Além do óbvio: a análise forense de imagens e a investigação do conteúdo implícito e explícito de fotografias digitais

Tiago Carvalho[1], **Ewerton Silva**[1], **Filipe Oliveira Costa**[1], **Anselmo Ferreira**[1], **Anderson Rocha**[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Campinas – SP – Brazil

{tjose, ewerton, filipe, ra023169, anderson.rocha}@ic.unicamp.br

Abstract. *Nowadays the use of tools for manipulating images and videos is increasingly common. Such tools facilitate the task of creating manipulations and deceiving the perception of observers on the semantics of these documents. Although there are image manipulations considered innocent (*e.g.*, correction of brightness), there are those considered malicious, such as the copy-paste and composition operations. In this paper, we discuss the main challenges present in the forensic authentication of digital documents such as images and videos as well as our most recent contributions in this context.*

Resumo. *Atualmente, torna-se cada vez mais comum a utilização de ferramentas para manipulação de imagens e vídeos. Tais ferramentas facilitam a criação de alterações em documentos, enganando a percepção de observadores quanto à semântica desses documentos. Apesar de existirem alterações consideradas inocentes (como uma correção de brilho), existem aquelas consideradas maliciosas como, por exemplo, operações de cópia e colagem e composição. Neste trabalho, discutimos os principais desafios tratados no cenário forense de autenticação de documentos digitais como imagens e vídeos, bem como as nossas mais recentes contribuições nesse contexto.*

1. Introdução

Atualmente, uma das formas mais comuns de comunicação presente na vida das pessoas é a comunicação visual, a qual retrata fatos e imortaliza momentos utilizando vídeos e imagens como meios. Graças aos avanços da tecnologia, produzir um vídeo ou uma foto torna-se mais fácil a cada dia, dado que o número de dispositivos para esse tipo de tarefa disponíveis no mercado vem aumentando significativamente e se tornando cada vez mais sofisticados.

Além do aumento no número de dispositivos para captura, o número de ferramentas para manipulação de imagens e vídeos também vem crescendo. Ferramentas de software, tais como o Adobe Photoshop, GIMP e outros, popularizam, cada vez mais, técnicas de manipulação de imagens antes empregadas apenas por pessoas com grande e conhecimento especializado. Tais ferramentas têm se tornado mais robustas e de fácil manipulação, possibilitando que usuários com um conhecimento mínimo sejam capazes de fazer grandes adulterações em imagens. Estas adulterações podem ir muito além de um simples ajuste de brilho ou a correção de pequenos defeitos na pele. Manipulações de imagens visando enganar o observador por meio da introdução de pessoas ou artefatos,

de modo a criar momentos que nunca ocorreram, são hoje uma prática comum em praticamente todos os meios de comunicação.

A falsificação de imagens, de modo a representar um momento histórico que nunca existiu é quase tão antiga quanto a arte da fotografia em si. Pouco depois que o francês Nicéphore Niepce criou a primeira fotografia em 1814, já apareciam as primeiras fotografias adulteradas [Rocha et al. 2011]. Stalin, Mao, Hitler, Mussolini, Castro e Brezhnev possuíam fotografias manipuladas (fosse para criar uma aparência mais heroica, para apagar inimigos ou remover garrafas de cerveja) [Rocha et al. 2011]. Tanto as alterações dos dias de Stalin, como a maior parte das adulterações anteriores à era digital, necessitavam de alta capacidade técnica e muitas horas (talvez dias) de trabalho em salas escuras de fotografia [Popescu 2004]. Hoje, mesmo um leigo com um computador pode facilmente produzir falsificações de difícil detecção [Farid 2008].

O poder da influência de uma foto fica evidente no trabalho publicado por Sacchi et al. [Sacchi et al. 2007], em que os autores exibem os resultados de um estudo que mostra como imagens de fotografias de eventos públicos ocorridos no passado afetam a memória das pessoas em relação a estes eventos.

Todo esse conjunto de fatores torna necessário um esforço cada vez maior da comunidade forense para criar métodos capazes de detectar quaisquer tipos de adulterações em imagens, gerando, desta forma, uma "disputa armamentista" entre os métodos forenses e os métodos de adulteração. Isso faz com que nossa principal motivação neste trabalho seja contribuir com a comunidade forense na exposição dos métodos por nós desenvolvidos para a identificação de adulterações em imagens.

Nesse trabalho, apresentamos as mais recentes contribuições desenvolvidas no laboratório *Reasoning for Complex Data* (RECOD) da Universidade Estadual de Campinas (UNICAMP), algumas das quais já foram publicadas em eventos e periódicos de renome internacional. Tais contribuições abrangem os problemas de atribuição de fonte para imagens, *linking* de dispositivos, identificação de cópia e colagem bem como identificação de composições.

2. Estado da arte em manipulações de imagens e vídeos

A Computação Forense é uma área da computação que vem crescendo e ganhando muita importância. Segundo Delp et al. [Delp et al. 2009], ela pode ser definida como *"a coleção de técnicas científicas para a preservação, coleção, validação, identificação, análise, interpretação, documentação e preservação de evidências digitais derivadas de fontes digitais com a proposta de facilitar ou promover a reconstrução de eventos, na maioria das vezes de natureza criminal"*.

Uma subárea que vem ganhando muito espaço dentro de Computação Forense é a Análise Forense de Documentos Digitais (AFD). Rocha e Goldenstein [Rocha and Goldenstein 2010] definem AFD como *o campo de pesquisas relacionado à análise de documentos digitais para verificação de sua autenticidade e integridade*. Ela tem produzido diversos tipos de métodos focados em objetivos específicos tais como a identificação de dispositivos de origem, identificação de criações sintéticas de imagens, identificação de adulterações em documentos, entre outras.

Rocha et al. [Rocha et al. 2011] indicam duas classes de adulterações podem ocor-

rer em uma imagem: as que melhoram a qualidade (ou Melhoria de Imagens) e aquelas que procuram enganar os observadores (Manipulação Intencional de Imagens) através da mudança da semântica da imagem. Apresentamos algumas técnicas desta última categoria e também técnicas de edição de imagens que facilitam esse processo a seguir:

- **Composição**. Consiste na união de partes de duas ou mais imagens numa única imagem. Esse tipo de falsificação pode afetar o comportamento das bordas, da iluminação, da compressão e também a do ruído do dispositivo de aquisição da imagem.
- **Ajuste fino de bordas (*feather edges*)**. Consiste no ajuste das fronteiras (bordas) dos objetos após uma operação de, por exemplo, composição, visando adequá-los às regiões da imagem onde estes se encontram e remover artefatos indesejáveis provenientes das imagens originais. Esse tipo de falsificação pode afetar o comportamento das bordas da imagem.
- **Casamento de padrões de iluminação (*light matching*)**. Auxilia no ajuste de iluminação de uma composição de forma a homogeneizar o aspecto da iluminação da imagem para torná-la verossímil. Esse tipo de falsificação pode afetar o comportamento da iluminação da imagem.
- **Realce de nitidez (*sharpening*)**. Consiste no realce de certos detalhes da imagem de forma a torná-los mais, ou menos, visíveis. Esse tipo de falsificação pode afetar o comportamento das bordas e cores da imagem.
- **Geração em computador**. Consiste na criação de objetos tridimensionais a partir de imagens ou vídeos. Aos objetos podem ser adicionadas texturas, cores, iluminação, etc. O objetivo desse processo é enriquecer os objetos com detalhes e aproximá-los de suas representações reais.
- **Cópia-colagem ou Clonagem (*cloning*)**. Consiste na alteração de partes de uma imagem usando segmentos ou propriedades da própria imagem. Esse tipo de falsificação pode afetar o comportamento de iluminação da imagem. Esse tipo de falsificação pode afetar o comportamento de iluminação da imagem.
- **Retoque e Conciliação (*retouching* e *healing*)**. Permitem o ajuste de regiões da imagem em termos de sombras, texturas, brilho, contraste, iluminação etc. Usando estas técnicas, é possível rejuvenescer uma pessoa, bem como eliminar vestígios da execução de outras operações, como cópia-colagem. Esse tipo de falsificação pode afetar o comportamento da iluminação da imagem.
- ***Lazy-Snapping***. Proposta por Li et al. [Li et al. 2004], esta técnica permite separar um objeto de interesse (*foreground*) da informação restante (*background*) em uma imagem. O método utiliza duas etapas: primeiramente o usuário define o que é fundo e o que é objeto através de curvas e, através de algoritmos de corte de grafos, a borda do objeto é computada. Na segunda etapa há a edição das bordas por parte do usuário caso o mesmo deseje corrigir erros da etapa anterior. Essa operação pode ser utilizada para selecionar de maneira precisa o objeto de uma imagem e utilizá-lo para realizar uma operação de composição.
- **Propagação Estrutural (*Structure Propagation*)**. Proposta por Sun et al. [Sun et al. 2005], essa técnica consiste em completar um espaço vazio da imagem de tal forma que o preenchimento resulte numa imagem consistente. Este método requer a especificação de informações estruturais pelos usuários através de segmentos de linha ou curvas, utilizadas para propagação de informações das regiões conhecidas

para a região vazia [Rocha and Goldenstein 2010]. Após essa etapa, as regiões desconhecidas restantes são completadas com informações de textura provenientes da vizinhança, processo este que facilita operações de cópia-colagem.

- **Cópia-Colagem Baseada em Conteúdo (*Content-Aware Fill)***: Nova ferramenta integrante do *Photoshop CS5* para cópias e colagens sofisticadas que levam o conteúdo da região analisada em questão no processo de clonagem. Nesse processo, um objeto pode ser removido da cena e a informação perdida pela sua remoção é recuperada a partir da seleção, por parte do usuário, do objeto e também de parte do fundo. A informação estrutural do fundo é então sintetizada em remendos (*patches*), que preenchem o vazio criado pela remoção da imagem.

Com essas numerosas ferramentas de manipulação de imagens, faz-se necessária a criação de metodologias de detecção eficazes da utilização das mesmas. Discutiremos esse assunto na próxima seção.

3. Metodologias

Uma vez que adulterações de vídeos e imagens vêm se tornando cada vez mais sofisticadas e eficazes, a necessidade de desenvolver métodos capazes de detectar tais adulterações torna-se eminente. Nesta seção, apresentamos os principais métodos desenvolvidos em nosso laboratório, os quais incrementam o estado da arte com diversas contribuições.

3.1. Atribuição de Fonte de Imagens

Uma forma de verificar a integridade e a autenticidade de imagens é identificar a fonte geradora da imagem em questão. Isso geralmente é feito pela detecção de "marcas" deixadas na imagem pelo dispositivo gerador no momento da captura e geração da imagem. Estas marcas são provenientes de características próprias do dispositivo gerador, como defeitos de fabricação, modo de interação entre os componentes da câmera e a luz, algoritmos de geração de imagem implementados nos componentes do dispositivo, entre outros fatores.

Pesquisas na área de atribuição de fonte em imagens digitais procuram identificar a marca ou fabricante do dispositivo utilizado na geração de uma imagem, bem como o dispositivo exato. Em particular, a tarefa de se encontrar o dispositivo específico utilizado na captura de uma imagem é a mais estudada. Existem propostas na literatura voltadas para a identificação da origem de uma imagem por meio do padrão de ruído deixado na imagem pelo dispositivo [Lukas et al. 2006,Li 2010,Goljan et al. 2008], artefatos gerados por imperfeições dos sensores de captura de um dispositivo [Kurosawa et al. 1999,Geradts et al. 2001], e presença de partículas de poeira no sensor [Dirik et al. 2008].

Identificar o dispositivo que gerou uma determinada imagem é uma forma de se garantir, por exemplo, que um documento foi gerado por uma câmera e não é resultado de qualquer manipulação digital; a informação de que uma foto foi obtida por uma câmera digital apreendida sob possessão de um suspeito poderia classificá-lo não mais como um consumidor mas sim como produtor de, por exemplo, fotos de pornografia infantil.

Atualmente, as técnicas mais efetivas para a identificação do dispositivo de captura específico analisam os efeitos do ruído inserido no processo de captura de imagens. Nossa abordagem para atribuição de fonte é baseada na proposta de Lukáš et al. [Lukas et al.

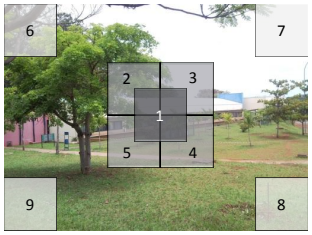


Figura 1. Regiões de interesse (ROIs) de dimensão 512×512 *pixels*.

2006], onde os autores propõem uma maneira de se realizar a estimativa do padrão de ruído dos sensores para identificar o dispositivo gerador de uma imagem. Em suma, o ruído residual médio de várias imagens da mesma câmera é obtido, gerando um padrão de referência da câmera sob investigação. Em seguida, é efetuada a correlação entre o ruído de uma imagem a ser avaliada e o padrão de referência calculado. Se o valor dessa correlação ultrapassa um limiar determinado em uma etapa de treinamento, os autores consideram que a imagem foi obtida pela câmera referente ao padrão usado na correlação.

Embora essa abordagem seja eficaz para a identificação da câmera que gerou uma imagem sob investigação, a pesquisa foi realizada considerando um cenário fechado (*closed-set*), no qual os autores assumem que uma imagem sob investigação foi gerada por uma entre n câmeras disponíveis durante a etapa de treinamento. Na prática, uma imagem a ser avaliada pode ter sido gerada por uma câmera totalmente desconhecida que não faz parte de nosso grupo de câmeras suspeitas, o que torna importante a identificação deste fato. Portanto, é importante modelar o problema de atribuição de fontes considerando um cenário aberto (*open-set*), no qual temos acesso somente a um conjunto limitado de câmeras suspeitas e temos que treinar o modelo de classificação considerando somente este conjunto enquanto buscamos classificar corretamente imagens geradas por câmeras às quais não necessariamente temos acesso. Assim sendo, propomos uma abordagem para resolver o problema de atribuição de fonte em cenário aberto, que consiste em três etapas:

A. Definição de regiões de interesse (ROIs). De acordo com Li e Satta [Li and Sata 2011], diferentes regiões da imagem podem ter diferentes informações sobre o padrão de ruído da câmera. Assim, visamos considerar várias regiões de uma imagem. Para cada imagem, foram extraídas nove regiões de interesse (*Regions of Interest* – ROI) de tamanho 512×512 *pixels*, de acordo com a Figura 1. A escolha dessas regiões de interesse nos permite trabalhar com imagens de diferentes resoluções.

B. Definição de características. Para cada região apresentada na Figura 1, nós calculamos o padrão de ruído conforme apresentado em Lukas et al. [Lukas et al. 2006], utilizando o filtro para extração de ruído no domínio da Transformada Discreta de Wavelet (DWT) proposto em [Mihcak et al. 1999], considerando os canais de cores R (vermelho), G (verde), B (azul). Foram realizados experimentos com outros espaços de cores e notamos que, para o problema de atribuição de fonte, a utilização do canal Y (luminância, do espaço de cor YCbCr [Wang and Weng 2000]) em conjunto com os canais R, G e B se mostrou eficaz.

Para cada ROI, extraímos o ruído residual de cada canal de cor utilizando um fil-

tro baseado na DWT. Em seguida, calculamos a média entre os ruídos de mesmo canal de várias imagens, gerando o padrão de ruído para cada canal de cor que representa a câmera sob investigação. Com isso, temos 36 padrões de ruído para representar uma câmera. Para cada imagem, calculamos seu ruído residual e criamos um vetor de características considerando a correlação entre cada ROI de uma imagem e o padrão de referência correspondente para cada câmera. Com essas correlações, temos 36 características para cada imagem, considerando uma câmera, rotulando imagens geradas pela câmera sob investigação como a classe positiva e as câmeras restantes disponíveis como classe negativa. Observe que algumas dessas imagens serão consideradas como sendo pertencentes à classe negativa desconhecida, ou seja, são imagens geradas por câmeras às quais não temos acesso na etapa de treinamento.

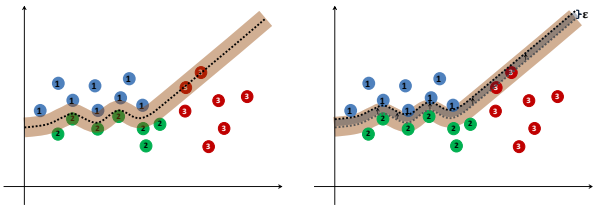
C. Atribuição de fonte em um cenário aberto. A principal contribuição deste trabalho é a utilização de aprendizado de máquina visando fazer a identificação da origem de imagens em um cenário aberto. Primeiramente, encontramos um classificador para treinar um conjunto de amostras considerando a classe de interesse e outras classes as quais temos acesso.O classificador escolhido se baseia no clássico algoritmo de Máquina de Vetores de Suporte (*Support Vector Machine* – SVM) [Bishop 2006] que transporta as amostras para um espeço de alta dimensão de forma que seja possível encontrara um hiperplano que faça a separação entre os dados da classe de interesse e das demais classes conhecidas.

Após o cálculo do hiperplano na etapa de treinamento, propomos um meio de classificar corretamente as classes desconhecidas na classificação por meio da movimentação do hiperplano de decisão por um valor ε se aproximando da classe positiva ou se afastando da(s) classe(s) negativa(s). A lógica é que, movendo o hiperplano, podemos ser mais restritos para o que nós sabemos como amostras positivas e, portanto, classificar qualquer outra amostra "muito diferente" como negativa, ou podemos ser pouco rigorosos sobre o que sabemos em relação às amostras positivas e aceitar pontos mais distantes do hiperplano como possíveis amostras positivas. Essa movimentação de plano tem como objetivo minimizar o erro de classificação na etapa de treinamento.

A Figura 2 descreve um exemplo para o caso não-linear. A Figura 2(a) apresenta o hiperplano de separação calculado, considerando amostras das classes azul e verde como as classes de interesse (1) e a classe negativa conhecida (2), respectivamente, e as amostras em vermelho representam as classes desconhecidas (3). A região laranja representa a distância entre as margens dos vetores de suporte das classes positiva e negativa. A Figura 2(b) mostra a operação de DBC sobre o hiperplano calculado, representado pela região azul.

Para os experimentos, nós construímos um conjunto de dados com 8500 imagens proveniêntes de 35 câmeras diferentes. As imagens foram obtidas considerando a resolução nativa das câmeras, em diversas configurações de iluminação, *zoom* e foco. O conjunto de dados é público e se encontra disponível para *download* em `http://www.recod.ic.unicamp.br/~filipe/image-source-attribution.zip`.

Nossa abordagem se mostrou bastante eficaz na identificação da câmera que capturou uma fotografia, obtendo um acerto médio de, aproximadamente, 98%, considerando um cenário onde temos acesso a 15 das 35 câmeras no treinamento, mas uma imagem a



(a) (b)

Figura 2. Nossa implementação do cenário aberto para atribuição de fonte de imagens utilizando *Decision Boundary Carving* (DBC).

ser avaliada pode ter sido gerada por qualquer uma das 35 câmeras. Publicamos esse método recentemente em [Costa et al. 2012].

3.2. Detecção de Cópia-Colagem

As possibilidades proporcionadas pelas ferramentas de software atuais, aliadas à criatividade e uma pitada de intenção (boa ou má) de seus usuários, são suficientes para que imagens digitais possam passar de genuínas a fraudulentas sem muito trabalho. Dentre outras, a Cópia-colagem (ou Clonagem) é uma forma de adulteração de imagens que pode gerar resultados supreendentes por meio de um esforço reduzido.

O objetivo da clonagem é ocultar ou multiplicar elementos presentes em uma cena, tais como folhagem, objetos e pessoas. Este efeito é alcançado (i) copiando-se um segmento da imagem, (ii) aplicando-se uma transformação geométrica neste, (iii) posicionando-se o segmento em outra região da mesma imagem e (iv) aplicando-se uma operação global nesta. A Figura 3 retrata o potencial da cópia-colagem.

(a) (b)

Figura 3. Exemplo de cópia-colagem. (a) é imagem original e (b) é a imagem manipulada.

Identificar uma cópia-colagem consiste em investigar uma imagem suspeita em busca de segmentos idênticos ou similares. Neste último caso, transformações como *Rotação*, *Redimensionamento* (ou *Escala*), *Espelhamento* (horizontal ou vertical) e *Suavização de Bordas* podem ser empregadas para modificar a região clonada e eliminar vestígios visuais da adulteração. Adicionalmente, operações globais como *Ruídos* e *Compressão JPEG* da imagem podem ser aplicadas com o mesmo propósito.

Neste trabalho, nós apresentamos uma nova heurística para detecção de cópia-colagem. Nossa estratégia tem como base os algoritmos *PatchMatch* e *PatchMatch Generalizado*, que são utilizados na busca por correspondências de *patches* em uma ou mais

imagens. A seguir, fornecemos uma breve descrição desses algoritmos e da nossa proposta para identificação de clonagens.

3.2.1. PatchMatch Generalizado

O *PatchMatch* [Barnes et al. 2009] é um algoritmo randomizado que se propõe a encontrar correspondências aproximadas de *patches* (blocos de *pixels* de tamanho definido, e.g., 7×7) em uma ou mais imagens por meio de um mecanismo baseado em Propagação e Busca Aleatória de correspondências na imagem. Uma correspondência (ou pareamento) para um *patch* pode ser entendida como um *patch* similar a este dentro da imagem.

Dando continuidade ao método original, em [Barnes et al. 2010] Barnes et al. propõem o *PatchMatch Generalizado*. O método consiste em averiguar a imagem de entrada de maneira similar à efetuada no *PatchMatch*, com a diferença de que, nesta abordagem, são encontradas não apenas uma, mas K correspondências para cada *patch* existente na imagem. Além disso, novas etapas (Enriquecimento Direto e Inverso) contribuem positivamente nos resultados. Essas etapas visam propagar boas correspondências de um *patch* para o espaço de correspondências desse mesmo *patch* (em oposição à etapa de propagação, que efetua tal difusão através das dimensões espaciais da imagem). Logo, o método passa a considerar um conjunto mais rico de pareamentos satisfatórios que podem aprimorar o NNF.

O *PatchMatch Generalizado* pode acomodar diversas métricas de similaridade entre *patches* e é facilmente adaptável para o emprego de descritores de imagens, tais como SIFT [Lowe 1999] e SURF [Bay et al. 2006].

3.2.2. Detecção de Cópia-colagem usando o PatchMatch Generalizado

Para esse problema, apresentamos uma heurística para identificação de clonagens que diverge daquelas existentes na literatura. Primeiramente, dada uma imagem suspeita, nós aplicamos o *PatchMatch Generalizado* com o propósito de encontrar, para cada um de seus *patches*, um conjunto de correspondências. Estas são tomadas como potenciais candidatas à duplicação dentro da própria imagem. Examinamos cada conjunto de *patches* candidatos a partir da verificação da vizinhança em que estes se encontram.

Dado que o *PatchMatch Generalizado* comporta medidas de similaridade diversas, nós adotamos uma métrica baseada em comparação de histogramas de intensidade/cor. A razão para isto é que a SSD não é robusta a pequenas variações nas regiões duplicadas. Para imagens em tons de cinza, computamos os histogramas de intensidade de ambos os *patches* sendo inspecionados e efetuamos o cálculo da Soma das Diferenças Absolutas (*Sum of Absolute Differences* – SAD) entre os bins de mesma posição. O valor desta soma é a distância de similaridade entre os dois *patches*. Caso a imagem possua informação de cor, calculamos os histogramas de cada canal (R, G e B) e cada *patch* passa a ser descrito por 3 histogramas. Em seguida, realizamos os cálculos de SAD para os histogramas de mesmo canal e aplicamos os valores encontrados na fórmula para conversão de cor em intensidade conforme $I = 0.299R + 0.587G + 0.114B$, na qual I representa a distância de similaridade. Descrevemos o método a seguir e o esquematizamos na Figura 4.

1. Empregamos o *PatchMatch Generalizado* na busca pelas K correspondências para cada *patch* da imagem. A inicialização do NNF é aleatória;
2. Percorremos o NNF final obtido em *scan order*;
3. Visitamos a lista de correspondências de cada *patch* (x, y);
4. Caso uma correspondência (x_i, y_i), sendo $i \leq K$, de (x, y) esteja a uma distância física deste inferior a um limiar T, não prosseguimos com a análise e passamos para a correspondência seguinte;
5. Comparamos a região ao redor do *patch* (x, y) com a região ao redor de cada uma de suas correspondências (x_i, y_i). Esta região de comparação abrange o *patch* mais um acréscimo de dois *pixels* em todas as direções. Caso um *patch* não possa crescer em quaisquer direções (borda da imagem), a mesma restrição se aplicará ao outro *patch* em comparação;
6. Se as duas regiões (ao redor de (x, y) e (x_i, y_i)) forem similares, isto é, se a distância de similaridade entre elas for menor do que um limiar D, marcamos ambas as regiões como duplicadas;
7. Se uma região já foi marcada, ela não é examinada novamente.

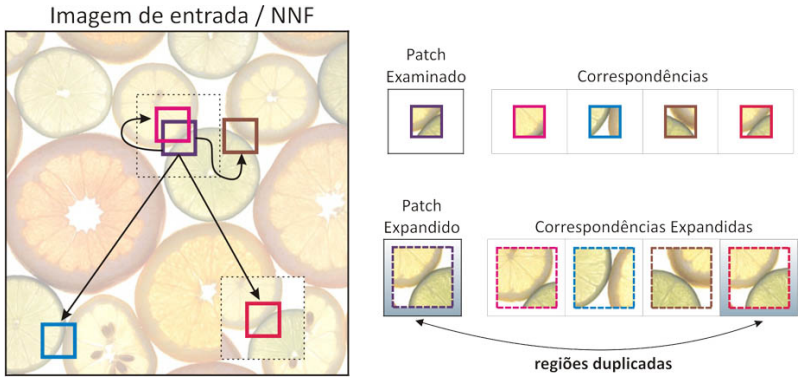


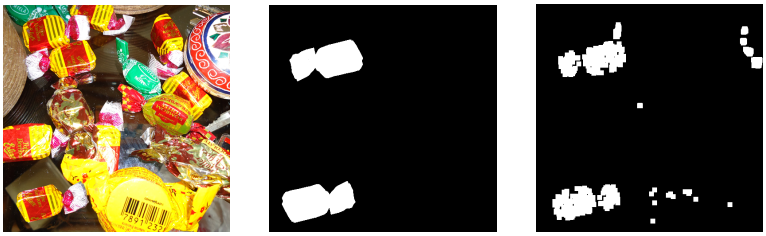
Figura 4. Esquematização da nossa proposta para detecção de clonagens baseada no *PatchMatch Generalizado*. O método examina todos os conjuntos de correspondências do NNF em busca daquelas que se encontram em uma vizinhança similar à vizinhança de um *patch*(x, y). Na Figura, a vizinhança do *patch* de borda roxa sendo avaliado é similar à vizinhança do *patch* de borda vermelha (duplicação).

Com esta abordagem, nós conseguimos contemplar, além de clonagens sem transformações adicionais, operações de espelhamento horizontal e vertical e rotações de 90, 180 e 270 graus porventura aplicadas nos segmentos duplicados. Este resultado é alcançado em decorrência da utilização de histogramas de intensidade/cor, que são invariantes a tais operações. Acreditamos que o método possui elevado potencial para detecção de clonagens nos cenários de operações de escala, suavização, compressão JPEG e de rotações em graus diversificados. Uma possível solução para isto seria a utilização de descritores SIFT ou SURF e a execução de modificações na ideia central do algoritmo.

A Figura 5 mostra um exemplo de clonagem e o resultado de detecção encontrado pelo algoritmo proposto.

3.3. Detecção de Composições (*Splicing*) em Imagens

Uma das formas mais comuns de adulteração de imagens são as realizadas através de uma operação conhecida como composição. Consiste na construção de uma nova im-



(a) (b) (c)

Figura 5. Resultado de uma detecção positiva utilizando o algoritmo proposto. (a) é a cópia-colagem, (b) é o mapa de referência que mostra as regiões duplicadas (em branco) e (c) é o resultado da detecção usando o nosso método.

agem utilizando partes do conteúdo de outras. Para identificar este tipo de adulteração, os métodos desenvolvidos pela comunidade forense baseiam-se em diversos tipos de características: inconsistências em descritores [Popescu 2004], inconsistências no processo de aquisição [Lin et al. 2005], inconsistências no processo de compressão [Luo et al. 2010] e inconsistências no processo de iluminação [Kee and Farid 2010].

Em especial, as abordagens baseadas em inconsistências de iluminação vêm ganhando espaço no cenário forense. Este tipo de abordagem destaca-se uma vez que um ajuste perfeito de iluminação em uma composição digital é extremamente difícil de se obter, uma vez que ao se realizar a composição utilizando partes de duas ou mais imagens, o fato de cada uma delas ser obtida em uma condição de iluminação diferente dificulta a criação deste tipo de falsificação. Outra vantagem desta classe de métodos é que eles podem ser utilizados para analisar imagens analógicas [Rocha et al. 2011].

A seguir, apresentamos uma de nossas contribuições para detecção de composições utilizando inconsistências de iluminação. O trabalho foi publicado em 2011 [Saboia et al. 2011] e utiliza os reflexões da luz nos olhos de pessoas para detectar falsificações em imagens contendo pessoas.

3.3.1. Eye Specular Highlights Telltales For Digital Forensics: A Machine Learning Approach

Nosso trabalho para detecção de composições através de inconsistências de iluminação presentes na imagem é baseado em uma extensão do artigo proposto por Johnson e Farid [Johnson and Farid 2007]. Nele, os autores se baseiam no fato de que a posição dos raios de luz refletidos nos olhos em imagens contendo pessoas é determinada pela posição relativa da fonte de luz, da superfície de reflexão do olho e do visualizador (neste caso, a câmera).

A Figura 6 exibe os três estágios que compõem o método original. O primeiro estágio estima a direção da fonte de luz para cada olho presente na fotografia de pessoas em análise. O segundo estágio (caracterização) procura estimar a posição da fonte de luz da imagem baseado nos raios de luz refletidos nos olhos e na correspondente fonte de luz estimada para cada olho. A posição calculada da fonte de luz é utilizada para calcular o erro angular para cada raio de luz. Finalmente, o terceiro estágio (decisão) calcula o erro angular médio e utiliza o teste de hipótese clássico com 1% de significância para decidir

se uma imagem sob investigação é uma composição.

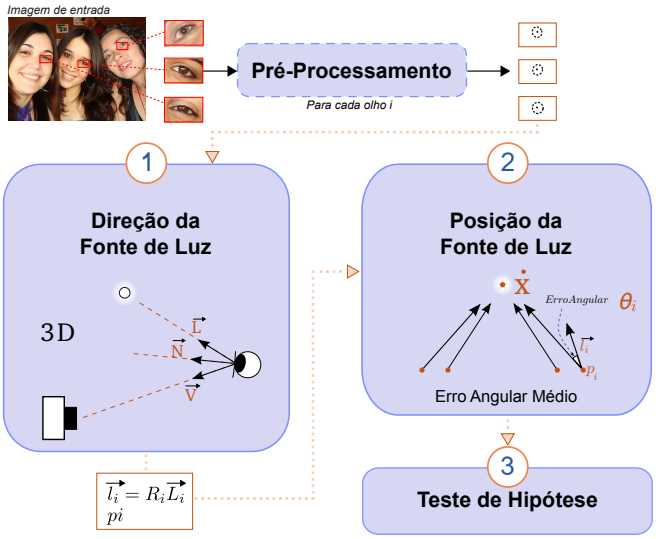


Figura 6. Diagrama mostrando os três estágios do método original proposto por Johnson e Farid [Johnson and Farid 2007]. Imagem retirada de [Saboia et al. 2011].

Procuramos por características mais robustas além das já propostas na literatura. No trabalho original de Johnson e Farid [Johnson and Farid 2007], os autores levaram em consideração apenas as informações referentes à iluminação da cena. No entanto, descobrimos que uma outra informação, antes ignorada, também era de grande importância para a identificação de manipulações. Tal informação diz respeito à posição do visualizador (neste caso, o visualizador é a câmera fotográfica) da cena. Essa informação, juntamente com a posição da fonte de luz, foi utilizada para descrever a imagem.

Outra característica importante é que, dado que os métodos que estimam a posição do observador e da fonte de luz são não determinísticos, eles podem produzir uma pequena variação em seu resultado. Nós exploramos tal variação para produzir diferentes caracterizações para cada imagem.

Uma vez que cada imagem possuía diferentes caracterizações (em nossos experimentos, utilizamos 5 descritores por imagem), foi possível substituir o teste de hipótese clássico da etapa de classificação por uma combinação de classificadores, em que cada um dos vetores de características era classificado por um classificador independente (em nossos experimentos, utilizamos o SVM [Bishop 2006]). Isso produziu classificações diferentes para uma mesma imagem. Por fim, combinamos os resultados dessas classificações para obter a classificação final.

Como principais contribuições decorrentes deste trabalho, podemos citar:

- um resultado com uma taxa de erro na classificação 20% menor que o trabalho original;
- a proposta de novas características não levadas em conta pelos autores originais.

A Figura 7 mostra um exemplo de composição e o resultado de detecção encontrado pelo algoritmo proposto.

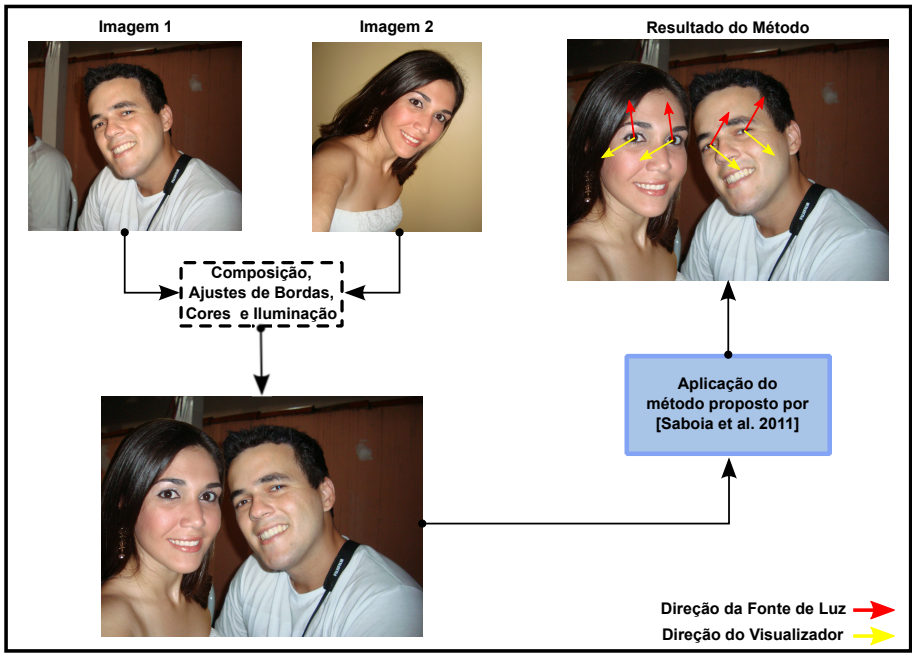


Figura 7. Resultado de uma detecção positiva utilizando o método proposto por [Saboia et al. 2011]. As direções estimadas para o visualizador e para a fonte de luz de cada uma das pessoas apontam em direções opostas, caracterizando uma imagem de composição.

4. Conclusões e Trabalhos Futuros

No cenário forense atualmente, existem diversos problemas em aberto esperando por soluções eficientes e eficazes. Neste trabalho, discutimos nuances de tais problemas, bem como algumas contribuições desenvolvidas em nosso laboratório na Universidade Estadual de Campinas para solucioná-los.

A atribuição de fonte de dispositivos tem como objetivo identificar qual foi o dispositivo gerador de uma imagem. Nesse contexto, nós exploramos soluções para o problema de atribuição de fonte de imagens geradas por câmeras digitais, uma tarefa de fundamental importância em um cenário criminal. Consideramos um cenário mais realístico, denominado cenário aberto, onde uma imagem sob investigação pode ter sido gerada por qualquer dispositivo, e não somente pelos dispositivos disponíveis no momento do treinamento. A abordagem proposta apresentou bons resultados, e com ela é possível analisar imagens de diferentes resoluções. Além disso, podemos identificar a fonte de imagens considerando métodos de caracterização complementares, tirando vantagem de todos os potenciais métodos de classificação de padrões por aprendizado de máquina.

O desafio da detecção de cópia-colagem em imagens digitais ainda possui diversas lacunas a serem preenchidas. Contemplar operações de rotação e escala tem sido uma das principais frentes de estudo dos pesquisadores acerca do tema. Acreditamos que a nossa metodologia, baseada no mecanismo aleatório e de propagação de correspondências do *PatchMatch Generalizado*, contribui com a literatura ao apresentar uma heurística inovadora para identificação de clonagens. Uma extensão direta do método proposto englobaria a combinação de descritores invariantes àquelas operações, tais como SIFT e SURF, para caracterização dos *patches* inspecionados.

Finalmente, a detecção de composições em imagens utilizando inconsistências de iluminação é uma das formas mais promissoras existentes atualmente. No entanto, o atual estado da arte, incluindo nossa contribuição, ainda trata o problema baseando-o em diversas premissas, as quais muitas vezes restringem o cenário de aplicação do método a cenários específicos, como imagens contendo pessoas por exemplo.

Mais especificamente, nosso método apresentado para detecção de composições é aplicável, apenas, em imagens contendo duas ou mais pessoas, e que tais pessoas estejam com os olhos visíveis. Assim, torna-se necessário o desenvolvimento de métodos capazes de utilizar inconsistências de iluminação para detectar composições em qualquer tipo de cenário, ou pelo menos com um número menor de restrições. Para isso, como trabalhos futuros abordaremos uma nova forma de representação menos restritiva para a iluminação da cena tentando desta forma aumentar o domínio de aplicação de nossos futuros métodos.

Agradecimentos

Agradecemos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Microsoft Research, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento dessa pesquisa.

Referências

Barnes, C., Shechtman, E., Finkelstein, A., and Goldman, D. B. (2009). Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM ToG*, pages 24:1–24:11.

Barnes, C., Shechtman, E., Finkelstein, A., and Goldman, D. B. (2010). The generalized patchmatch correspondence algorithm. In *ECCV*, pages 29–43.

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *ECCV*, pages 404–417.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning, 1st edition*. Springer.

Costa, F. O., Eckmann, M., Scheirer, W. J., and Rocha, A. (2012). Open set source camera attribution. In *SIBGRAPI*, pages 71–78.

Delp, E., Memon, N., and Wu, M. (2009). Digital Forensics [From the Guest Editors]. *IEEE SPM*, 26:14–15.

Dirik, A. E., Sencar, H. T., and Memon, N. (2008). Digital single lens reflex camera identification from traces of sensor dust. *IEEE T.IFS*, 3(3):539–552.

Farid, H. (2008). Digital Image Forensics. *Scientific American*, 6:66–71.

Geradts, Z. J., Bijhold, J., Kieft, M., Kurosawa, K., Kuroki, K., and Saitoh, N. (2001). Methods for identification of images acquired with digital cameras. *Enabling Technologies for Law Enforcement and Security*, 4232:505–512.

Goljan, M., Fridrich, J., and LukÃ¡š, J. (2008). Camera identification from printed images. In *SPIE Conference on Security, Forensics, Steganography, and Watermarking of Multimedia Contents*, volume 6819.

Johnson, M. and Farid, H. (2007). Exposing Digital Forgeries Through Specular Highlights on the Eye. In *Information Hiding*, pages 311–325.

Kee, E. and Farid, H. (2010). Exposing Digital Forgeries from 3-D Lighting Environments. In *IEEE WIFS*, pages 1–6.

Kurosawa, K., Kuroki, K., and Saitoh, N. (1999). CCD fingerprint method – identification of a video camera from videotaped images. In *IEEE ICIP*, pages 537–540.

Li, C.-T. (2010). Source camera identification using enhanced sensor pattern noise. *IEEE T.IFS*, 5(2):280–287.

Li, C.-T. and Sata, R. (2011). On the location-dependent quality of the sensor pattern noise and its implication in multimedia forensics. In *ICDP*, pages 1–6.

Li, Y., Sun, J., Tang, C.-K., and Shum, H.-Y. (2004). Lazy snapping. *ACM ToG*, pages 303–308.

Lin, Z., Wang, R., Tang, X., and Shum, H. (2005). Detecting Doctored Images Using Camera Response Normality and Consistency. In *IEEE CVPR*, pages 1087–1092.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *ICCV*, page 1150.

Lukas, J., Fridrich, J., and Goljan, M. (2006). Digital Camera Identification from Sensor Pattern Noise. *IEEE T.IFS*, 2:205–214.

Luo, W., Huang, J., and Qiu, G. (2010). JPEG Error Analysis and Its Applications to Digital Image Forensics. *IEEE T.IFS*, 5:480–491.

Mihcak, M. K., Kozintsev, I., Ramchandran, K., and Moulin, P. (1999). Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Processing Letters*, 6(12):300–303.

Popescu, A. (2004). *Statistical Tools for Digital Image Forensics*. PhD thesis, Department of Computer Science, Dartmouth College.

Rocha, A. and Goldenstein, S. (2010). CSI: Análise Forense de Documentos Digitais. *Atualizações em Informática (JAI)*, pages 263–317.

Rocha, A., Scheirer, W., Boult, T., and Goldenstein, S. (2011). Vision of the Unseen: Current Trends and Challenges in Digital Image and Video Forensics. *ACM Computer Survey*, 43(4):1–42.

Saboia, P., Carvalho, T., and Rocha, A. (2011). Eye Specular Highlights Telltales for Digital Forensics: a Machine Learing Approach. In *IEEE ICIP*, pages 1937–1940.

Sacchi, D., Agnoli, F., and Loftus, E. (2007). Changing History: Doctored Photographs Affect Memory for Past Public Events. *Applied Cognitive Psychology*, 21(8):1005–1022.

Sun, J., Yuan, L., Jia, J., and Shum, H.-Y. (2005). Image completion with structure propagation. *ACM ToG*, pages 861–868.

Wang, X. and Weng, Z. (2000). Scene abrupt change detection. In *Canadian Conference on Electrical and Computing Engineering*, pages 880–883.