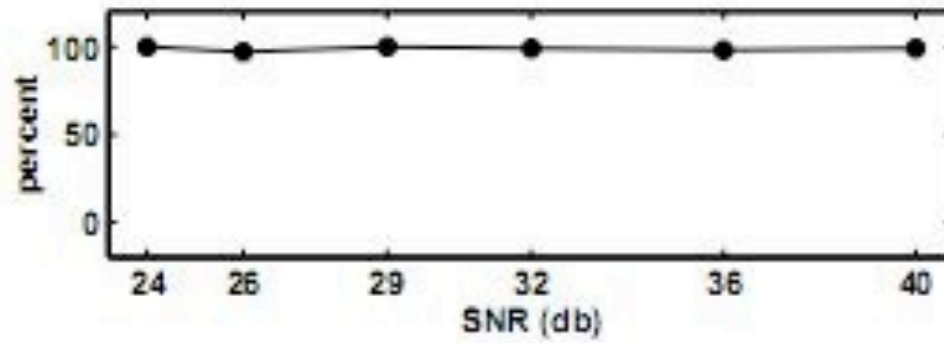
32×32



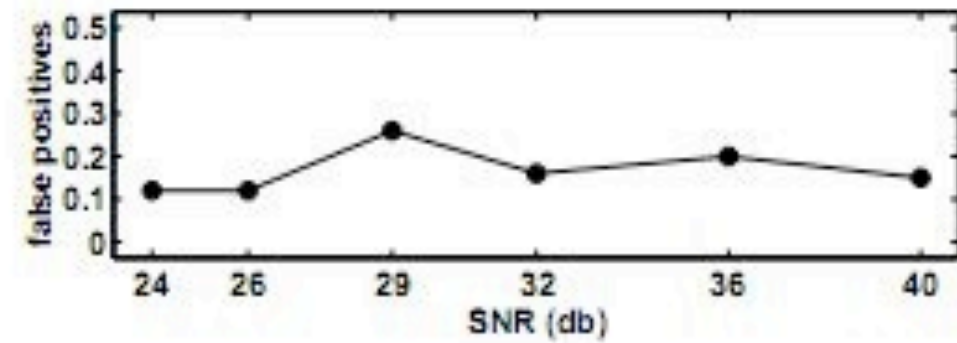
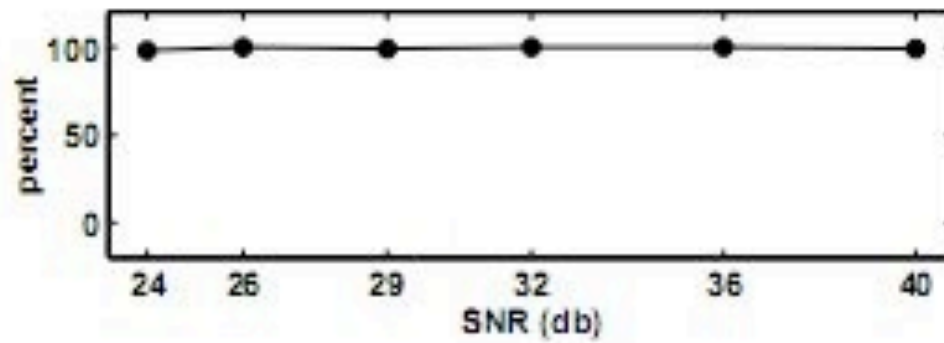
64×64



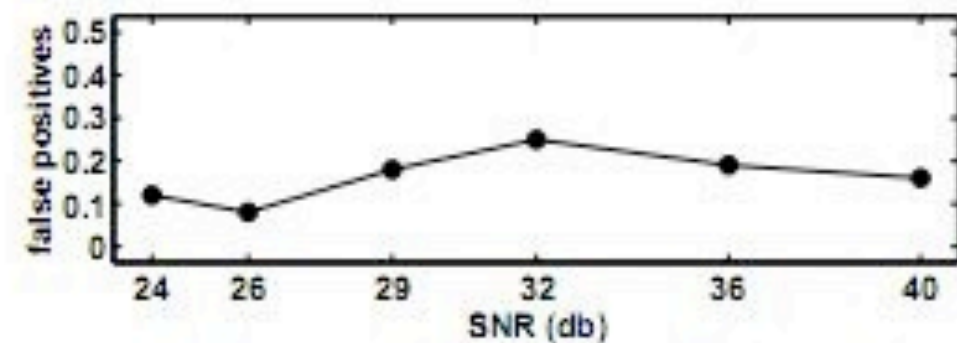
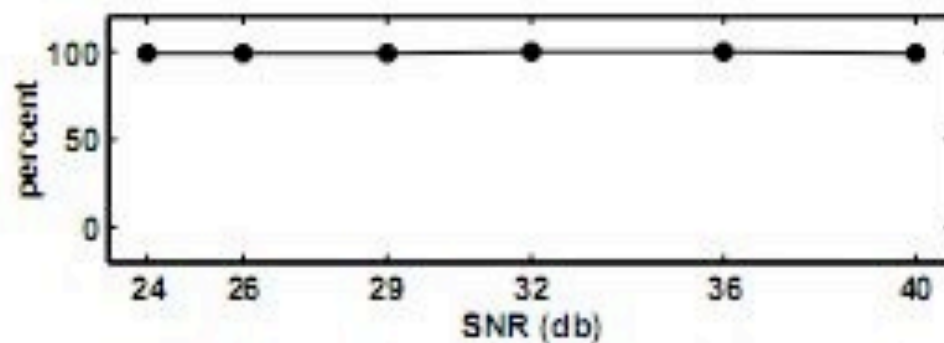
96×96

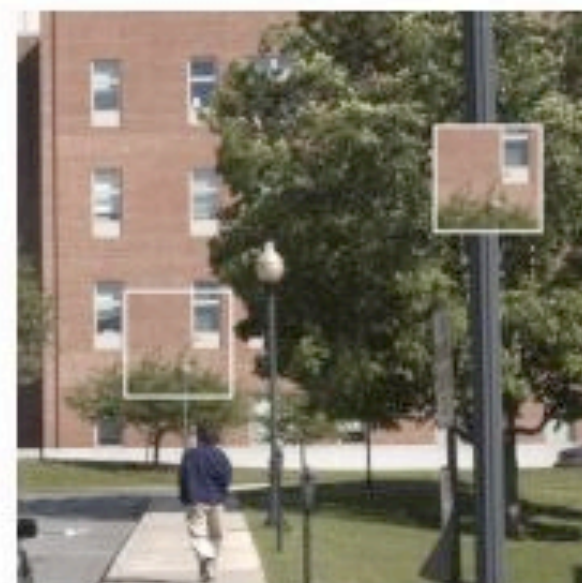


128×128

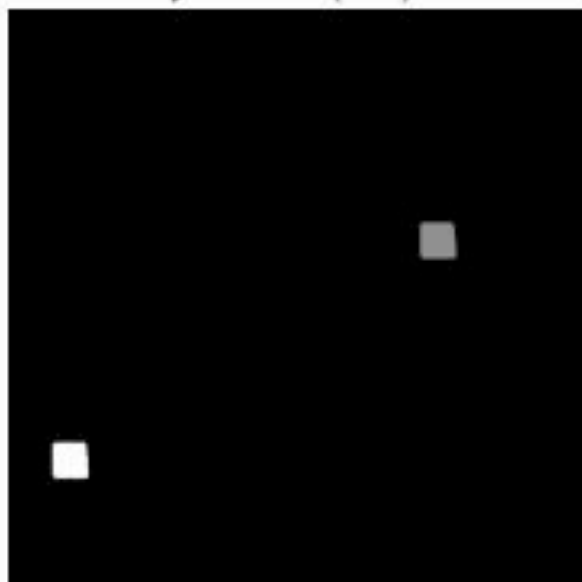


160×160





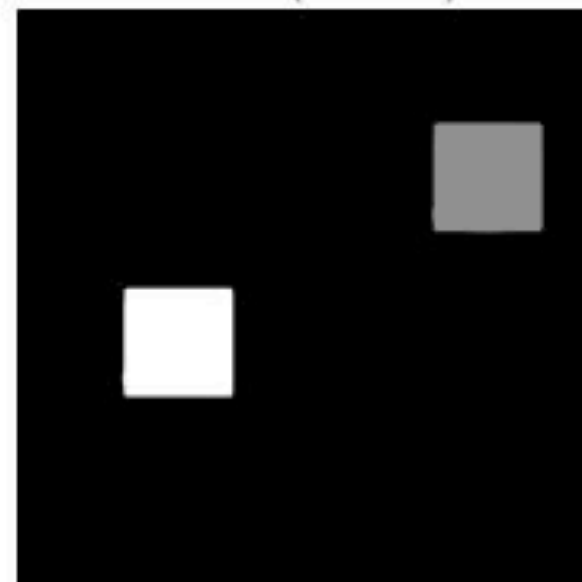
JPEG (85)



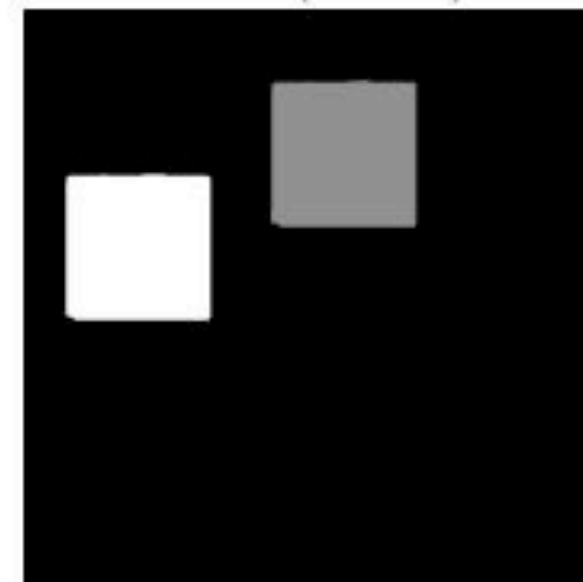
JPEG (65)



SNR (36 db)



SNR (29 db)



© Popescu e Farid

Estado da Arte

- ▶ Detecção de Manipulações de Cópia-colagem em Imagens Digitais [Fridrich et al. 2003]
- ▶ Expondo Manipulações Digitais por meio de Detecção de Regiões Duplicadas em Imagens [Popescu e Farid 2004]
- ▶ Detectando Duplicação de Regiões de Imagens Utilizando Características SIFT [Pan e Lyu 2010]

Detectando Duplicação de Regiões de Imagens Utilizando Características SIFT

[Pan e Lyu 2010]

- ▶ As abordagens anteriores (baseadas em blocos de *pixels*) não são sensíveis a operações geométricas nas regiões clonadas
- ▶ Pan e Lyu propõem um método baseado na correspondência de características SIFT da imagem

Scale Invariant Features Transform (SIFT)

- ▶ Encontra pontos-chave, que são características locais únicas da imagem
- ▶ Robusto a vários tipos de degradação:
 - Rotação
 - Escala
 - Mudança global de contraste
 - Variações de iluminação
 - Ruídos

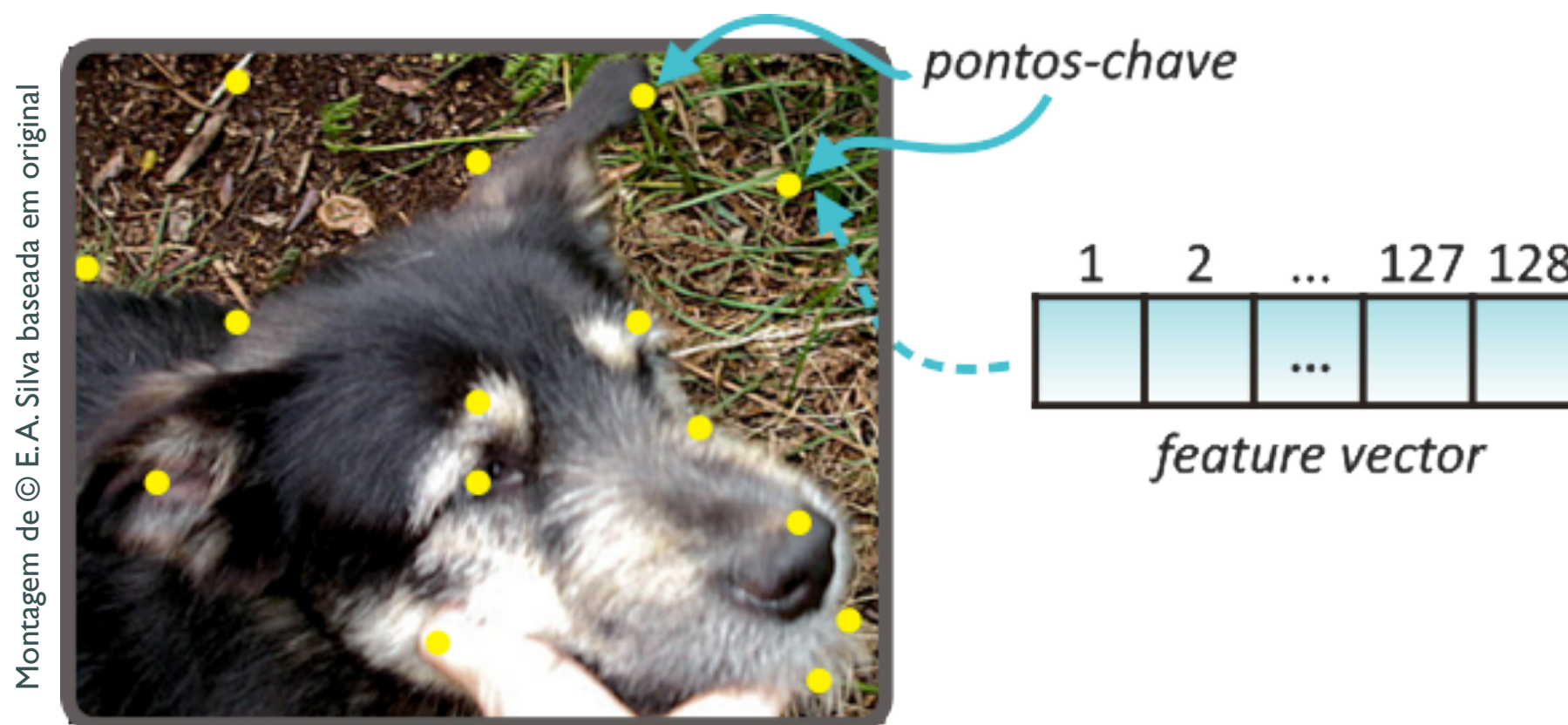
Descrição do Método

1. Coleta de características SIFT
 - Determinação dos pontos-chave
 - Determinação dos vetores de características para cada ponto-chave
2. Casamento de características e poda de correspondências SIFT
3. Estimação das transformações geométricas efetuadas
4. Correlação entre as regiões empregando as estimativas calculadas

I. Coleta de características SIFT

- ▶ Utiliza-se o algoritmo SIFT para coleta dos pontos e determinação dos vetores de características de cada um deles
- ▶ Número de pontos é bem menor que o número de *pixels*
- ▶ Vetores possuem 128 dimensões

I. Coleta de características SIFT



2. Casamento de características e poda de correspondências SIFT

- ▶ A imagem é dividida em vários blocos de *pixels* sem sobreposição
- ▶ Para cada par de pontos-chave de blocos distintos calcula-se a distância euclidiana entre os vetores dos dois pontos
 - Isto é, encontra-se o vizinho mais próximo ao ponto sendo analisado

2. Casamento de características e poda de correspondências SIFT

- ▶ Na poda, correspondências de pontos que casam com múltiplos pontos na imagem são removidas
 - Podem ser causadas por ruídos
- ▶ O número de correspondências corretas para cada bloco é armazenado
- ▶ O bloco com o maior número de acertos e o bloco contendo sua correspondência servirão como base para estimar a transformação entre regiões.

3. Estimação das transformações geométricas efetuadas

► Escala:

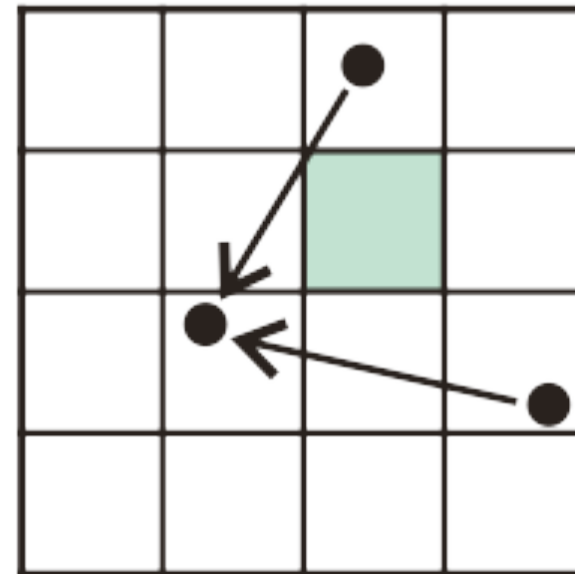
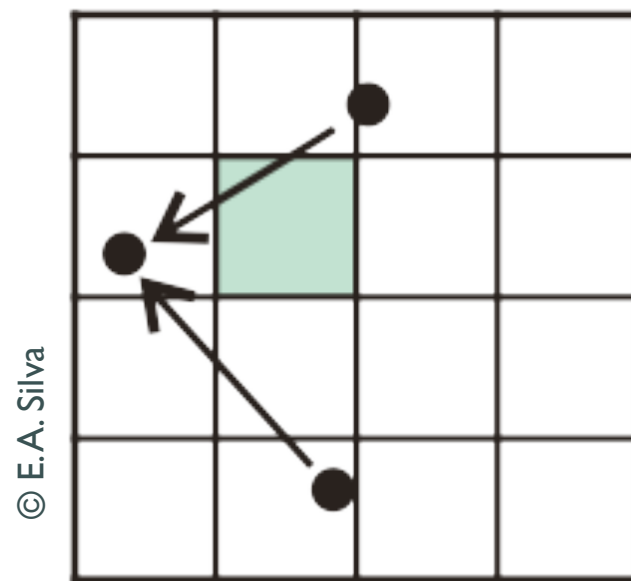
- Dois pares de pontos correspondentes são obtidos (A, B) e (C, D)
- Calculamos a razão: $\|A - B\| / \|C - D\|$ para todos os pares pertencentes aos blocos
- A razão com maior número de ocorrências é uma estimativa para o fator de escala

3. Estimação das transformações geométricas efetuadas

► Rotação:

- Três pontos não-colineares e suas correspondências na região com maiores taxas de correspondência
- É definido um sistema de coordenadas para cada região (original e duplicada)
- Transformamos os pontos de um sistema para o outro

3. Estimação das transformações geométricas efetuadas



4. Correlação entre as regiões

- ▶ A imagem é segmentada em blocos de 4 x 4 *pixels* com sobreposição
- ▶ Efetuamos a transformação na região duplicada
- ▶ Computamos a correlação entre os blocos que fazem parte do contorno das regiões
- ▶ O valor da correlação é comparado a um limiar, que define se o bloco faz parte do contorno da região duplicada.

Experimentos e Resultados

- ▶ Testes com 100 imagens coloridas de tamanho 512 x 512 *pixels*
- ▶ Regiões duplicadas de tamanho 64 x 64 e 96 x 96
- ▶ Compressão JPEG (60% a 100%)
- ▶ Ruídos aditivos (20dB a 40 dB)

Experimentos e Resultados

JPEG	Q = 60	Q = 70	Q = 80	Q = 90	Q = 100
64 × 64	86.23/1.74	85.04/1.66	89.79/1.68	90.49/1.75	92.76/2.15
96 × 96	91.42/0.93	92.35/0.96	93.02/0.95	93.85/1.02	95.05/1.20
SNR	20 dB	25 dB	30 dB	35 dB	40 dB
64 × 64	89.76/1.78	92.06/2.03	92.43/2.08	92.55/2.07	92.61/2.13
96 × 96	92.84/1.02	94.24/1.13	94.62/1.18	94.70/1.18	94.78/1.19

©Pan e Lyu

Experimentos e Resultados

copy-move

scaling & copy-move

rotation & copy-move



$p_{\text{acc}} = 91.3\%$, $p_{\text{fpr}} = 0.3\%$

$p_{\text{acc}} = 93.4\%$, $p_{\text{fpr}} = 1.2\%$

$p_{\text{acc}} = 81.4\%$, $p_{\text{fpr}} = 1.9\%$

©Pan e Lyu

Considerações Finais

Considerações Finais

- ▶ Um dos grandes desafios da detecção de cópia-colagem reside na identificação em cenários de rotação e escala
- ▶ O método de Pan e Lyu (2010) aparenta ser robusto a essas operações, mas os autores não mencionam os ângulos e fatores de escala
- ▶ As três abordagens não mostram experimentos com operações de pós-processamento combinadas
- ▶ Outro desafio é a detecção de cópia-colagem realizada por meio de *image completion*

Referências

Referências

1. [Fridrich et al. 2003] Jessica Fridrich, David Soukal, and Jan Lukas. **Detection of copy-move forgery in digital images**. In Digital Forensic Research Workshop (DFRWS), Cleveland, USA, 2003.
2. [Pan e Lyu 2010] Xunyu Pan and Siwei Lyu. **Detecting image region duplication using sift features**. In IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1706-1709, 2010.
3. [Popescu e Farid 2004] Alin C. Popescu and Hany Farid. **Exposing digital forgeries by detecting duplicated image regions**. Technical Report TR 2004-515, Department of Computer Science – Dartmouth College, Hanover, USA, 2004.

Obrigado!
