Capítulo 6

CSI[1]: Análise Forense de Documentos Digitais

Anderson Rocha, Siome Goldenstein[2]

Abstract

In this chapter, we introduce the reader to the emerging field of Digital Media Forensics which aims at uncovering and analyzing the underlying facts about a digital object (e.g., images and videos). We pay special attention to source camera identification and forgery detection research topics. Source camera's identification goal is to identify the particular model of a camera, or the exact camera, that captured a document. Forgery detection's goal is to establish the authenticity of a digital document, or to expose any potential tampering it might have undergone. We describe each of these components of digital media forensics in detail, along with a critical analysis of the state-of-the-art, and recommendations for the direction of future research.

Resumo

Neste capítulo, apresentamos um estudo crítico das principais técnicas existentes no emergente campo de pesquisas denominado análise forense de documentos digitais (e.g., imagens e vídeos) para verificação de sua autenticidade e integridade. Focamos em duas sub-áreas de pesquisa: a identificação da câmera que originou uma determinada imagem ou vídeo bem como a identificação de adulterações em documentos digitais. Com relação à identificação da origem de uma imagem ou vídeo, o objetivo é identificar o modelo particular de uma câmera ou filmadora, ou a câmera exata. O objetivo da detecção de adulterações em documentos é estabelecer a autenticidade dos mesmos, ou expor quaisquer tipos de adulterações sofridas. Finalmente, apresentamos também tendências e recomendações para projetos futuros.

[1] *Crime Scene Investigation*

[2] Os autores estão juntos ao *Reasoning for Complex Data* (RECOD) Lab., Instituto de Computação – Universidade Estadual de Campinas (UNICAMP), Caixa Postal 6176, CEP 13.083-970, Campinas, SP, Brasil e podem ser contatados nos endereços {anderson.rocha,siome}@ic.unicamp.br.

6.1. Introdução

Uma imagem vale mil mentiras — Anônimo

Com o advento da *internet* e das câmeras de alta performance e de baixo custo juntamente com poderosos pacotes de *software* de edição de imagens e vídeos (e.g., *Adobe Photoshop* e *Illustrator*, *Apple Final Cut Pro*), usuários comuns tornaram-se potenciais especialistas na criação e manipulação de documentos digitais. Quando estas modificações deixam de ser inocentes e passam a implicar questões legais, torna-se importante o desenvolvimento de abordagens eficientes e eficazes para sua detecção [Rocha et al. 2011, Rocha e Goldenstein 2010].

A identificação de imagens que foram digitalmente adulteradas é de fundamental importância atualmente. Ao campo de pesquisas relacionado à análise de documentos digitais para verificação de sua autenticidade e integridade denominamos *Análise Forense de Documentos Digitais*. O julgamento de um crime, por exemplo, pode estar sendo baseado em evidências que foram fabricadas especificamente para enganar e mudar a opinião de um júri. Um político pode ter a opinião pública lançada contra ele por ter aparecido ao lado de um traficante procurado mesmo sem nunca ter visto este traficante antes. Um banco pode aceitar, ingenuamente, como verdadeiro um cheque falsificado de um cliente.

Com o avanço tecnológico, as adulterações digitais têm atingido os mais variados meios de comunicação, inclusive o meio científico. Existem indícios de que boa parte das imagens científicas publicadas em veículos respeitados são adulteradas [Rocha et al. 2011]. Outra preocupação se refere à perspectiva histórica. Recentemente, cientistas levantaram a hipótese de que a adulteração de imagens de eventos históricos afetam a memória das pessoas em relação a tais eventos [Sacchi et al. 2007]. Finalmente, existe a preocupação econômica: com a existência de *software* e equipamentos de impressão de qualidade, tornou-se muito mais fácil a falsificação de documentos financeiros como, por exemplo, cheques (re-impressão, modificação de valores, falsificação de assinatura, etc.).

Nesse capítulo, discutimos as principais formas utilizadas para criação/adulteração de conteúdo digital atualmente, bem como apontamos algumas limitações das técnicas existentes na detecção dessas falsificações. Estamos interessados na proposição de abordagens que permitam a normatização da pesquisa existente na área de detecção de falsificações em imagens e vídeos digitais dado que, atualmente, os trabalhos existentes ainda são insulares e não possuem uma metodologia ou conjunto de dados padrão para testes e comparação.

Organizamos o restante do capítulo da seguinte forma. A Seção 6.2 nos traz alguns fatores históricos ligados à manipulação de documentos analógicos e digitais. A Seção 6.3 discute algumas formas de manipulação de imagens e vídeos digitais presentes atualmente. A Seção 6.4 apresenta o estado da arte

na análise forense de imagens e vídeos. Finalmente, a Seção 6.5 conclui o capítulo.

6.2. Aspectos Históricos

Quem disse que a câmera nunca mente foi um mentiroso — Russell Frank

Nesta seção, apresentamos os principais fatos históricos relacionados à falsificação de documentos desde as primitivas combinações e adulterações analógicas de filmes fotográficos às mais recentes edições de imagens em ferramentas de *software* como o Adobe Photoshop.

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 [Kossoy 2006] criou a primeira fotografia em 1814[3], já apareciam as primeiras fotografias adulteradas [Rocha et al. 2011]. A Figura 6.1 mostra um dos primeiros exemplos de falsificação de imagens. A fotografia conhecida como *The two ways of life* é de Oscar G. Rejland, 1857. Esta montagem analógica consiste em uma composição de 30 imagens.

Figura 6.1. Composição analógica de 30 imagens. Oscar Rejland, 1857.

O regime Stalinista usou e abusou de técnicas de adulteração em imagens para "moldar" a história de acordo com o que lhe conviesse [Farid 2007]. Se algum indivíduo, outrora fotografado em algum evento, tornava-se desafeto do regime, todas as suas aparições em registros fotográficos oficiais eram eliminadas. A Figura 6.2 mostra um exemplo. Nesta fotografia, o ditador soviético Josef Stalin aparece com e sem a presença do comissário de água e transporte Nikolai Yezhov. Yezhov foi executado em 1940.

[3] Estudos recentes demonstram que a fotografia foi, na verdade, inventada concorrentemente por vários pesquisadores tais como Nicéphore Niepce, Louis Daguerre, Fox Talbot, e Hercule Florence. Este último, por sinal, realizando experiências no interior de São Paulo, na vila de São Carlos, hoje Campinas [Kossoy 2006].

Figura 6.2. Josef Stalin com (original) e sem (adulterada) a presença de Nikolai Yezhov.

A maior parte das adulterações anteriores à era digital necessitava de alta capacidade técnica e muitas horas (talvez dias) de trabalho em salas escuras de fotografia [Rocha et al. 2011, Popescu 2004]. No entanto, após a era digital, esse tipo de adulteração tornou-se muito comum e, hoje convivemos diariamente com exemplos de imagens modificadas digitalmente. Essas alterações variam de simples correções de brilho, cor e contraste feitas por usuários comuns querendo recuperar uma fotografia de família mal capturada à atividades criminais ou com interesses escusos [Sencar e Memon 2008].

A Figura 6.3 mostra um exemplo recente. A fotografia é de Brian Walski e apareceu no jornal Los Angeles Times em 2003. Walski combinou duas imagens para retratar um momento histórico na guerra do Iraque. No entanto, como pode ser observado nas imagens originais, o momento histórico e único em que o cidadão iraquiano segurando uma criança nos braços olha com esperança para o soldado britânico nunca existiu. Walski foi despedido após o incidente.

Nos últimos anos tivemos um crescimento de casos relacionados ao processamento questionável de imagens digitais; principalmente para fins políticos. Pouco tempo após a indicação da americana Sarah Palin como possível candidata a vice presidente dos Estados Unidos pelo partido republicano, uma imagem[4] foi amplamente distribuída na *internet* mostrando Sarah Palin de biquíni segurando um rifle (Figura 6.4(a)). Tempos depois, descobriu-se que a imagem era uma composição da cabeça de Palin com o corpo de outra pessoa.

Casos dessa natureza não estão tão distantes de nós cidadãos brasileiros. Em abril de 2009, o jornal *Folha de São Paulo* publicou um artigo sobre como a então ministra da Casa Civil Dilma Rousseff (possível candidata a presidente em 2010 pela situação) participara de ações de resistência e terrorismo durante o governo militar. Como parte da matéria, o jornal divulgou a imagem de uma alegada ficha policial (Figura 6.4(b)) da então ministra afirmando que a mesma foi retirada dos arquivos do Departamento de Ordem Política e Social (DEOPS) junto ao Arquivo Público de São Paulo. Em uma análise detalhada,

[4] Consulte o Apêndice A.1 para uma definição formal de imagem.

Figura 6.3. Soldado britânico "orienta" iraquianos. Fotografia e adulteração de Brian Walski.

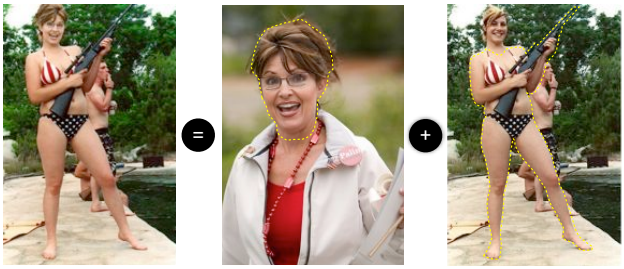
Goldenstein e Rocha [Goldenstein e Rocha 2009] provaram que o documento é falso. A fotografia no documento é o resultado de uma operação de composição (*splicing*) de uma outra imagem em tons de cinza, o texto é resultado de manipulação e inserção digital e, finalmente, o documento não é resultado de um processo de escaneamento.

A comunidade científica também tem sido vítima de falsificações digitais. Dois casos, em particular, chamam a atenção. Em 2004, o professor sul-coreano Hwang Woo-Suk e colegas publicaram um artigo na renomada revista científica *Science* reportando importantes avanços na pesquisa sobre células tronco. Menos de um ano depois, um painel investigativo apontou que nove das onze colônias que Hwang afirmara serem verdadeiras foram fabricadas a partir de duas outras autênticas [Rocha et al. 2011, Choe Sang-Hun 2006]. Outra caso alarmante veio à tona em julho de 2007 quando o professor R. Michael Roberts e colegas da *Missouri University* retiraram seu artigo[5] publicado pela *Science* após um painel investigativo revelar que as imagens publicadas eram adulteradas [Kavanagh 2006].

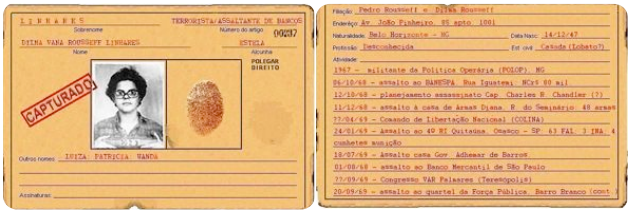
Infelizmente, os casos acima não são isolados. Em pelo menos um veículo importante de comunicação[6] estima-se que pelo menos 20% das publicações aceitas contenham imagens com manipulações impróprias. Ainda mais preocupante é o fato de que aproximadamente 1% dessas mesmas publicações contém manipulações fraudulentas [Pearson 2005]. Para se ter uma idéia, em 1996, cerca de 6% das análises feitas pelo *U.S. Office of*

[5] *Cdx2 Gene Expression and Trophectoderm Lineage Specification in Mouse Embryos.*

[6] *Journal of Cell Biology.*



(a) Montagem buscando denegrir a imagem da candidata republicana Sarah Palin às eleições americanas em 2008.



(b) Ficha falsa publicada pelo jornal Folha de São Paulo.

Figura 6.4. Exemplos recentes de foto-montagens com fundo político.

Research Integrity, que monitora as publicações científicas americanas, envolvia imagens científicas contestadas. Em 2005, este número subiu para 44% [Parrish e Noonan 2009].

Com o avanço das tecnologias de captura de vídeos bem como a facilidade de compartilhamento (e.g., *Youtube*), vídeos digitais estão cada vez mais presentes em nossas atividades cotidianas. Embora falsificações em vídeos sejam relativamente mais difíceis de serem feitas, temos encontrado diversos casos nos últimos anos. A Figura 6.5 mostra um quadro do vídeo parte do programa televisivo russo *The People Want to Know*. Neste vídeo, o analista político Mikhail Delyagin foi removido (pelo menos em grande parte) após fazer duras críticas ao primeiro ministro russo Vladimir Puttin [Clifford J. Levy 2008]. Note que apenas parte de Delyagin foi removida (sua perna e mão permanecem visíveis à direita do homem que segura o microfone). Emissoras de televisão têm, cada vez mais, utilizado técnicas de edição de imagens e vídeos segundo suas necessidades. Recentemente, em pelo menos dois casos reportados, uma rede de televisão brasileira utilizou recursos de edição para eliminar informações em matérias que foram ao ar [Folha de São Paulo 2010, UOL Notícias 2009].

Mais de 30 anos de pesquisa relacionadas à distorção de memórias mos-

Figura 6.5. Resquícios de uma edição mal feita em que o analista político russo Mikhail Delyagin foi removido (pelo menos em parte) de um programa de televisão.

tram que o ato de "relembrar" não se trata apenas de um mecanismo de recuperação de uma peça particular de informação em um banco de dados. Ao contrário, é um processo de reconstrução pelo qual a memória original pode ser continuamente modificada. Por exemplo, por questões ligadas a *estresse* pós-traumático, não é incomum vítimas de roubos ou sequestros descreverem características erradas sobre seus agressores. Neste sentido, recentemente, uma publicação chamou a atenção de pesquisadores forenses. Em estudo publicado no periódico *Applied Cognitive Psychology*, cientistas italianos levantaram a hipótese de que não só o mecanismo de memória é continuamente modificado mas também suscetível ao erro. Nesse estudo, os autores mostraram que a adulteração de imagens de eventos históricos, afetam a memória, atitudes e comportamentos das pessoas em relação a tais eventos [Sacchi et al. 2007]. De forma preocupante, os participantes da pesquisa que viram imagens adulteradas de eventos históricos passaram a ter ou manifestar uma visão diferente sobre tal evento mesmo, em alguns casos, tendo participado de tais eventos em pessoa.

Após esse apanhado geral de casos forenses conhecidos na literatura, nas próximas seções, apresentamos algumas técnicas para análise forense de documentos bem como discutimos suas limitações. Em especial, damos mais atenção às técnicas relacionadas a imagens e vídeos.

6.3. Técnicas de manipulação de imagens e vídeos

Não duvide que nossa percepção de beleza é distorcida — Dove Inc.

A atividade forense precisa, antes de mais nada, distinguir simples operações de melhoria de imagens ou vídeos de alterações com intuito de falsificação. É importante ressaltar que qualquer operação de processamento de imagem pode ser utilizada para enganar o visualizador. No entanto, a distinção pre-

cisa ser feita em relação ao objetivo da edição em si. Em [Rocha et al. 2011], os autores fazem a distinção das principais operações de processamento de imagens em duas categorias.

De um lado encontram-se as operações de melhoria de imagem com o objetivo de melhorar sua visibilidade. Não há combinação de *pixels* ou mesmo qualquer tipo de operação localizada. Alguns exemplos de operações nesta categoria são: ajuste de brilho e contraste, correção gamma, redimensionamento, rotação e outros.

Por outro lado, operações de adulteração são aquelas com *intenção* de enganar o visualizador de alguma forma. Algumas operações comuns nesta categoria envolvem operações locais tais como combinação e modificação de *pixels*, cópia e colagem (*cloning*), composição com outras imagens (*splicing*), ajuste fino de bordas (*feather edges*), retoque e conciliação (*healing and retouching*), casamento de padrões de iluminação (*light matching*), entre outras.

Para complicar um pouco mais a linha tênue que diferencia a natureza dessas operações, existem aquelas que estão entre essas categorias mas que, por si só, não configuram operações simples pois envolvem combinação de *pixels* mas também não configuram operações de adulteração. No entanto, dependendo do objetivo da edição em uma determinada imagem, tais operações podem ser consideradas ou combinadas com a finalidade de gerar imagens falsas ou adulteradas. Alguns exemplos são: realce (*sharpening*), borramento (*blurring*) e compressão.

Dentre as operações com o objetivo de enganar o visualizador podemos destacar:

- **Composição (*splicing*)**. Consiste na composição de uma imagem utilizando partes do conteúdo de uma ou mais imagens. Um político, por exemplo, em uma fotografia F_1 pode ser colocado ao lado de uma pessoa em uma outra fotografia F_2, mesmo sem nunca ter visto tal indivíduo antes.
- **Ajuste fino de bordas (*feather edges*)**. Consiste no ajuste das bordas de um objeto após uma operação de composição, por exemplo, de modo a diminuir o máximo possível os artefatos gerados pela composição.
- **Casamento de padrões de iluminação (*light matching*)**. Consiste em ajustar a iluminação de uma composição de modo a eliminar artefatos de iluminação que possam levar à identificação das adulterações.
- **Realce (*sharpening*)**. Embora não altere a semântica geral de uma imagem ou vídeo, pode mudar a maneira como interpretamos os mesmos. Detalhes podem ser realçados ou obscurecidos de acordo com o interesse do adulterador.
- **Geração em computador**. Consiste na construção de modelos tridimensionais a partir de imagens ou vídeos de base. Pode-se aplicar cor e textura para dar mais realidade à cena criada.

- **Cópia e colagem (*cloning*)**. Consiste na cópia de algumas partes de uma imagem e posterior colagem em outras partes. Pode ser utilizado para eliminar detalhes ou objetos, por exemplo.
- **Retoque e conciliação (*healing and retouching*)**. Consiste em uma operação de clonagem mais sofisticada. Permite o casamento não apenas dos valores dos *pixels* tais como na clonagem mas também leva em consideração a textura, iluminação e sombras dos *pixels* amostrados. A partir desta técnica pode-se rejuvenescer uma pessoa em alguns anos ou mesmo alterar a disposição da cena de um determinado crime.

A Figura 6.6 mostra alguns exemplos das operações discutidas acima. A composição foi feita em *Adobe Photoshop* em menos de 30 minutos.

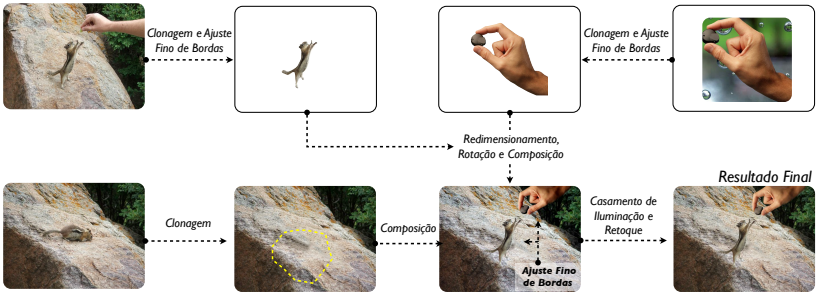


Figura 6.6. Exemplo contendo combinações de possíveis técnicas de edição de imagens tais como: clonagem, ajuste fino de bordas, composição e retoque.

Quando técnicas de edição e composição são utilizadas em conjunto, o trabalho de investigação se torna mais difícil. Para complicar, nesse dinâmico jogo de guerra entre os investigadores forenses e adulteradores, a cada dia aparecem operações de composição e criação de conteúdo mais sofisticadas.

Alguns exemplos recentes de técnicas avançadas de composição e edição são: *Lazy snapping* [Li et al. 2004], *Paint Selection* [Liu et al. 2009], *Poisson Matting* [Liu et al. 2009] e propagação estrutural [Sun et al. 2005].

1. *Lazy Snapping* [Sun et al. 2004] consiste em um método interativo para separar um objeto de um determinado fundo. O método possui duas etapas: uma etapa rápida de marcação e uma etapa de edição de bordas. O método ajusta (*snap*) a marcação grosseira do passo inicial ao contorno real do objeto mesmo em situações com arestas de pouco contraste ou ambíguas. O método utiliza um algoritmo de segmentação baseado em técnicas de cortes em grafos combinado com técnicas de

super-segmentação. Os experimentos realizados mostram que a técnica provê segmentações mais precisas em determinadas situações que técnicas presentes no estado da arte tais como o *Lasso Magnético* disponível no *Adobe Photoshop*.

2. *Paint Selection* [Liu et al. 2009], consiste em uma versão aperfeiçoada do *Lazy Snapping*. Nesta abordagem, os autores propõe dois algoritmos de otimização para tornar possível operações em imagens com resolução em *mega-pixels*: corte em grafo *multi-core* e reamostragem adaptativa por bandas da imagem. Devido às otimizações propostas, a abordagem precisa de menos *pixels* para seu processamento. A grande vantagem da nova abordagem está numa constatação aparentemente óbvia: a seleção interativa de objetos é um processo progressivo em que usuários podem ser envolvidos passo a passo. Dessa forma, não é necessário resolver um problema global de otimização para cada interação do usuário e sim uma série de otimizações locais que estejam na direção das intenções dos usuários.
3. *Poisson Matting* [Sun et al. 2004] consiste na formulação do problema de composição de imagens, também conhecido como *alpha matting*, utilizando equações de Poisson com restrições em relação ao campo de gradiente da composição (*matte*) bem como uma série de algoritmos de filtragem para permitir o ajuste fino por parte do usuário.
4. *Propagação Estrutural* [Sun et al. 2005] consiste em uma técnica capaz de melhorar os efeitos de clonagem de partes de uma imagem. Basicamente, a técnica permite a expansão de regiões conhecidas para regiões a serem removidas. Esta técnica também é conhecida como *image completion*. O usuário manualmente especifica informações estruturais da parte a ser eliminada utilizando um conjunto de curvas ou segmentos de linha a partir de regiões conhecidas e que devem ser "propagadas" para a nova região. A abordagem sintetiza regiões (*patches*) de imagem ao longo das marcações do usuário. A propagação estrutural é formulada como um problema de otimização em relação às diversas restrições estruturais e de consistência. Os *patches* de imagem são calculados utilizando-se programação dinâmica [Cormen et al. 2001] quando uma curva de restrição é especificada ou o algoritmo *Belief Propagation* [Yedidia et al. 2003] para duas ou mais curvas. Os *patches* achados são preenchidos utilizando-se técnicas de síntese de textura [Liang et al. 2001].

A união das quatro técnicas anteriores permite a criação de falsificações altamente realísticas. Por exemplo, um indivíduo pode utilizar a técnica de *Lazy Snapping* ou mesmo de *Paint Selection* para selecionar um objeto a ser eliminado de uma cena. Em seguida, utilizando a técnica de *Propagação Estrutural*, o indivíduo conseguirá eliminar este objeto minimizando os artefatos resultantes de tal operação. Finalmente, tomando uma terceira imagem em conjunto

com a técnica de *Poisson Matting*, o indivíduo conseguirá uma composição bastante realista.

6.4. Análise Forense de Imagens e Vídeos — Estado da Arte

Ver é acreditar? — Anônimo

Nessa seção, apresentamos as principais técnicas para detecção de falsificações em imagens e vídeos disponíveis na literatura bem como algumas de suas limitações. Por questões de consistência em relação às notações de diferentes contextos em diversos trabalhos, a notação doravante adotada pode ser diferente das publicações originais.

De forma geral, na análise forense de documentos, dado um objeto (e.g., imagem), queremos responder questões tais como [Sencar e Memon 2008]:

- Este objeto é original ou foi criado a partir da composição (cópia/colagem) de outros objetos digitais?
- Este objeto realmente representa um momento único ou foi digitalmente adulterado para enganar o visualizador?
- Qual é o histórico de processamento deste objeto?
- Quais partes do objeto sofreram adulterações e qual o impacto dessas modificações?
- O objeto foi adquirido pela câmera do fabricante $\mathcal{F}_1$ ou do fabricante $\mathcal{F}_2$?
- Este objeto realmente é originário da câmera, filmadora ou *scanner* $\mathcal{C}_1$ como afirmado?

Atualmente, não existem metodologias estabelecidas para verificar a autenticidade e integridade de objetos digitais de forma automática [Sencar e Memon 2008]. Embora a marcação digital (*watermarking*) possa ser utilizada em algumas situações, sabemos que a grande maioria das imagens e vídeos digitais não possui marcação. Adicionalmente, qualquer solução baseada em marcação digital implicaria a implementação de tal abordagem diretamente nos sensores de aquisição das imagens ou vídeos o que tornaria seu uso restritivo. Além disso, possivelmente haveria perdas na qualidade do conteúdo da imagem devido à inserção das marcações. Assim, as técnicas propostas na literatura para análise forense de imagens e vídeos podem ser categorizadas em três grandes áreas de acordo com o seu foco principal:

1. Identificação da origem do objeto;
2. Distinção entre objetos naturais e sintéticos;
3. Identificação de adulterações.

Uma segunda característica destas técnicas é que elas são chamadas técnicas de detecção cega e passiva. A detecção é cega no sentido de que não é necessário a presença do conteúdo original para comparação e é passiva no sentido de que não é necessário a utilização de nenhuma forma de marcação digital no processo geral [Rocha et al. 2011].

6.4.1. Identificação da origem do documento

Técnicas de identificação da origem do documento dizem respeito às abordagens para investigação e identificação das características do dispositivo de captura de um objeto (e.g., câmera digital, *scanner*, gravadora). Para estas técnicas, normalmente esperamos dois resultados: (1) a classe ou modelo da fonte utilizada e (2) as características da fonte específica utilizada. É importante ressaltar que os dispositivos normalmente codificam as condições de aquisição no cabeçalho da imagem (e.g., cabeçalho EXIF). No entanto, devido à facilidade com que tal informação pode ser destruída ou alterada, ela não tem muita utilidade para a análise forense.

As pesquisas nesta área têm focado a identificação da câmera digital que capturou uma determinada imagem ou vídeo, bem como a identificação do *scanner* que capturou uma imagem [Rocha et al. 2011, Sencar e Memon 2008].

O desenvolvimento de uma abordagem de identificação da fonte originadora de uma imagem ou vídeo requer conhecimentos das propriedades físicas e de operação de tais dispositivos. Normalmente, o processo de aquisição de uma imagem ocorre da seguinte maneira: a luz (representando a cena a ser fotografada) entra na câmera através das lentes, e passa por uma combinação de filtros que incluem, pelo menos, os filtros de infra-vermelho e anti-serrilhamento para garantir qualidade visual. A luz é então focada no sensor de captura que nada mais é que uma matriz de *sensels* ou *pixels* (elementos foto-sensíveis).

Os sensores mais utilizados são os baseados em CCDs (*charge-coupled devices*) ou CMOS (*complimentary metal-oxide semiconductor*). Cada ponto da matriz de captura (*sensel*) integra a luz incidente em relação ao espectro completo e obtém um sinal elétrico representando a cena fotografada. No entanto, por razões econômicas, normalmente cada *sensel* é monocromático. Desta forma, as máquinas digitais empregam CFA (*color filter array*) que arranjam os *pixels* em mosaico de forma que cada elemento tenha um filtro espectral e capte apenas uma banda do comprimento de onda. Os CFAs mais comuns empregam três sensores: vermelho (***R**ed*), verde (***G**reen*), e azul (***B**lue*). Como cada ponto tem apenas uma cor, as duas cores ausentes são inferidas por interpolação utilizando uma operação conhecida como demosaico. Após o demosaico, a imagem passa por outras operações tais como: correção pontual, realce, correção de abertura, correção gamma e compressão. A Figura 6.7 ilustra o processo de aquisição de uma imagem enquanto a Figura 6.8 mostra um exemplo de arranjo de *pixels* em mosaico utilizando CFAs.

A partir do modelo básico de aquisição de imagens e vídeos, cada conjunto de técnicas para identificação da origem de um objeto busca descobrir as propriedades que tornam um determinada origem única em relação às demais.

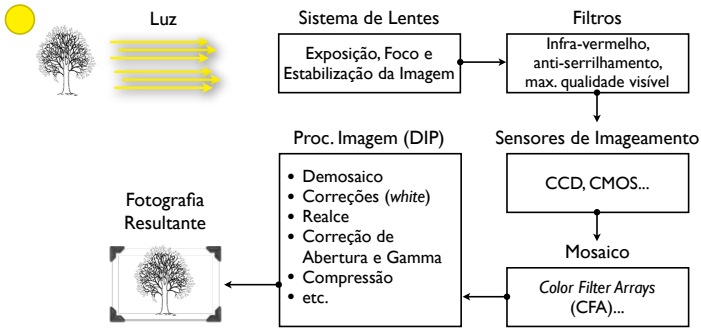


Figura 6.7. Possível *pipeline* do processo de aquisição de uma imagem (via câmera digital).

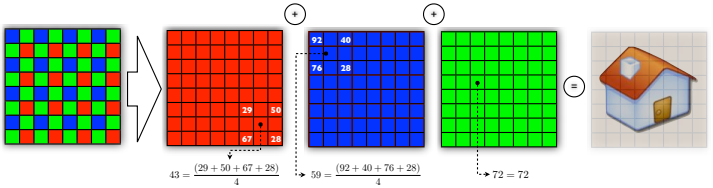


Figura 6.8. Arranjo de *pixels* em mosaico utilizando CFAs e posterior operação de demosaico. O filtro de mosaico/demosaico ilustrado é conhecido como filtro de Bayer [Gonzalez e Woods 2007].

6.4.1.1. Identificação do modelo de aquisição

O principal objetivo das técnicas para identificação do modelo de aquisição é apontar o modelo ou o fabricante de um dispositivo que capturou uma determinada imagem. No contexto de câmeras digitais, normalmente utilizamos informações relacionadas ao processo de aquisição da imagem tais como: informações das lentes, tipo e tamanho dos sensores de aquisição, tipo de filtro de mosaico e demosaico empregado, algoritmos de processamento de imagens implementados na lógica de processamento da câmera entre outras.

Para a identificação do modelo de aquisição, alguns pesquisadores têm utilizado: descritores para avaliar possíveis pós-processamentos nas imagens [Kharrazi et al. 2004]; artefatos decorrentes da escolha do sensor CFA e do algoritmo de demosaico [Bayaram et al. 2005b, Popescu 2004]; diversidade nas tabelas de quantização JPEG [Popescu 2004]; e distorções causadas pe-

las lentes [Choi et al. 2006]. Normalmente, estas abordagens apontam para uma classe ou modelo de câmera utilizado na captura e não uma câmera em específico.

Muitos fabricantes, no entanto, utilizam os mesmos componentes diminuindo o poder de discriminação de tais técnicas. A maior parte das técnicas nessa classe de identificação consiste na extração de características a respeito do modelo de câmera analisado para posterior utilização com alguma técnica de aprendizado de máquina. A seguir, apresentamos mais detalhes a respeito de algumas técnicas utilizadas para identificar o modelo de aquisição de uma determinada imagem.

Uma característica presente em fotografias capturadas por câmeras e lentes de baixo custo é a presença de distorções radiais. Essas distorções podem levar à identificação do modelo de câmera utilizado em uma determinada captura. Choi et al. [Choi et al. 2006] apresentam um método para extrair aberrações e distorções de imagens que, posteriormente, podem ser utilizadas em um classificador de padrões (c.f., Apêndice A.2). Os autores propõem modelar as distorções radiais presentes nas imagens a partir de estatísticas de segunda ordem

$$r_u = r_d + d_1 r_d^3 + d_2 r_d^5, \tag{1}$$

onde d_1 e d_2 são os parâmetros de distorção de primeira e segunda ordem e r_u e r_d são os raios com e sem distorção, respectivamente. O raio é a distância radial $\sqrt{x^2 + y^2}$ de algum ponto (x, y) a partir do centro de distorção (e.g., o centro da imagem) [Rocha et al. 2011, Choi et al. 2006]. Os parâmetros d_1 e d_2 são características (*features*) a partir das quais podemos treinar um classificador tal como o *Support Vector Machine* (SVM) [Bishop 2006] (c.f., Apêndice A.2). Adicionalmente, os autores propõem utilizar 34 características de descrição presentes em [Kharrazi et al. 2004] e apresentadas abaixo, em uma abordagem combinada. Os autores reportam um resultado em torno de 91% de acurácia quando separando três modelos de câmeras de diferentes fabricantes.

Os artefatos de cor inseridos durante o processo de demosaico também podem fornecer pistas importantes para a separação de modelos de câmeras. [Kharrazi et al. 2004] apresentam um conjunto de características para imagens coloridas que levam a um bom índice discriminatório entre diversos modelos de câmeras. Algumas características apresentadas pelos autores são: valores médios de *pixels*, correlação de pares RGB, distribuição de centros de massa de *pixels* vizinhos, razão de energia entre pares RGB, estatísticas no domínio de *wavelets*, e características de qualidade de imagem. Adicionalmente, os autores também utilizam medidas baseadas em diferenças de *pixels* (e.g., erro médio quadrático e erro médio absoluto), medidas de correlação (e.g., correlação cruzada), distância de Czenakowski entre outras. Os autores reportam entre 78% e 95% de acurácia para 5 modelos de câmera. Esses resultados foram confirmados em [Tsai e Wu 2006].

A escolha dos sensores para a operação de mosaico, CFA, du-

rante a captura de uma imagem, bem como a abordagem de demosaico utilizada também nos oferece pistas a respeito do modelo de câmera utilizado [Bayaram et al. 2005b, Popescu 2004, Celiktutan et al. 2005]. Em [Celiktutan et al. 2005], os autores estudaram as características da operação de demosaico para utilização em um classificador de padrões (c.f., Apêndice A.2). A motivação é que os algoritmos proprietários de demosaico deixam correlações ao longo de planos de *bits* adjacentes das imagens. Para analisar tais efeitos, os autores definem um conjunto de medidas (abordagens) de similaridade $\{m_1, m_2, m_3\}$ que, posteriormente, são utilizadas em classificadores como k vizinhos mais próximos e SVM [Bishop 2006].

A primeira abordagem é uma medida de similaridade baseada em uma função

$$\delta_c^n(a,b) = \begin{bmatrix} 1 & se & p_c = 0 & p_n = 0 \\ 2 & se & p_c = 0 & p_n = 1 \\ 3 & se & p_c = 1 & p_n = 0 \\ 4 & se & p_c = 1 & p_n = 1 \end{bmatrix} \tag{2}$$

onde b é um plano de *bits* (matriz da imagem) e a denota um de quatro *scores*[7]: $1, 2, 3,$ e 4. O subscrito c define algum *pixel* central e o superescrito n denota um de quatro possíveis *pixels* vizinhos.

Em seguida, fazemos a soma $\delta_c^n(a,b)$ em quatro direções (n itera ao longo de seus vizinhos acima, abaixo, à direita e à esquerda), bem como sobre todos os *pixels* (c itera sobre $M \times N$ *pixels*). Após as somas, podemos omitir o sub- e superescrito e calcular os termos de concordância (*agreement score*)[8], normalizados e obter histogramas de quatro *bins* (função de densidade de probabilidade):

$$\mathcal{A}_a^b = \delta(a,b) / \sum_a \delta(a,b). \tag{3}$$

A partir desses histogramas de 4 *bins*, podemos definir a distância de Kullback-Leibler binária como

$$m_1 = -\sum_{n=1}^{4} \mathcal{A}_n^7 \log \frac{\mathcal{A}_n^7}{\mathcal{A}_n^8}, \tag{4}$$

onde $\mathcal{A}$ é o termo de concordância normalizado. A intuição da distância aqui utilizada é verificar o quanto dois planos de *bits* são similares/correlacionados dado a intuição inicial de que a combinação de *pixels* empregada pelos algoritmos de demosaico proprietários tradicionais deixam artefatos nos canais de *bits* que compõem a *resolução de cor* de uma imagem (Consulte o Apêndice A.1 para uma definição formal).

[7] *Score* pode ser entendido com uma nota a ser atribuída.

[8] O termo de concordância é um termo de classificação do tipo de *bit* analisado em relação a um *bit* vizinho. Por exemplo, se o *bit* tem valor 0 e seu vizinho, em um outro plano de *bits*, tem valor 0, o termo de concordância, como definido pelos autores, é 1 (Equação 2).

A segunda abordagem é também uma medida de similaridade que utiliza uma máscara de ponderação[9] restrita a uma determinada vizinhança de *pixels*. Cada imagem binária resulta um histograma de 512 *bins* computado em relação à máscara de ponderação. Cada *score* é computado com a seguinte função

$$S = \sum_{i=0}^{7} p_i 2^i, \tag{5}$$

onde p_i é o *pixel* analisado dentro da máscara. A máscara de ponderação restrita a uma vizinhança 3×3 é definida como

1	2	4
128	256	8
64	32	16

Por exemplo, o *score*, segundo a Equação 5, se torna $S = 2 + 4 + 8 = 14$ no caso em que os *bits* em um plano de *bits* e restrito à máscara acima, tem os pontos E, N, NE como 1 e todos os outros como 0.

A medida de similaridade binária final é computada baseada na diferença absoluta do $n^{ésimo}$ *bin* do histograma no 7^o e 8^o planos de *bits* (*bit planes*) (c.f., Apêndice A.1) após a normalização

$$m_2 = \sum_{n=0}^{511} |S_n^7 - S_n^8|. \tag{6}$$

Medidas de qualidade de imagem, como mencionado anteriormente, podem ser de muita valia para a análise forense. A distância de Czenakowski, por exemplo, é uma característica popular para identificação de operações de mosaico/demosaico porque ela é capaz de comparar efetivamente vetores com componentes não negativos. Neste contexto, a terceira abordagem definida por [Celiktutan et al. 2005] é baseada na distância de Czenakowski

$$m_3 = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(1 - \frac{2 \sum_{k=1}^{3} \min(I_k(i,j), \hat{I}_k(i,j))}{\sum_{k=1}^{3} (I_k(i,j) + \hat{I}_k(i,j))} \right). \tag{7}$$

Esta métrica de distância requer a utilização de um filtro de supressão/-redução de ruído (*denoising*). $I_k(i,j)$ representa o $(i,j)^{ésimo}$ *pixel* da $k^{ésima}$ banda de cor de uma imagem colorida, e $\hat{I}_k$ é sua versão filtrada (*denoised*).

Com estas três medidas de similaridade, [Celiktutan et al. 2005] geram 108 características de similaridade binária e 10 características de qualidade por imagem analisada. Os autores reportam (para esta técnica em conjunto com o classificador SVM) aproximadamente 100% de acurácia para dois modelos

[9] Uma máscara de ponderação nada mais é do que um conjunto de pesos associados a uma região pré-definida. No contexto da técnica apresentada, a máscara serve para priorizar elementos em certas regiões mais que em outras.

de câmeras diferentes. Adicionalmente, os autores reportam 95% de acurácia para um cenário com três modelos de câmeras e 62.5% para um cenário com seis modelos diferentes.

O maior problema com as técnicas apresentadas acima está na falta de rigor na análise dos resultados experimentais comparado com outras áreas relacionadas à Visão Computacional e Reconhecimento de Padrões tais como Biometria e Rastreamento [Rocha et al. 2011].

A grande maioria das abordagens reporta resultados diretamente para um conjunto limitado de modelos de câmeras, sendo difícil estabelecer o comportamento dessas abordagens na prática. Outra característica de interesse seria a análise de modelos de câmera desconhecidos pelo treinamento do sistema em questão. Como os sistemas se comportam nesses casos? Mais importante, as técnicas até agora mostraram certo grau de competência para imagens com baixa compressão JPEG. Como tais técnicas se comportam na presença de compressões mais severas?

Diferente das abordagens anteriores, [Popescu 2004] apresentam um algoritmo de Maximização de Esperança (*Expectation/Maximization* ou EM) muito poderoso para identificação do algoritmo de demosaico utilizado em uma determinada imagem bem como para identificação de falsificações (Seção 6.4.3.4). A abordagem de [Popescu 2004] não se baseia diretamente em um problema de classificação supervisionado embora possa ser melhorado quando associado a técnicas de aprendizado como mostrado nas extensões propostas por [Bayram et al. 2005, Bayram et al. 2006].

A hipótese de motivação para o algoritmo de EM é que linhas e colunas de imagens interpoladas provavelmente possuem correlações com seus vizinhos. Essa informação de vizinhança pode ser fornecida por *kernels* de tamanhos específicos (e.g., 3×3, 4×4, e 5×5).

O algoritmo em si pode ser dividido em dois estágios. No estágio de Esperança (E), estima-se a probabilidade de cada amostra pertencer a um modelo em particular. No estágio de maximização (M), estima-se a forma específica das correlações entre as amostras. Ambos os estágios são iterados até um critério de convergência ser atingido.

Mais especificamente, podemos assumir que cada amostra pertence a um de dois modelos possíveis. Se uma amostra é linearmente correlacionada com seus vizinhos, ela pertence ao modelo $\mathcal{M}_1$. Se a amostra não é correlacionada com seus vizinhos, ela pertence ao modelo $\mathcal{M}_2$. A função de correlação linear é definida como

$$f(x,y) = \sum_{u,v=-k}^{k} \alpha_{u,v} f(x+u, y+v) + \mathcal{N}(x,y), \tag{8}$$

onde $f(\cdot,\cdot)$ é um canal de cor (R, G, ou B) de uma imagem após a operação de demosaico, k é um inteiro, $\mathcal{N}(x,y)$ representa amostras independentes e identicamente distribuídas de uma distribuição normal com média zero e variância um e u, v denotam os *offsets* dos *pixels* (e.g., $(x+u, y+v)$, $u = 1, v = 0$, denota

o vizinho à direita de x). Adicionalmente, $\vec{\alpha}$ é um vetor de coeficientes lineares que expressa as correlações, com $\alpha_{0,0} = 0$.

O estágio de Esperança (E) estima a probabilidade de cada amostra pertencer ao modelo $\mathcal{M}_1$ usando a regra de Bayes

$$\Pr\{f(x,y) \in \mathcal{M}_1 | f(x,y)\} = \frac{\Pr\{f(x,y)|f(x,y) \in \mathcal{M}_1\}\Pr\{f(x,y) \in \mathcal{M}_1\}}{\sum_{i=1}^{2} \Pr\{f(x,y)|f(x,y) \in \mathcal{M}_i\}\Pr\{f(x,y) \in \mathcal{M}_i\}}, \quad (9)$$

onde $\Pr\{f(x,y) \in \mathcal{M}_1\}$ e $\Pr\{f(x,y) \in \mathcal{M}_2\}$ são as probabilidades *a priori* e são assumidas como iguais a $1/2$. Se assumirmos que uma amostra $f(x,y)$ é gerada por $\mathcal{M}_1$, a probabilidade de que isso ocorra é

$$\Pr\{f(x,y)|f(x,y) \in \mathcal{M}_1\} = \frac{1}{\sigma\sqrt{2\pi}}\left[-\frac{1}{2\sigma^2}\left(f(x,y) - \sum_{u,v=-k}^{k} \alpha_{u,v} f(x+u, y+v)\right)^2\right]. \quad (10)$$

Nós estimamos a variância σ^2 no estágio M. Adicionalmente, assumimos que $\mathcal{M}_2$ tem uma distribuição uniforme.

O estágio M calcula uma estimativa de $\vec{\alpha}$ usando o método dos mínimos quadrados ponderados (na primeira iteração do estágio E, $\vec{\alpha}$ é aleatoriamente escolhido)

$$E(\vec{\alpha}) = \sum_{x,y} w(x,y)\left(f(x,y) - \sum_{u,v=-k}^{k} \alpha_{u,v} f(x+u, y+v)\right)^2. \quad (11)$$

Os pesos $w(x,y)$ são equivalentes a $\Pr\{f(x,y) \in \mathcal{M}_1 | f(x,y)\}$. Esta função de erro é minimizada por um sistema de equações lineares antes de resultar uma estimativa. Ambos os estágios são executados até um valor estável de $\vec{\alpha}$ ser atingido.

Popescu et al. [Popescu 2004] afirmam que os mapas de probabilidade gerados pelo algoritmo EM podem ser usados para determinar o algoritmo de demosaico utilizado em uma câmera em particular. Estas probabilidades tendem a ser agrupar. Em um teste com 8 algoritmos de demosaico diferentes [Popescu 2004], o algoritmo EM apresentou um resultado de 97% de acurácia. No pior resultado reportado (algoritmo de demosaico baseado em filtro da mediana 3×3 *vs.* número de gradientes variável), o algoritmo conseguiu um resultado de 87% de acurácia.

Desde sua proposição, várias extensões foram elaboradas sobre o trabalho de [Popescu 2004]. Em [Bayram et al. 2005], os autores aplicam o algoritmo de EM para um problema de identificação de câmeras em conjunto com um classificador SVM para análise dos mapas de probabilidade. Os autores reportam resultados de 96% de acurácia para um problema de duas câmeras e de 89% de acurácia para um cenário multi-classe. Em [Bayram et al. 2006], os autores propõem uma abordagem de fusão das características resultantes dos mapas de probabilidade do algoritmo EM com técnicas adicionais de detecção de artefatos de suavidade (*smoothing*) nas imagens. Para um cenário com três câmeras, os autores reportam um resultado de aproximadamente 98% de acurácia. Algumas outras variações do modelo EM original incluem a modelagem de erro

ao invés do cálculo dos coeficientes de interpolação [Long e Huang 2006] bem como o cálculo do erro assumindo-se um determinado padrão CFA em uma imagem [Swaminathan et al. 2006].

6.4.1.2. Identificação do dispositivo específico

A identificação da câmera em si e não do modelo utilizado na captura, requer características únicas em relação à câmera utilizada. Estas características podem ser decorrentes, por exemplo, de imperfeições dos componentes, defeitos e falhas decorrentes de efeitos do ambiente e condições de operação. O maior desafio é estimar o fabricante e o tipo da câmera a partir de apenas uma imagem. As abordagens de maior relevância neste sentido, têm analisado os efeitos do ruído inserido no processo de captura de imagens [Lukas et al. 2006] ou os artefatos originados pela presença de poeira nos sensores no momento da aquisição [Dirik et al. 2007]. Finalmente, algumas abordagens relevantes para identificação de *scanners* como meios originadores de imagens são [Gou et al. 2007, Khanna et al. 2007, Khanna et al. 2009].

É importante ressaltar que alguns dos componentes utilizados para a identificação do dispositivo específico que capturou uma imagem podem ser temporais por natureza (sujeira no sensor, por exemplo). Um investigador forense precisa estar atento a essas informações sempre que possível.

Um dos primeiros autores a sugerir a utilização das imperfeições nos sensores para o cenário forense foi [Kurosawa et al. 1999]. Em seu trabalho, os autores propunham a identificação do ruído de padrão fixo causado por *dark currents* em câmeras digitais. Um *dark current* pode ser definido como a razão pela qual elétrons se acumulam em cada *pixel* devido à ação termal. Essa energia termal é achada nas junções inversas dos pinos e é independente da quantidade de luz incidente. Em seu trabalho, os autores apenas intensificam os ruídos de padrão fixo enquanto propõem a sua detecção como defeitos localizados dos *pixels*.

Em [Geradts et al. 2001], os autores apresentam uma análise mais completa a respeito das imperfeições presentes nos sensores de captura. Para detecção, os autores utilizam *pixels* supersaturados (*hot pixels*), *pixels* com pouca saturação (*cold/dead pixels*) e defeitos agrupados (*pixel traps*).

Pixels supersaturados são *pixels* individuais no sensor de captura com uma carga maior que a normal. *Pixels* com pouca saturação são aqueles que apresentam pouquíssima ou nenhuma carga. Defeitos agrupados são uma interferência com o processo de transferência de carga durante a captura levando a uma linha total ou parcialmente danificada na imagem (e.g., toda branca ou toda preta).

A maior limitação no uso dessas características no cenário forense reside em sua efemeridade. Câmeras mais sofisticadas possuem sensores especiais para corrigir tais defeitos no momento da captura das imagens. Dessa forma,

apresentamos agora uma das técnicas mais efetivas para identificação do dispositivo específico que capturou uma imagem.

Em seu trabalho [Lukas et al. 2006], os autores apresentam uma análise mais formal para identificação de dispositivos de captura baseado em padrões de ruído. Para um maior entendimento, considere a Figura 6.9 que apresenta a hierarquia do ruído presente em uma imagem digital. Vemos dois tipos principais de padrões de ruído: fixo e de foto-responsividade não uniforme. O ruído de padrão fixo (FPN) é causado pelos *dark currents* descritos acima e não são considerados no trabalho de [Lukas et al. 2006]. A razão é que o ruído do tipo FPN é relacionado à diferenças *pixel* a *pixel* quando o sensor não está exposto à luz. Basicamente, FPN é um ruído aditivo que depende do tempo de exposição e temperatura ambiente. Adicionalmente, FPN pode ser eliminado pelos sensores de câmeras mais sofisticadas extraindo-se um quadro preto (*dark frame*) da imagem após sua captura.

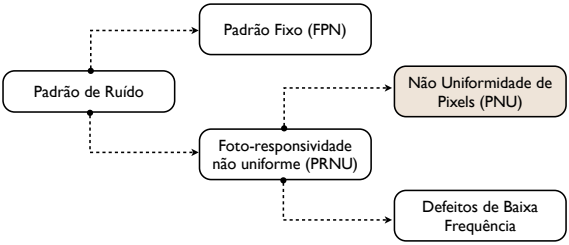


Figura 6.9. Hierarquia do padrão de ruído.

O ruído decorrente da foto-responsividade não uniforme (PRNU) é causado, basicamente, pelo ruído não uniforme dos *pixels* (PNU). PNU é definido como a frequência que diferentes *pixels* possuem à luz e é causado basicamente por inconsistências no sensor durante o processo de fabricação. Os defeitos de baixa frequência são causados por refração da luz nas partículas na ou próximas da câmera, superfície ótica e configurações de *zoom*. Lukas et al. [Lukas et al. 2006] não consideram este tipo de ruído, mas Dirik et al. [Dirik et al. 2008], em uma extensão posterior, consideram. A natureza efêmera dos defeitos de baixa frequência traz a questão da sua confiabilidade no cenário forense, exceto quando tratando de pequenas sequências de imagens de um mesmo período.

Para utilizar o ruído PNU para identificação dos sensores de captura, a natureza desse ruído precisa ser isolada. Um sinal de imagem μ exibe propriedades de um sinal branco com uma banda de atenuação de alta frequência. A atenuação é atribuída à natureza de filtro de passa baixas do algoritmo CFA (que, neste caso, não estamos interessados). Se uma grande porção da imagem é saturada (*pixels* com valor 255), não será possível separar o ruído PNU

do sinal da imagem. Em um cenário forense, certamente não teremos uma imagem de referência que facilmente nos permitiria recuperar as informações PNU. Dessa forma, o primeiro estágio do algoritmo de identificação de câmeras utilizando informações baseadas no ruído PNU consiste em estabelecer um padrão de referência P_c, uma aproximação do ruído PNU. No processo de aproximação, $\bar{I}^{(k)}$ é construído a partir da média de K diferentes imagens de uma cena uniforme (*lit scene*) $k = 1, \cdots, K$

$$\bar{I}^{(k)} = \frac{1}{K} \sum_{k=1}^{K} I^k. \tag{12}$$

A aproximação pode ser otimizada para suprimir o conteúdo da cena aplicando-se um filtro de supressão/redução de ruído λ, e fazendo-se a média dos resíduos $\xi^{(k)}$ ao invés das imagens originais $I^{(k)}$

$$\bar{\xi}^{(k)} = (\bar{I}^{(k)} - \lambda(I^{(k)}))/K. \tag{13}$$

Lukas et al. [Lukas et al. 2006] mostram que o filtro de supressão/redução de ruído baseado na transformada *wavelet* possui bons resultados.

Para determinar se uma dada imagem pertence a uma câmera em particular, calculamos a correlação ρ_c entre o ruído residual da imagem em questão $\xi = I - \lambda(I)$ e o padrão de referência P_c (a barra sobre o símbolo significa a média)

$$\rho_c(I) = \frac{(\xi - \bar{\xi}) \cdot (P_c - \bar{P}_c)}{\|\xi - \bar{\xi}\| \|P_c - \bar{P}_c\|}. \tag{14}$$

Lukas et al. [Lukas et al. 2006] apresentam resultados expressos em termos de falsos positivos e falsos negativos. Os autores reportaram uma taxa de falsos negativos entre 5.75×10^{-11} e 1.87×10^{-3} para uma taxa de falsos positivos fixa em 10^{-3} em um cenário com nove câmeras diferentes.

Uma melhoria para esta abordagem foi proposta por [Sutcu et al. 2007], com uma técnica capaz de fundir informações do ruído não uniforme dos *pixels* (PNU) com informações de demosaico coletadas a partir da imagem descritas anteriormente. Os autores reportam uma melhoria de 17% na acurácia no cenário multi-classe.

Uma desvantagem do método baseado em informações do ruído PRNU para o cenário forense é que sua detecção é condicionada a uma operação de sincronização apropriada. Uma pequena modificação de escala ou recorte na imagem pode levar a uma detecção incorreta [Goljan et al. 2008]. Transformações geométricas (e.g., escala e rotação) causam dessincronização e introduzem distorções devido à reamostragem.

Neste sentido, [Goljan et al. 2008] apresentam uma extensão ao trabalho original de [Lukas et al. 2006] para um cenário mais geral em que a imagem sob investigação tenha sofrido alguma operação de recorte e/ou escala. Antes de fazer a comparação dos padrões de referência, os autores empregam uma

etapa de força bruta para identificar os parâmetros de escala da imagem analisada. Em seguida, os autores utilizam métricas de estimação da correlação de pico (*Peak to Correlation Energy*, PCE) e correlação cruzada normalizada (*Normalized Cross-correlation*, NCC) entre os padrões de referência da imagem redimensionada e da câmera para estimar os parâmetros de recorte[10]. Isso é feito até um critério de parada ser atingido. Os autores reportam bons resultados para imagens com até 50% de redimensionamento e até 90% de área recortada.

Embora o trabalho de [Goljan et al. 2008] seja importante para nos conduzir a um cenário forense mais confiável na identificação da câmera que capturou uma imagem, é importante notar que: (1) a qualidade da resposta depende, em parte, do conteúdo da imagem e nível de compactação (e.g., JPEG); (2) é um procedimento computacionalmente intensivo, uma vez que precisamos de força bruta para localizar os parâmetros de escala. Neste sentido, visualizamos aqui mais uma direção de pesquisa que merece investigação. Poderíamos pensar em maneiras alternativas à força bruta para localizar os parâmetros de escala e recorte. Os autores [Goljan et al. 2008] propõem uma busca hierárquica para este fim mas não desenvolvem a idéia nem apresentam resultados.

6.4.1.3. Identificação de *scanners*

Recentes avanços nos dispositivos de captura de imagens analógicas via *scanners* de alta resolução trouxeram a necessidade de ferramentas forenses de identificação igualmente avançadas. Com a computação tornando-se mais e mais ubíqua a cada dia, não é incomum vermos uma imagem escaneada praticamente idêntica a uma fotografia original. Por outro lado, o processo de fabricação de qualquer equipamento de captura, seja uma câmera ou um *scanner*, introduz vários defeitos nos sensores de imageamento e, consequentemente, cria ruído no processo de aquisição dos pixels de uma imagem. Nesse sentido, vários pesquisadores têm procurado desenvolver técnicas forenses para identificação de *scanners* com relativo sucesso.

Para um melhor entendimento do funcionamento dessas técnicas, a Figura 6.10 apresenta o conjunto de estágios básico para a aquisição de uma imagem a partir de um *scanner* de mesa (*flatbed*) [Tyson 2001]. O documento é colocado no *scanner* e o processo de escaneamento começa. A lâmpada (fonte de luz) utilizada para iluminar o documento é do tipo fluorescente de cátodo frio (*cold cathode fluorescent lamp*, CCFL) ou xenon. Utilizando um estabilizador, uma correia e um motor de passo, a cabeça de escaneamento passa de forma linear sobre a imagem para capturá-la. O objetivo do estabilizador é garantir o movimento suave (sem desvios ou trepidações) da cabeça de escaneamento com relação ao documento. A cabeça de escaneamento

[10] Consulte o Apêndice A.3 para uma definição formal.

possui um conjunto de lentes, espelhos, filtros e o sensor de imageamento. A maioria dos *scanners* de mesa utiliza um sensor CCD ou CMOS. A resolução máxima do *scanner* é determinada pela resolução vertical e horizontal. O número de elementos no sensor CCD linear determina a resolução ótica horizontal. O tamanho do passo da cabeça de escaneamento dita a resolução vertical [Tyson 2001].

Existem duas maneiras básicas de se conseguir um escaneamento com uma resolução abaixo da resolução padrão de um *scanner*. Uma maneira é sub-amostrar o sensor de imageamento de modo a capturar apenas os *pixels* de interesse. Por exemplo, para escanear um documento a 600 DPIs em um *scanner* nativo de 1200 DPIs, basta amostrar apenas os *pixels* ímpares do sensor CCD. A segunda maneira consiste em escanear o documento na resolução nativa e fazer a redução à resolução desejada na memória do *scanner*. A maior parte dos bons *scanners* utiliza a segunda forma.

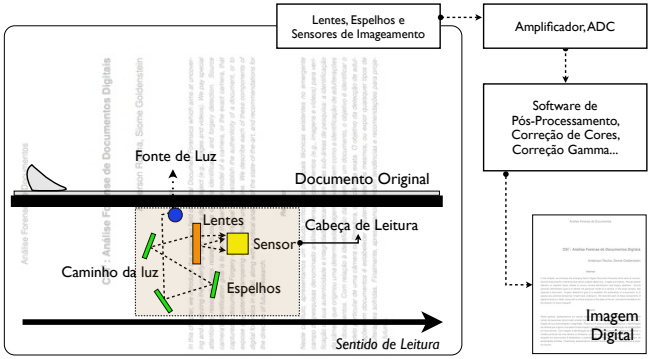


Figura 6.10. Possível *pipeline* do processo de aquisição de uma imagem (via *scanner*).

Gloe et al. [Gloe et al. 2007a] apresentam uma extensão do trabalho de [Lukas et al. 2006] para identificação de *scanners*. Experimentos com cinco *scanners* mostraram bons resultados. Os autores mostraram que, para imagens escaneadas em um padrão nativo, um padrão de referência 2-D produz melhores resultados. Por outro lado, para imagens escaneadas em resoluções não nativas, os autores mostraram que os padrões de referência 1-D são mais apropriados. Uma explicação possível é que pequenas perturbações tais como poeira no sensor e arranhões na placa de vidro nos padrões de referência 2-D são eliminadas pelas operações de redimensionamento na resolução não nativa.

Gou et al. [Gou et al. 2007] apresentam outra abordagem para identificação

de *scanners* baseada em padrões de ruído. Os autores utilizam três conjuntos de características extraídos das imagens escaneadas. Esta abordagem busca classificar o modelo do *scanner* em investigação e não o dispositivo exato. Infelizmente, os autores mostram resultados de treinamento e teste apenas para um conjunto limitado de imagens (< 50 imagens) e nenhum resultado conclusivo é possível a partir desses experimentos.

Recentes avanços na literatura científica descrevem uma técnica computacional para determinar o *scanner* de procedência de uma imagem [Khanna et al. 2009]. Esta técnica é usada para decidir qual equipamento, dentre um conjunto limitado, deu origem a uma imagem de teste.

Diferente das câmeras digitais, *scanners* usam um sensor de captura unidimensional. Essa constatação levou [Khanna et al. 2009] a propor a construção de um padrão de referência do *scanner* utilizando-se a média das linhas da estimativa de ruído da imagem. O padrão de ruído linear de uma imagem é conseguido fazendo-se a média de todas as linhas da imagem representando o ruído estimado da imagem sob investigação.

O padrão de referência de um *scanner* em particular (assinatura) é conseguido fazendo-se a média dos padrões de referências de múltiplas imagens escaneadas pelo mesmo *scanner*. Para identificar o *scanner* que capturou uma determinada imagem, comparamos seu padrão de referência com a assinatura dos *scanners* conhecidos em nossa base de conhecimento. O *scanner* que produzir a maior correlação é escolhido.

Para um melhor entendimento, considere I^k a k-ésima imagem de entrada de tamanho $M \times N$ *pixels* (M linhas, N colunas). Seja I^k_{noise} o ruído correspondente à imagem de entrada I^k. Seja $I^k_{denoised}$ o resultado da utilização de um filtro de redução/supressão de ruído na imagem I. Assim,

$$I^k_{noise} = I^k - I^k_{denoised} \tag{15}$$

Seja K o número de imagens utilizadas para a obtenção do padrão de referência de um *scanner* em particular. Dessa forma, o padrão de referência (2-D) do *scanner* é obtido como:

$$\widetilde{I}^{array}_{noise}(i,j) = \frac{1}{K}\sum_{k=1}^{K} I^k_{noise}(i,j); \qquad 1 \le i \le M \text{ e } 1 \le j \le N \tag{16}$$

Em seguida, calculamos o padrão de referência 1-D ou assinatura do *scanner* fazendo a média das linhas:

$$\widetilde{I}^{linear}_{noise}(1,j) = \frac{1}{M}\sum_{i=1}^{M} \widetilde{I}^{array}_{noise}(i,j); \qquad 1 \le j \le N. \tag{17}$$

Podemos utilizar a correlação entre a assinatura de um *scanner* e o padrão de referência de uma imagem para determinarmos a origem. A correlação entre dois vetores $X, Y \in \Re^N$ é definida como

$$C(X,Y) = \frac{(X - \overline{X}) \cdot (Y - \overline{Y})}{\|X - \overline{X}\| \cdot \|Y - \overline{Y}\|}. \tag{18}$$

Uma das grandes diferenças entre os sensores de captura em uma câmera e em um *scanner* é que em um *scanner* apenas partes do sensor são utilizadas dependendo do tamanho da imagem a ser capturada bem como sua posição na placa de vidro. Neste sentido, para a abordagem anterior ter sucesso, tanto a imagem sob investigação quanto as imagens de treinamento utilizadas para a construção da assinatura dos *scanners* têm que ter sido capturadas nas mesmas condições (tamanho e posição na placa de vidro). Infelizmente, isso não acontece na prática.

Khanna et al. [Khanna et al. 2009] propõem resolver esse problema calculando estatísticas básicas sobre o vetor de assinatura da imagem em análise. Para isso, os autores calculam a média das linhas e colunas na imagem $\widetilde{I}^{l}_{noise}$ e $\widetilde{I}^{c}_{noise}$. Em seguida, calcula-se a correlação $\rho_l(i)$ entre a média de todas as linhas ($\widetilde{I}^{l}_{noise}$) e a i-ésima linha de I^{noise}. Similarmente, calcula-se a correlação $\rho_c(j)$ entre e média de todas as colunas ($\widetilde{I}^{c}_{noise}$) e a j-ésima coluna de I_{noise}. Finalmente, os autores calculam estatísticas (e.g., média, variância, moda, curtose, etc.) sobre ρ_r, ρ_c, $\widetilde{I}^{r}_{noise}$ e $\widetilde{I}^{c}_{noise}$. Os vetores de descrição são fornecidos a um classificador de padrões. Os autores reportam resultados acima de 90% de acurácia para um cenário com 11 *scanners* analisados.

6.4.1.4. Técnicas contra-forenses na identificação de sensores

Tal como em qualquer outro campo de pesquisa forense, as técnicas de identificação dos dispositivos de captura de um objeto digital também estão suscetíveis às técnicas contra-forenses.

Gloe et al. [Gloe et al. 2007b] apresentam duas técnicas contra-forenses para manipular as informações de aquisição e identificação de dispositivos de captura discutidos em [Lukas et al. 2006]. Em seu trabalho, os autores observam que a utilização de um filtro de supressão/redução de ruído baseado em *wavelets* tal como utilizado em [Lukas et al. 2006] não é suficiente para criar uma imagem de qualidade e eliminar toda a informação de ruído necessária para a criação da assinatura dos dispositivos de captura. Como nem toda a informação de ruído é eliminada, um método conhecido como *flatfielding* pode ser aplicado de modo a estimar resquícios do ruído de padrão fixo (FPN) e do ruído de foto-responsividade não uniforme (PRNU). Como discutido antes neste capítulo, FPN é um ruído aditivo independente do sinal enquanto PRNU é um ruído multiplicativo dependente da fonte originadora. Para estimar o ruído do tipo FPN, pode-se utilizar um quadro preto $I_{dark\_estimate}$ representando a média de J imagens I_{dark} capturadas no escuro (e.g., sem retirar a tampa da lente).

$$I_{dark\_estimate} = \frac{1}{J}\sum_{J} I_{dark}. \tag{19}$$

Para a estimativa do ruído do tipo PRNU, são necessárias K imagens de uma cena homogeneamente iluminada I_{light} com a estimativa $I_{dark\_estimate}$

subtraída. Para estimar o quadro de *flatfield* $I_{flatfield}$, calcula-se a média destas imagens

$$I_{flatfield} = \frac{1}{K} \sum_{K} (I_{light} - I_{dark\_estimate}). \tag{20}$$

Tendo-se uma estimativa dos ruídos do tipo FPN e PRNU de uma câmera, um indivíduo mal-intencionado pode suprimir as características de ruído de uma imagem de uma câmera em particular para evitar quaisquer traços de identificação de origem. Uma imagem $\hat{I}$ com a assinatura de ruído retirada pode ser criada minimizando-se

$$\hat{I} = \frac{I - I_{dark\_estimate}}{I_{flatfield}}. \tag{21}$$

Felizmente, os autores argumentam que o efeito de *flatfielding* não pode ser facilmente estimado. A dificuldade está no grande número de parâmetros que precisam ser levados em conta (tempo de exposição, velocidade de captura, ISO, etc.) para gerar as estimativas $I_{dark\_estimate}$ e $I_{flatfield}$. Entretanto, fixando apenas um parâmetro, os autores mostram resultados contra-forenses convincentes para imagens em formato RAW (sem processamento algum) e TIFF.

Após a utilização da técnica acima, um outro ataque forense possível consiste em extrair a assinatura de ruído de uma outra câmera e substituir a assinatura da câmera verdadeira. O padrão de ruído de uma câmera pode ser substituído utilizando-se a operação de *flatfielding* inverso. Uma imagem com $\hat{I}_{forge}$ com assinatura de ruído falsificada pode ser criada a partir de informações pré-computadas de qualquer câmera

$$\hat{I}_{forge} = \hat{I} \cdot I_{flatfield\_forge} + I_{dark\_forge}. \tag{22}$$

Os autores também reportam resultados interessantes com esta técnica.

6.4.2. Identificação de criações sintéticas

Distinguir entre uma imagem natural e uma imagem feita em computador pode ser crucial em algumas situações. Por exemplo, segundo a lei americana, a possessão de imagens de menores de idade é considerada crime de pedofilia. Se estas imagens forem geradas em computador, não há crime algum [Rocha et al. 2011]. Entretanto, uma imagem pode ser copiada em computador e alterada de modo que seus traços se pareçam como se ela tivesse sido feita totalmente em computador [Farid 2007, Lyu 2005].

De forma geral, as abordagens para separar imagens geradas em computador de imagens naturais tem considerado: decomposição da imagem em filtros de quadratura em espelho e subsequente análise estatística dos artefatos de decomposição [Lyu 2005]; diferenças dos modelos de superfície em imagens naturais e geradas em computador [Ng et al. 2005], análise da presença de ruído de aquisição em imagens naturais e ausência em

imagens geradas em computador [Dehnie et al. 2006]; análise do comportamento de imagens naturais e geradas em computador mediante a sucessivas perturbações [Rocha e Goldenstein 2010, Rocha e Goldenstein 2006, Rocha e Goldenstein 2007]; e artefatos resultantes da operação de demosaico em imagens naturais e sua ausência em imagens sintéticas [Dirik et al. 2007].

Apresentada por Lyu e Farid no contexto de detecção de mensagens escondidas em imagens e depois aplicada no contexto de separação de imagens geradas em computador e imagens naturais [Lyu e Farid 2002, Lyu e Farid 2004, Lyu 2005, Lyu e Farid 2005], esta abordagem de detecção consiste na construção de modelos estatísticos de alta ordem para imagens naturais e na busca por desvios nestes modelos.

As imagens naturais possuem regularidades que podem ser detectadas com estatísticas de alta ordem através de uma decomposição *wavelet*, por exemplo [Lyu e Farid 2002]. O processo de criação de uma imagem em computador insere artefatos estatísticos fazendo com que seja possível separar essa classe de imagens de imagens naturais. Após a construção dos modelos, é necessário utilizarmos classificadores capazes de dizer se uma dada imagem é natural ou gerada em computador.

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.

Os autores aplicam uma decomposição baseada nos *filtros de quadratura em espelho* (QMFs – *Quadrature Mirror Filters*) [Vaidyanathan 1987]. Esta decomposição divide a imagem no domínio da frequência em múltiplas escalas e orientações. Esta decomposição é feita aplicando-se filtros de passa-baixas e passa-altas sobre a imagem gerando quatro sub-bandas: *vertical, horizontal, diagonal* e de *passa-baixas*. Escalas subsequentes são criadas aplicando-se o processo novamente sobre a sub-banda de *passa-baixas*.

A partir desta decomposição da imagem, os autores propõem um modelo estatístico composto por dois conjuntos de descritores. O primeiro conjunto consiste em descritores como média, variância, moda e curtose calculados sobre os histogramas dos coeficientes das sub-bandas. O segundo conjunto de características é composto por estatísticas de alta ordem calculadas sobre os erros de um preditor linear de coeficientes de magnitude. Um preditor linear de erro consiste na combinação de um *pixel* com seus vizinhos em escalas e orientações diferentes. Para um maior entendimento, considere a sub-banda vertical, $V_i(x, y)$, na escala i. Um preditor linear para a magnitude destes coe-

ficientes em um subconjunto de todos os possíveis vizinhos é dado por

$$\begin{aligned} V_i(x,y) &= w_1 V_i(x-1,y) + w_2 V_i(x+1,y) + w_3 V_i(x,y-1) \\ &\quad + w_4 V_i(x,y+1) + w_5 V_{i+1}(\frac{x}{2},\frac{y}{2}) + w_6 D_i(x,y) \\ &\quad + w_7 D_{i+1}(\frac{x}{2},\frac{y}{2}), \end{aligned} \tag{23}$$

onde w_k denota os valores escalares de peso dos coeficientes. Os coeficientes do erro são calculados utilizando-se uma função de minimização do erro

$$E(w) = [V - Qw]^2, \tag{24}$$

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 23.

Ao final, o modelo possui dois conjuntos de valores de descritores (diretos e resultantes do preditor), cada um com quatro descritores aplicados em três orientações (vertical, horizontal, diagonal) e n escalas resultando em $F = 2 \times 3 \times 4 \times s = 72$, para $s = 3$ escalas. Este vetor de características deve ser utilizado em um classificador de padrões para a elaboração do resultado final. Lyu e Farid mostraram que esse modelo foi capaz de classificar 67% de imagens geradas em computador enquanto a taxa de classificação errada foi mantida fixa em 1% para um cenário com 40.000 imagens naturais e 6.000 imagens geradas em computador.

Em seu trabalho [Rocha e Goldenstein 2007, Rocha e Goldenstein 2010], Rocha e Goldenstein apresentam um novo meta-descritor de imagens denominado Randomização Progressiva (PR) para o contexto de categorização de imagens. Uma das aplicações do meta-descritor apresentado é a separação entre imagens naturais e imagens geradas em computador. PR é um meta-descritor que captura as diferenças entre classes gerais de imagens usando os artefatos estatísticos inseridos durante um processo de perturbação sucessiva das imagens analisadas. A observação mais importante é que classes diferentes de imagens possuem comportamentos distintos quando submetidas a sucessivas perturbações.

Uma perturbação pode ser definida como a alteração de alguns *pixels* selecionados na imagem de acordo com alguma sequência de *bits*. Para inserir a perturbação, basta alterar os valores dos *bits* menos significativos (*Least Significant Bits*, LSBs) de alguns *pixels*. As $T(I, P_i)$ transformações são perturbações de diferentes porcentagens (pesos) nos LSBs disponíveis. No trabalho base, os autores utilizam $n = 6$ perturbações onde $P = \{1\%, 5\%, 10\%, 25\%, 50\%, 75\%\}$, $P_i \in P$ denota os tamanhos relativos dos conjuntos de *pixels* selecionados para terem seus LSBs alterados. A Figura 6.11 mostra um exemplo de perturbação para uma sequência de *pixels* $B = 1110$. Como a sequência de perturbação possui quatro *bits*, selecionamos quatro *pixels* na imagem. Para cada *pixel* selecionado, verificamos seu

bit menos significativo. Caso a imagem seja colorida, essa seleção pode levar em conta os canais de cor. Nesse caso, cada *pixel* possui três *bits* menos significativos, um para cada canal de cor. Caso o *pixel* selecionado tenha LSB igual ao LSB com o qual estamos fazendo a perturbação, este *pixel* permanece inalterado.

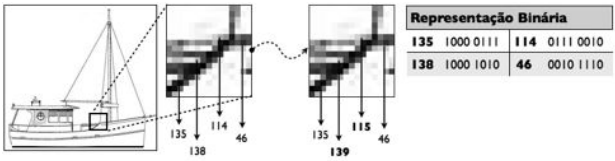


Figura 6.11. Um exemplo de perturbação LSB usando os *bits* $B = 1110$.

A descrição das imagens é feita utilizando-se uma abordagem baseada em regiões e, para cada região, calculando-se descritores estatísticos tais como χ^2 (teste do Chi-quadrado) [Westfeld e Pfitzmann 1999] e U_T (teste Universal de Ueli Maurer) [Maurer 1992]. Os descritores, de forma geral, são calculados sobre histogramas de *Pares de Valores* (PoVs). Para entender melhor, considere um *pixel* com L *bits* representando 2^L valores possíveis. Quando consideramos possíveis mudanças apenas no canal LSB, temos 2^{L-1} classes de invariância. Chamamos estas classes de invariância de *Pares de Valores*. Por exemplo, quando perturbamos todos os LSBs disponíveis em uma imagem com uma sequência B, a distribuição de valores 0/1 de um par de valor será a mesma da distribuição 0/1 em B.

A seleção de regiões pode ser tão simples quanto a seleção de quadrantes sem sobreposição, quanto pode utilizar filtros que localizam porções da imagem com maior riqueza de detalhes [Rocha e Goldenstein 2010]. Após a aplicação das n perturbações, e análise dos dois descritores nas r regiões, cada imagem produz $F = n \times r \times 2$ valores. Esses valores podem ser passados a um classificador de padrões já treinado para efetuar a classificação da imagem analisada. O treinamento desse classificador consiste em utilizar vetores de características resultantes do processo Randomização Progressiva (F_i) para imagens representando cada classe a ser categorizada. Por exemplo, se queremos diferenciar imagens geradas em computador e imagens naturais, utilizamos, no processo de treinamento do classificador, um conjunto F_{CGI} de vetores de características provenientes de um conjunto de imagens geradas em computador e um conjunto F_{Nat} de vetores de características provenientes de um conjunto de imagens naturais. Os autores reportam resultados na faixa de 90% de acurácia para um cenário com mais de 40.000 imagens naturais e 5.000 imagens geradas em computador.

Ng et al. [Ng et al. 2005] apresentam uma técnica para separação de imagens naturais e geradas em computador motivada pelas diferenças físicas nos

processos de captura e geração de tais imagens. Os autores desenvolveram dois níveis de separação: (1) autenticidade a nível de processamento e (2) autenticidade a nível de cena. Autenticidade a nível de processamento compreende as imagens capturadas por um sensor de captura (e.g., *scanner* ou câmera). Autenticidade de cena é definida como um instantâneo de um campo físico de luz. Os autores apresentam uma série de características para dar suporte ao modelo tais como:

1. *Dimensão fractal local*; para capturar a complexidade de texturas em fotografias.
2. *Vetores de patches locais*; para capturar características de arestas e bordas.
3. *Superfície gradiente*; para capturar a forma de resposta de uma câmera.
4. *Geometria quadrática local*; para capturar artefatos devido ao modelo poligonal utilizado por objetos computadorizados.
5. *Vetor de fluxo de Beltrami*; para capturar artefatos devido à suposição de independência de cores em computação gráfica.

As características acima são consideradas em conjunto e produzem um vetor de descrição utilizado em um classificador de padrões como o SVM. Os autores reportam um acerto de aproximadamente 84% para um cenário com 3.200 imagens.

Características específicas da câmera de captura, tais como as que vimos na Seção 6.4.1.2, também podem ser utilizadas para distinção entre imagens naturais e sintéticas. Revisitando o trabalho de [Lukas et al. 2006], Dehnie et al. [Dehnie et al. 2006] utilizam as características do ruído para distinguir imagens geradas em computador e imagens naturais. A idéia é que mesmo que diferentes câmeras possuam diferentes características de ruído durante o processo de captura, ainda existem propriedades estatísticas que permanecem ao longo de diferentes câmeras tornando-se possível a separação destas como um todo de uma imagem sintética. O problema com essa abordagem aparece quando os falsificadores criam um conteúdo sintético a partir de imagens naturais alterando apenas propriedade localizadas.

Dirik et al. [Dirik et al. 2007] associam características decorrentes do algoritmo de demosaico [Popescu e Farid 2005b, Bayram et al. 2005, Swaminathan et al. 2006] com características referentes à presença de aberrações cromáticas para melhorar a qualidade de detecção da presença de algum algoritmo de demosaico utilizado em uma imagem. Os autores reportam resultados acima de 90% de acurácia.

6.4.2.1. Técnicas contra-forenses na identificação de imagens sintéticas

Os métodos para distinção entre imagens naturais e sintéticas também estão sujeitos a ataques contra-forenses. Uma medida simples que pode ser

tomada por um agressor consiste na recaptura da imagem utilizando uma câmera digital [Ng et al. 2005].

Ng et al. [Ng et al. 2005] buscam resolver esse problema utilizando esse tipo de dado na etapa de treinamento do classificador escolhido. Yu et al. [Yu et al. 2008] apresentam outra técnica para detecção de ataques de recaptura. A motivação para esse trabalho é que a especularidade de uma fotografia recapturada é modulada pela mesoestrutura da superfície da fotografia. Assim, a sua distribuição espacial pode ser usada para a classificação.

Assim como os sistemas de identificação de câmeras, as técnicas contraforenses ainda estão em sua infância, e nós esperamos encontrar ataques mais sofisticados em um futuro próximo.

6.4.3. Identificação de adulterações

O maior objetivo em análise forense de documentos consiste na detecção de adulterações em documentos digitais. Tipicamente, documentos (ou suas partes) tais como imagens sofrem uma ou mais manipulações digitais: operações afins (e.g., aumento, redução, rotação), compensação de cor e brilho, supressão ou modificação de detalhes (e.g., filtragem, adição de ruído, compressão). Embora muitas operações de adulteração gerem documentos sem artefatos visuais, elas afetam as estatísticas inerentes dos mesmos [Rocha et al. 2011, Sencar e Memon 2008].

As abordagens propostas na literatura para resolver este problema ainda estão em seus primórdios [Sencar e Memon 2008]. A análise forense de documentos digitais é recente e seus principais trabalhos foram publicados a partir de 2004. De forma geral, podemos agrupar as abordagens propostas em:

1. Técnicas de detecção de clonagem;
2. Técnicas que analisam variações em descritores de características;
3. Técnicas que analisam inconsistências em descritores de características;
4. Técnicas que analisam inconsistências relacionadas ao processo de aquisição;
5. Técnicas que analisam inconsistências de iluminação;
6. Técnicas que analisam inconsistências de compressão;

6.4.3.1. Técnicas de detecção de clonagem

Clonagem é uma das operações de adulteração mais simples que uma imagem ou vídeo pode sofrer. Também conhecida como cópia/colagem, esta operação está presente em operações mais sofisticadas tais como o retoque e conciliação (c.f., Seção 6.3).

O objetivo mais comum da operação de clonagem é fazer com que um objeto em uma cena “desapareça” utilizando propriedades da própria cena tais

como padrões de textura e cor na vizinhança do objeto em questão. Por utilizar elementos da própria cena para eliminar detalhes da mesma, a operação de clonagem é tecnicamente simples de detectar utilizando-se busca exaustiva. No entanto, soluções de força bruta são computacionalmente caras.

Fridrich et al. [Fridrich et al. 2003] apresentam uma técnica para detecção rápida de regiões duplicadas em imagens. Os autores utilizam uma janela deslizante sobre a imagem e calculam, para cada bloco de *pixels*, a transformada discreta do cosseno (DCT).

O conjunto de coeficientes resultantes de cada aplicação de DCT é armazenado como uma linha em uma matriz $A_{\mathcal{D}}$ de coeficientes. Os autores propõem utilizar a transformação quantizada para maior robustez e habilidade de fazer casamentos não exatos para regiões duplicadas. Ao aplicar as transformações sobre todos os possíveis blocos da imagem, os coeficientes são ordenados lexicograficamente. Em seguida, busca-se por linhas semelhantes. Para reduzir o número de falsos positivos, os autores propõem uma etapa de pós-processamento em que uma região é considerada duplicada se e somente se mais linhas da matriz partilham da mesma condição e são próximas no espaço da imagem.

Popescu e Farid [Popescu e Farid 2004a] apresentam uma abordagem semelhante trocando a transformada discreta do cosseno pela análise dos componentes principais (PCA) dos blocos. Resultados comparáveis foram reportados. A Figura 6.12 ilustra o processo. A partir de uma imagem de entrada, duplicamos o barco próximo ao coqueiro à esquerda. Como houve a duplicação de uma região, podemos utilizar a técnica para detecção de regiões duplicadas proposta por Fridrich et al. [Fridrich et al. 2003] ou Popescu e Farid [Popescu e Farid 2004a]. Ambas as técnicas se baseiam em uma análise por regiões feita sobre a imagem sob investigação. Nesta análise, uma janela deslizante é aplicada sobre a imagem. Para cada região de *pixels* sob a janela deslizante, aplicamos a sumarização desta região (e.g., PCA ou DCT). Em seguida, com cada região sendo representada por um conjunto de coeficientes resultantes da sumarização, ordenamos estes conjuntos de coeficientes lexicograficamente. Por exemplo, a região $r_m = \langle 18, 25, 5, 4 \rangle$ vem antes de uma região $r_n = \langle 18, 25, 6, 1 \rangle$. Finalmente, analisamos os blocos ordenados para detectar eventuais duplicações. Regiões lexicograficamente ordenadas que estejam muito próximas podem indicar a existência de uma duplicação. No entanto, o investigador precisa levar em consideração que regiões de valor muito próximo também podem se referir a regiões homogêneas parecidas na imagem em questão.

Um dos problemas das abordagens anteriores é sua complexidade para ser utilizada diretamente na detecção de regiões duplicadas em vídeos [Wang e Farid 2007].

Podemos definir a clonagem em um vídeo da seguinte forma: dado um par de quadros (*frames*) de um vídeo $I(x, y, \tau_1)$ e $I(x, y, \tau_2)$, provenientes de uma câmera estacionária ou não, o objetivo é estimar o deslocamento espacial

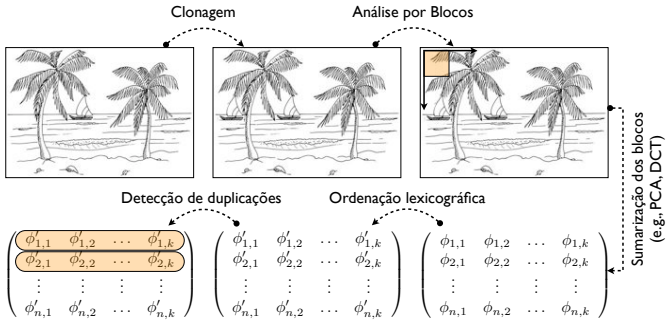


Figura 6.12. Detecção de regiões duplicadas em imagens.

(Δ_x, Δ_y) correspondente a uma região duplicada de um quadro colocada em outro quadro em uma posição diferente.

Wang e Farid[Wang e Farid 2007] apresentam uma técnica para detectar tais operações em câmeras estacionárias utilizando estimativas de correlação de fase [Castro e Morandi 1987]. Para isso, primeiro definimos o espectro de potência cruzado normalizado (*normalized cross power spectrum*)

$$\Psi(\omega_x, \omega_y) = \frac{F(\omega_x, \omega_y, \tau_1) F^*(\omega_x, \omega_y, \tau_2)}{\|F(\omega_x, \omega_y, \tau_1) F^*(\omega_x, \omega_y, \tau_2)\|}, \tag{25}$$

onde $F(\cdot)$ é a transformada de Fourier de um quadro, $*$ é o complexo conjugado, e $\|\cdot\|$ é a magnitude.

Técnicas baseadas em correlação de fase estimam os deslocamentos espaciais analisando picos em $\psi(x, y)$, a inversa da transformada de Fourier de $\Psi(\omega_x, \omega_y)$. Um pico é esperado na origem $(0, 0)$ dado que a câmera em questão é estacionária. Quaisquer picos em outras posições são pistas para alinhamentos secundários que podem representar uma região duplicada. No entanto, tais picos também podem ser referentes a simples movimentos de translação no cenário com câmeras não estacionárias.

A localização espacial dos picos correspondem a deslocamentos espaciais candidatos (Δ_x, Δ_y). Para cada deslocamento candidato, a técnica calcula a correlação entre $I(x, y, \tau_1)$ e $I(x, y, \tau_2)$ para determinar se um deslocamento corresponde a uma determinada região duplicada.

Para isso, os autores dividem cada quadro em pequenas regiões de 16×16 *pixels* com sobreposição de um *pixel*. Em seguida, calcula-se o coeficiente de correlação entre cada par de blocos correspondentes. Blocos acima de um determinado limiar são marcados como duplicados.

Os autores também propõem uma possível extensão para câmeras não estacionárias. Para esse objetivo, calcula-se uma medida aproximada do movimento da câmera o qual deve ser compensado em tempo de execução do

algoritmo. Uma desvantagem dessa abordagem é que as operações de duplicação são simples cópias seguidas de colagem sem nenhuma sofisticação adicional tais como retoque ou conciliação.

6.4.3.2. Técnicas que analisam variações em descritores de características

Abordagens nesta categoria analisam descritores de imagens e vídeos sensíveis ao processo de adulteração e os comparam com o comportamento analisado e aprendido a partir de outras imagens/vídeos normais não alterados. Na maioria das vezes, estas soluções empregam classificadores no processo de decisão.

Alguns trabalhos relevantes consideram variações em: métricas de qualidade de imagens (IQMs[11]) [Avcibas et al. 2004]; bicoerência para análise das correlações de alta ordem em imagens [Ng e Chang 2004]; estatísticas de coeficientes wavelet (HOWS[12]) [Lyu 2005]; métricas de similaridade binária de imagens (BSM[13]) [Bayaram et al. 2005a]; IQMs, QMFs e BSMs combinados [Bayaram et al. 2006]; estatísticas de momento e de Markov [Shi et al. 2007].

Avcibas et al. [Bayaram et al. 2006] abordam a detecção de adulterações como um problema de classificação. Os autores argumentam que adulterações em imagens normalmente envolvem uma sequência de múltiplos passos, que frequentemente demandam uma sequência de operações de processamento de imagens mais simples tais como: escala, rotação, mudanças de contraste, suavização, etc. Neste sentido, os autores desenvolvem um conjunto de classificadores *experts* em detectar cada uma das operações elementares. Ao final, os resultados são combinados de modo a produzir uma resposta mais confiável. As características de descrição das imagens utilizadas no processo de treinamento variam desde métricas de qualidade de imagem [Acbibas et al. 2003] e medidas de similaridade binária [Bayaram et al. 2005a] provenientes da literatura de esteganálise [Cox et al. 2007] a filtros de quadratura em espelho de alta ordem [Lyu e Farid 2004].

A maior limitação desta abordagem é que operações elementares de processamento de imagens em si não representam operações de adulteração de conteúdo. Um investigador forense precisa estar ciente dessas condições e utilizar tal técnica no sentido de localizar variações nas imagens que possam apontar para falsificações. Por exemplo, mudanças abruptas de brilho e contraste em uma imagem podem ser indicações de composição.

Ng et al. [Ng e Chang 2004] propõem um sistema de classificação binário baseado em estatísticas de alta ordem para detecção de composições de

[11] *Image Quality Metrics.*

[12] *High Order Wavelet Statistics.*

[13] *Binary Similarity Measures.*

imagens. Os autores fazem uso de características de bicoerência motivados pelo sucesso de tais características na identificação de composições em áudio [Nemer et al. 2001].

Bicoerência é a correlação de terceira ordem de três frequências harmonicamente relacionadas de Fourier de um sinal $\Xi(\omega)$ conhecido como bi-espectro normalizado. Os autores reportam um resultado de $\approx 71\%$ de acurácia no banco de dados de composição da Universidade de Colúmbia (*Columbia Splicing data set*) [Columbia DVMM Research Lab. 2004]. Uma limitação desta abordagem é que o cálculo das características de bicoerência é computacionalmente caro, frequentemente na ordem de $O(n^4)$ onde n é o número de *pixels* da imagem sendo investigada.

Shi et al. [Shi et al. 2007] apresentam um modelo para separar imagens normais de imagens resultantes de operações de composição. O modelo é representado por características extraídas de um conjunto de imagens e matrizes resultantes do cálculo da transformada de cosseno multi-escala por blocos (*multi-size block discrete cosine transform*, MBCT) sobre as imagens analisadas. Para cada matriz, os autores calculam os coeficientes de erro, suas subbandas *wavelet* e estatísticas de momentos uni- e bidimensionais. Os autores também calculam matrizes de transição probabilística de Markov. Embora efetiva para procedimentos simples de composição tais como as que fazem parte do banco de dados de imagens DVMM com 92% de acurácia, a abordagem não parece ser muito eficaz para composições mais sofisticadas que utilizam arestas adaptativas e propagação estrutural [Sun et al. 2005]. Isso se deve ao fato de que as matrizes de transição são frequentemente incapazes de capturar as mudanças sutis nas arestas resultantes da propagação estrutural. Adicionalmente, a abordagem proposta não é capaz de apontar a região onde provavelmente ocorreu a operação de composição.

6.4.3.3. Técnicas que analisam inconsistências em descritores de características

Abordagens nesta categoria analisam inconsistências a respeito de um determinado conjunto de descritores ao longo de uma imagem ou vídeo. Estas inconsistências podem ser desvios abruptos de um ponto a outro ou a presença de similaridades inesperadas ao longo do objeto analisado.

Alguns trabalhos relevantes têm considerado inconsistências inseridas por: presença de artefatos devido a dupla compressão JPEG [He et al. 2006, Popescu 2004]; correlação linear periódica devido a reamostragem [Popescu 2004], iluminação ambiente [Johnson 2007], reflexos oculares [Johnson e Farid 2007b] e presença de regiões repetidas nas imagens [Popescu 2004].

No momento da criação de uma imagem composta, frequentemente é necessário fazer a reamostragem de uma imagem em uma grade de amostragem (*lattice*) utilizando alguma técnica de interpolação (e.g., bicúbica). Embora im-

perceptível, a reamostragem contem correlações específicas que, quando detectadas, podem representar evidências de adulteração.

Popescu e Farid [Popescu e Farid 2005a] descrevem a forma destas correlações e propõem um algoritmo para detectá-las. Os autores mostram que a forma específica das correlações pode ser determinada achando-se o tamanho da vizinhança, Φ, em que ocorre a combinação dos *pixels* e o conjunto de coeficientes, $\vec{\beta}$, representando os parâmetros dessas combinações. Tanto Φ quanto $\vec{\beta}$ devem satisfazer a restrição

$$\vec{M}_i = \sum_{j=-\Phi}^{\Phi} \beta_j \vec{M}_{i+j} \tag{26}$$

na equação

$$\left(\vec{M}_i - \sum_{j=-\Phi}^{\Phi} \beta_j \vec{M}_{i+j} \right) \cdot \vec{\mu} = 0, \tag{27}$$

onde $\vec{\mu}$ é o sinal analisado e $\vec{M}_i$ é a i-ésima linha da matriz de reamostragem.

Na prática, os autores apontam que nem as amostras que são correlacionadas nem a forma específica das correlações são conhecidas. Os autores propõem utilizar um algoritmo de Maximização de Esperança (EM) similar ao discutido na Seção 6.4.1.1 no contexto de identificação de câmeras para, simultaneamente, estimar um conjunto de amostras correlacionadas com seus *pixels* vizinhos bem como uma aproximação para a forma destas correlações.

Os autores assumem que cada amostra pertence a um de dois modelos possíveis. O primeiro modelo, $\mathcal{M}_1$, corresponde às amostras s_i que são correlacionadas com seus *pixels* vizinhos e são geradas a partir do seguinte modelo

$$\mathcal{M}_1 : s_i = \sum_{k=-\Phi}^{\Phi} \beta_k s_{i+k} + \mathcal{N}(i), \tag{28}$$

onde $\mathcal{N}(i)$ denota amostras independentes e identicamente distribuídas de uma distribuição normal com média zero e variância desconhecida σ^2. No passo E do método, a probabilidade de cada amostra s_i pertencer ao modelo $\mathcal{M}_1$ pode ser estimada utilizando-se o teorema de Bayes similar à Equação 9, Seção 6.4.1.1, onde s_i substitui $f(x, y)$.

Na abordagem proposta, assume-se que a probabilidade de observação de amostras geradas pelo modelo alternativo, $\Pr\{s_i | s_i \in \mathcal{M}_2\}$, é uniformemente distribuída sobre o intervalo de valores possíveis de s_i. No passo M da abordagem, a forma específica das correlações entre amostras é estimada minimizando-se uma função de erro quadrática.

É importante ressaltar que a reamostragem em si não constitui um ato de adulteração. Um indivíduo poderia, simplesmente, buscar economia de espaço ao reamostrar todas as imagens de sua coleção particular para a metade da resolução original. No entanto, quando diferentes correlações estão presentes na mesma imagem, um investigador forense tem em mãos uma forte evidência

de adulteração de imagem por composição. Os autores reportam resultados promissores para imagens com baixa compressão. À medida em que a taxa de compressão aumenta, a eficácia do método diminui.

Os autores afirmam que a generalização do algoritmo proposto para imagens coloridas é simples e propõem a análise de cada canal de cor independentemente. Entretanto, os autores não mostram experimentos sob estas condições.

Um possível contra-ataque para a técnica acima foi proposto por [Gloe et al. 2007b]. Os autores propõem antecipar a detecção dos traços de reamostragem. Para isso, o método proposto procura destruir as correlações dos *pixels* fazendo uso de pequenas distorções geométricas super impondo um vetor aleatório de perturbação sobre cada posição de *pixel*. Para lidar com possíveis problemas de *jitter*[14], os autores apresentam uma abordagem adaptativa com relação ao conteúdo da imagem.

6.4.3.4. Técnicas que analisam inconsistências relacionadas ao processo de aquisição

Abordagens nesta categoria analisam inconsistências relacionadas ao processo de aquisição das imagens. Tais características também podem ser usadas para inferir o dispositivo que capturou a imagem ou vídeo em análise [Sencar e Memon 2008].

Alguns trabalhos relevantes têm considerado inconsistências decorrentes da: interpolação CFA [Popescu e Farid 2005b]; padrão inerente de ruídos [Lukas et al. 2007]; função de resposta não linear das câmeras digitais e subsequentes inconsistências nas arestas [Lin et al. 2005].

Zhouchen et al. [Lin et al. 2005] apresentam uma abordagem para identificação de falsificações em imagens baseada na análise de consistência/inconsistência das funções de resposta da câmera que capturou a imagem sob investigação. Uma imagem é apontada como adulterada se as funções de resposta são anormais ou inconsistentes umas com as outras. A função de resposta da câmera é um mapeamento entre a irradiância de um *pixel* e o valor do *pixel* após a aquisição. Por exemplo, suponha que um *pixel* esteja em uma aresta e a radiância da cena muda ao longo da aresta mas é constante em ambos os lados da mesma (Figura 6.13(a)). Assim, a irradiância do *pixel* na aresta deveria ser uma combinação linear dos *pixels* fora da aresta (Figura 6.13(b)). No entanto, devido à não linearidade da função de resposta da câmera, esta relação "linear" é quebrada durante a leitura dos valores destes *pixels* (Figura 6.13(c)). Em seu trabalho, os autores estimam o relacionamento linear original calculando a função inversa de resposta da câmera [Lin et al. 2004].

[14] *Jitter* pode ser entendido como o desvio ou deslocamento de algum aspecto de um sinal digital. Por exemplo, em uma imagem, *Jitter* produz um efeito de tremor.

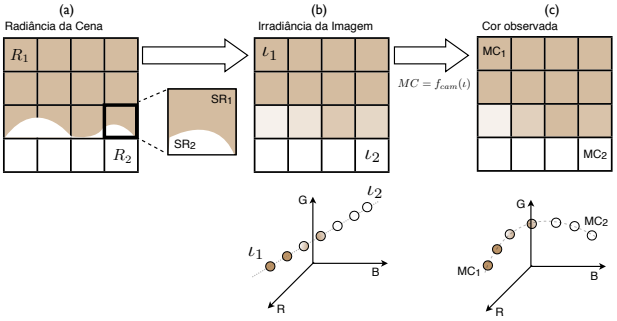


Figura 6.13. Estimativa da função de resposta da câmera. (a) R_1 e R_2 são duas regiões com radiância constante. A terceira linha representa uma combinação de R_1 e R_2. (b) As irradiâncias dos *pixels* em R_1 são mapeadas para o mesmo ponto τ_1, no espaço de cores RGB. O mesmo ocorre para os *pixels* em $R_2 \rightarrow \tau_2$. Entretanto, as cores dos *pixels* na terceira linha são o resultado de combinações lineares de τ_1 e τ_2. (c) A função de resposta da câmera f_{cam} deforma o segmento de linha em (b) em uma curva durante a leitura.

A limitação da técnica proposta está na dificuldade de se calcular a função inversa de resposta da câmera. Para isso, é necessário calcular uma função inversa que requer o aprendizado sobre um modelo de mistura de Gaussianas proveniente de um banco de dados contendo diversas funções de resposta conhecidas (DoRF[15]) [Lin et al. 2005, Grossberg e Nayar 2010]. Se a imagem analisada for uma composição de regiões provenientes de câmeras não presentes no banco de dados, o modelo será incapaz de identificar a função de resposta apropriada. Uma outra limitação do método é a necessidade de interação do usuário para marcar pontos em arestas candidatas de composição. Além disso, é possível que a abordagem não funcione com câmeras recentes presentes no meio comercial que fazem uso de sensores CMOS adaptativos capazes de dinamicamente calcular a função de resposta da câmera de modo a produzir fotografias mais agradáveis.

[15] *Dababase of camera response functions.*

6.4.3.5. Técnicas que analisam inconsistências de iluminação

Ao criar uma imagem composta (por exemplo, duas pessoas colocadas lado a lado), frequentemente é necessário casar as informações e condições de iluminação das fotografias individuais.

Neste sentido, Johnson e Farid [Johnson e Farid 2005] apresentam uma técnica para revelar traços de adulteração em imagens a partir de inconsistências de iluminação. Abordagens tradicionais para estimação da direção de iluminação assumem que a superfície em análise: (1) é Lambertiana (reflete a luz isotropicamente); (2) tem um valor constante de reflectância; (3) é iluminada por uma fonte localizada no infinito, entre outras.

Com estas restrições, podemos representar a intensidade na imagem como

$$I(x,y) = \Lambda(\vec{U}(x,y) \cdot \vec{\Theta} + A), \tag{29}$$

onde Λ é o valor constante de reflectância, $\vec{\Theta}$ é um vetor de tamanho três apontando na direção da origem de iluminação, $\vec{U}$ é um vetor de tamanho três representando a superfície normal no ponto (x, y) e A é o termo representando a iluminação constante do ambiente. Se estivermos interessados apenas na direção da iluminação, então o termo de reflectância Λ pode ser considerado unitário. A equação linear resultante possui uma restrição e quatro variáveis: os três componentes de $\vec{\Theta}$ e o termo de ambiente A.

Com pelo menos quatro pontos com a mesma reflectância Λ e superfícies normais distintas $\vec{U}(x, y)$, a direção da luz e o termo ambiente podem ser resolvidos utilizando-se mínimos quadrados.

Entretanto, para estimar a direção de iluminação, as abordagens tradicionais requerem o conhecimento das superfícies normais 3-D de, pelo menos, quatro pontos distintos na superfície analisada contendo a mesma reflectância o que é muito restritivo com apenas uma imagem para análise e objetos desconhecidos na cena. Para contornar esse problema, os autores utilizam uma abordagem desenvolvida por [Nillius e Eklundh 2001] que permite estimar dois componentes da direção de iluminação a partir de uma única imagem. Os autores relaxam a restrição de reflectância constante em toda a cena adotando agora um modelo em que a imagem possui regiões (*patches*) com reflectância constante. Essa suposição requer que a técnica seja capaz de estimar as direções da fonte de iluminação para cada região ao longo da superfície da imagem em análise. A Figura 6.14 mostra um exemplo onde as inconsistências relativas à iluminação podem levar à identificação de adulterações.

Johnson e Farid [Johnson e Farid 2007a] estenderam a solução acima para lidar com ambientes de iluminação mais complexos (mais de uma fonte originadora de luz). Sob as condições de simplificação já mencionadas, um ambiente com iluminação arbitrária pode ser expresso como uma função não negativa em uma esfera $\Theta(\vec{A})$. $\vec{A}$ é um vetor unitário em coordenadas Cartesianas e o valor de $\Theta(\vec{A})$ é a intensidade da luz incidente ao longo da direção $\vec{A}$. Como resultado, a irradiância, $\iota(\vec{U})$, pode ser parametrizada pela normal de superfície

$\vec{U}$ e escrita como uma convolução da função de reflectância sobre a superfície, $\Lambda(\vec{A}, \vec{U})$, e a iluminação ambiente $\Theta(\vec{A})$

$$\iota(\vec{U}) = \int_{\Omega} L(\vec{A})\Lambda(\vec{A}, \vec{U})d\Omega, \tag{30}$$

onde Ω representa a superfície. Para uma superfície Lambertiana, a função de reflectância é um cosseno aproximado

$$\Lambda(\vec{A}, \vec{U}) = \max(\vec{A} \cdot \vec{U}, 0). \tag{31}$$

A convolução na Equação 30 pode ser simplificada expressando-se a iluminação ambiente e as funções de reflectância em termos de harmônicos esféricos (*spherical harmonics*).

Ao analisar os contornos de oclusão de objetos em imagens reais, é comum encontrarmos um número limitado de superfícies normais. Dessa forma, pequenas quantidades de ruído nas superfícies normais ou mesmo nas intensidades medidas podem causar variações significativas na estimativa da iluminação ambiente [Johnson e Farid 2007a]. Uma das desvantagens do método proposto é que a identificação dos contornos de oclusão (bons candidatos para a determinação das superfícies normais) precisam ser marcados manualmente exigindo um certo conhecimento do operador.

Sistemas automáticos e semi-automáticos para identificação das fontes de iluminação podem representar um grande passo à frente na análise forense de documentos dado que o sistema visual humano pode ser incapaz de julgar inconsistências de iluminação e sombras como já estudado anteriormente [Ostrovsky et al. 2005]. Em [Farid 2009], o autor apresenta um caso interessante. Desde o assassinato do presidente americano John Kennedy, surgiram inúmeras teorias a respeito de seu assassinato. Em algumas delas, o assassino acusado, Lee Harvey Oswald, agiu como parte de uma conspiração.

Foi sugerido, por exemplo, que fotografias de incriminação de Oswald foram manipuladas, tornando-se evidências de um plano maior. Especificamente, foi argumentado que a iluminação e sombras nestas fotografias são fisicamente impossíveis. Dado que o sistema visual humano é incapaz de julgar iluminação e geometria tridimensional apropriadamente, Farid apresenta um estudo de caso em que prova que os elementos presentes na cena, ao contrário do que se pensava, são coerentes. Para sua análise, o autor constrói um modelo tridimensional da fotografia em questão a partir de fotografias adicionais do suspeito para determinar se as sombras na foto podem ser explicadas por uma única fonte de iluminação, como ilustra a Figura 6.15.

Recentemente, Johnson e Farid [Johnson e Farid 2007b] também investigaram inconsistências de iluminação analisando reflexos especulares nos olhos (pequeno branco na íris) para identificar imagens compostas de pessoas. A posição de um reflexo ocular é determinada pela relativa posição da fonte de luz, a superfície de reflexão e o visualizador (câmera). De acordo com os autores, reflexos oculares fornecem uma informação poderosa quanto à forma, cor e localização da fonte de iluminação em uma cena.

Figura 6.14. Exemplo de inconsistências de iluminação em uma composição de duas imagens. Observe a inconsistência entre a direção de iluminação Θ_1 na imagem destino (*host*) e a direção de iluminação Θ_2 na imagem composta (*spliced*).

A lei da reflexão em Física diz que um raio de luz reflete a partir de uma superfície em um ângulo de reflexão θ_r igual ao ângulo de incidência θ_i, onde estes ângulos são medidos em relação à superfície normal $\vec{U}$. Assumindo vetores unitários, a direção do raio refletido $\vec{R}$ pode ser escrita em termos da direção da luz $\vec{\Theta}$ e a normal de superfície $\vec{U}$

$$\vec{R} = \vec{\Theta} + 2(cos(\theta_i)\vec{U} - \vec{\Theta}) \quad (32)$$

$$= 2cos(\theta_i)\vec{U} - \vec{\Theta}. \quad (33)$$

A Figura 6.16 ilustra o procedimento. Assumindo-se um refletor perfeito ($\vec{V} = \vec{R}$), a restrição acima resulta

$$\vec{\Theta} = 2cos(\theta_i)\vec{U} - \vec{V} \quad (34)$$

$$= 2(\vec{V}^T\vec{U})\vec{U} - \vec{V}. \quad (35)$$

Com isso, a direção da luz $\vec{\Theta}$ pode ser estimada a partir da normal de superfície $\vec{U}$ e a direção de visualização $\vec{V}$ em um reflexo ocular.

É importante ressaltar, no entanto, que reflexos oculares tendem a ser relativamente pequenos permitindo a um falsificador habilidoso fazer manipulações

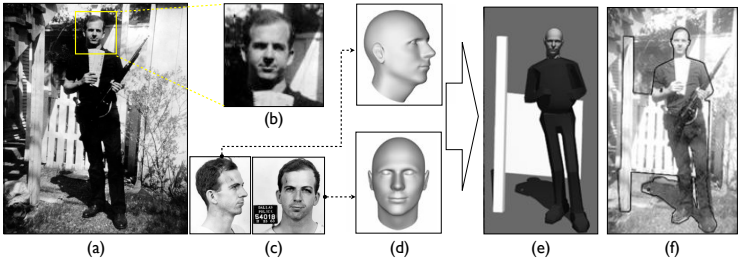


Figura 6.15. Ilustração do caso de Lee Harvey Oswald. (a) Oswald em seu quintal (uma das fotografias sob investigação). (b) *zoom* da área da cabeça mostrando sombras no queixo e nariz aparentemente inconsistentes com as sombras no chão. (c) informações auxiliares ao caso (e.g., fotos do registro de prisão). (d) Reconstrução tridimensional a partir das imagens auxiliares anteriores. (e) combinação da face 3-D reconstruída em (c) com um corpo articulado genérico e informações de fundo da fotografia analisada para criar uma cena tridimensional fidedigna. (f) Super-imposição da reconstrução 3-D com a fotografia investigada – a geometria da cena e as sombras tem um casamento quase perfeito, levando à conclusão de que a fotografia pode ser explicada por uma única fonte de luz, o sol.

de modo a esconder modificações nas imagens. Para isso, cor, forma e localização dos reflexos têm que ser construídos de modo a serem globalmente consistentes com a iluminação presente no restante da cena.

6.4.3.6. Técnicas que analisam inconsistências de compressão

Algumas técnicas forenses são desenvolvidas com algum alvo específico. Por exemplo, se uma determinada operação de adulteração modifica certa propriedade estrutural de uma imagem ou documento, é natural o desenvolvimento de uma técnica particular para detecção deste tipo de anomalia.

Neste sentido, Popescu e Farid [Popescu e Farid 2004b] analisam os efeitos da dupla quantização de imagens codificadas no formato JPEG e apresentam uma técnica para a detecção deste tipo de atividade.

A dupla compressão ou quantização JPEG introduz artefatos específicos não presentes em imagens comprimidas uma única vez. Os autores, argumen-

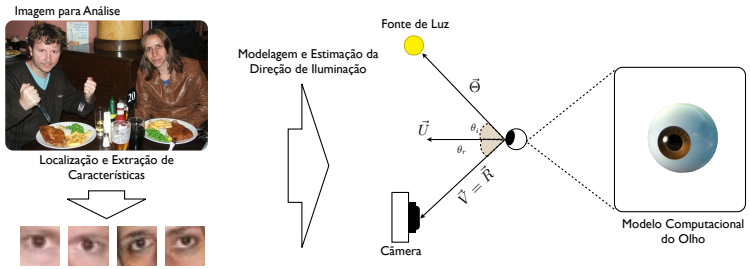


Figura 6.16. Exemplo de análise de uma imagem a partir de reflexos oculares. Dada a imagem a ser investigada, localiza-se os olhos e modela-se o sistema. A posição do reflexo ocular em um dado olho é determinada pela superfície normal $\vec{U}$ e as direções relativas da fonte de iluminação Θ e o visualizador V.

tam, no entanto, que a presença de dupla quantização não necessariamente implica um ato malicioso. Por exemplo, um usuário pode desejar simplesmente economizar espaço em disco rearmazenando suas imagens com a metade da resolução original. A Figura 6.17 ilustra efeito da dupla quantização sobre um pequeno sinal sintético unidimensional $\mu[t]$ com distribuição normal no intervalo $[0, 127]$.

No cenário forense, Lin et al. [He et al. 2006] propuseram uma abordagem para detectar regiões adulteradas em imagens analisando o efeito da dupla quantização escondidos nos coeficientes da transformada de cosseno em uma extensão do trabalho de Popescu e Farid [Popescu e Farid 2004b]. A idéia é que, dado que a imagem adulterada contém partes adulteradas e também partes sem alteração alguma, os histogramas da transformada de cosseno da parte inalterada sofrem os efeitos da dupla quantização. Esta parte da imagem é a mesma da imagem original JPEG (imagem destino ou *host*). Por outro lado, os histogramas da parte modificada (imagem para composição ou *spliced image*) não têm o mesmo efeito da dupla quantização se esta parte da imagem provém de uma câmera com formato diferente, ou de uma imagem JPEG diferente. Algumas razões para isso acontecer são:

1. ausência da primeira compressão JPEG na parte composta;
2. diferenças da grade de amostragem da imagem destino e da imagem composta;
3. a composição dos blocos resultantes da transformada de cosseno ao longo da borda pode esconder traços das partes originais e compostas dado que é improvável que a parte composta contenha blocos exatos de 8×8 *pixels*.

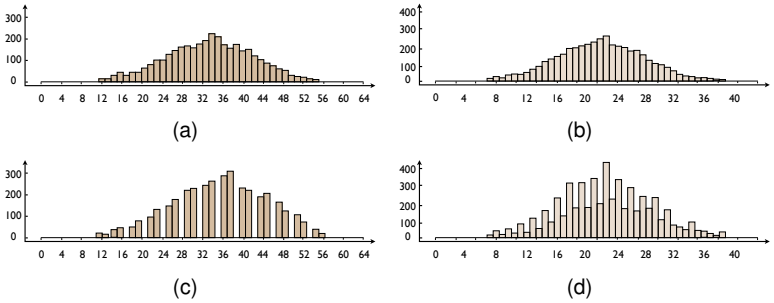


Figura 6.17. A linha superior ilustra histogramas de sinais com quantização simples de passos 2 (a) e 3 (b). A linha inferior ilustra histogramas de sinais duplamente quantizados com passos 3 seguidos de 2 (c), e 2 seguidos de 3 (d). Observe os artefatos nos histogramas de dupla quantização (picos e intervalos).

É importante ressaltar que esta abordagem não funcionará em alguns casos. Por exemplo, se a imagem original de destino não é uma imagem JPEG, o efeito da dupla quantização na parte sem modificações não será detectado. Além disso, os níveis de compressão também afetam a qualidade de resposta. Em termos gerais, quanto menor a razão entre o grau de intensidade da segunda quantização em relação à primeira, mais difícil será a detecção dos efeitos da dupla compressão.

Um ataque contra-forense neste cenário consiste em reamostrar a imagem adulterada em uma nova grade (deslocamento de um ou dois *pixels*, por exemplo). Esse tipo de operação provavelmente diminuirá os traços/pistas da dupla quantização pois irá gerar uma nova tabela de quantização.

No cenário forense, algumas vezes temos a necessidade de apontar se uma determinada imagem foi modificada de qualquer forma (criminosamente ou não) desde a sua captura, incluindo operações simples como correções de brilho e contraste, por exemplo. Esse problema é conhecido como autenticação de imagem. Neste contexto, Kee e Farid [Kee e Farid 2010] apresentam uma abordagem que explora a formação e armazenamento de um *thumbnail* de imagem para autenticação de imagens no formato JPEG.

Um *thumbnail* de imagem é uma representação da imagem de alta resolução utilizando uma versão reduzida (tipicamente na ordem de 160×120 *pixels*). Esta representação, em geral, é salva juntamente com o cabeçalho da imagem no formato JPEG de modo a facilitar a visualização da mesma em computadores, nas próprias câmeras etc.

A abordagem de [Kee e Farid 2010] consiste em modelar *thumbnails* de

imagens a partir de uma série de operações de filtragem, ajuste de contraste e compressão. Os autores automaticamente calculam os parâmetros do modelo de estimativa e mostram que estes parâmetros, embora não únicos, diferem significativamente entre câmeras e pacotes de *software* tais como *Adobe Photoshop*.

Dada uma imagem $I(x, y)$, seu *thumbnail* é criado por uma série de seis passos: (1) recorte; (2) pré-filtragem; (3) redimensionamento; (4) ajuste de brilho; (5) ajuste de contraste e (6) compressão JPEG. Ao estimar os parâmetros de criação do *thumbnail* de uma imagens, estes representam uma espécie de assinatura da imagem em questão. Construindo-se uma base de informações com as assinaturas de diversas câmeras e diversos pacotes de edição de imagens, é possível apontar, analisando-se variações nestas assinaturas, se uma determinada imagem em formato JPEG é autêntica ou modificada de alguma forma.

Um possível contra-ataque para esta técnica consiste em substituir o *thumbnail* gerado pelo *software* de edição por um *thumbnail* estimado para o modelo de câmera que capturou a imagem.

6.5. Conclusões

The very nature of photography was to record events — Hany Farid, Darthmouth College, EUA

Concluímos esse capítulo apresentando algumas tendências na área de análise forense de documentos. Adicionalmente, discutimos proposições para melhorar as técnicas existentes bem como propor novas soluções.

Há uma grande demanda por soluções eficientes e eficazes para resolver problemas em análise forense de documentos. Entretanto, existem muitos desafios aos quais podemos explorar e contribuir para esta área

- **Avaliação de performance e *benchmarking***: a maior parte das técnicas apresentadas possui validação insuficiente consistindo de uma prova de conceito e alguns exemplos onde a técnica se aplica. Muitas vezes não há comparação entre as abordagens previamente apresentadas. Não há um conjunto de dados padrão para ser analisado.
- **Robustez**: normalmente as técnicas apresentadas não procuram discutir as suas limitações frente a possíveis ataques de robustez. Basicamente, queremos responder à pergunta: “podemos acreditar na análise forense de imagens” [Gloe et al. 2007b].

Técnicas contra-forenses não devem ser vistas como um atraso às atividade forenses. Pelo contrário, com a identificação de ataques contra-forenses às abordagens existentes, seremos capazes de aprender suas limitações e tirar conclusões importantes no desenvolvimento de técnicas forenses cada vez mais avançadas.

Ao desenvolvermos soluções para detecção de adulterações é bastante comum analisarmos propriedades físicas dos equipamentos de captura dos

dados e suas propriedades estatísticas. Além disso, dentro da própria Ciência da Computação, precisamos de técnicas de mineração, indexação, clusterização, e resumo de dados (Bancos de Dados); técnicas de análise de padrões, aprendizado de máquina e heurísticas (Estatística e Inteligência Artificial); técnicas de Processamento de Imagens e Vídeos e de Visão Computacional, no caso dos dados analisados serem imagens ou vídeos; bem como abordagens de teoria da computação e algoritmos, pois sempre buscamos abordagens eficientes e eficazes para nossas soluções.

Durante seu encontro em 2006, a Sociedade Brasileira de Computação (SBC) identificou o tema de Gestão da informação em grandes volumes de dados distribuídos como um dos grandes desafios da computação no Brasil. Esse capítulo vem ao encontro de tal constatação.

Grandes bancos de dados começam a surgir em diversas partes do mundo e do Brasil. Como evitar o roubo e falsificação de documentos em tais sistemas? Certamente, precisamos saber como armazenar nossas informações de modo a aumentar sua segurança. Políticas de proteção da privacidade e de verificação de autenticidade têm que ser discutidas e implementadas.

Nesse sentido, um trabalho colaborativo e multi-disciplinar com a utilização de diversas áreas do conhecimento bem como a conscientização crítica de pesquisadores e entusiastas podem nos conduzir um passo à frente.

Agradecimentos

Agradecemos aos revisores e editores por suas considerações construtivas que permitiram a melhoria do presente capítulo. Adicionalmente, gostaríamos de expressar nossa gratidão à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio em nossos projetos de pesquisa.

A. Conceitos importantes

Nesta seção, apresentamos alguns conceitos considerados importantes para o entendimento do texto de forma auto-contida.

A.1. Imagem Digital

De acordo com [Gomes e Velho 1996], devemos estabelecer um universo matemático no qual seja possível definir diversos modelos abstratos de uma imagem. Em seguida, precisamos criar um universo de representação onde procuramos esquemas que permitam uma representação discreta desses modelos, com o objetivo de codificar tal imagem em um computador.

Quando observamos uma fotografia, ou uma cena no mundo real, recebemos de cada ponto do espaço um impulso luminoso que associa uma informação de cor a esse ponto [Gomes e Velho 1996]. Nesse sentido, podemos definir uma *imagem contínua* (não discreta) como a aplicação $\mathcal{I} : \mathcal{U} \rightarrow \mathcal{C}$, onde $\mathcal{U} \subset \mathbb{R}^3$ é uma superfície e $\mathcal{C}$ é um espaço vetorial. Na maioria das aplicações, $\mathcal{U}$ é um subconjunto plano e $\mathcal{C}$ é um espaço de cor. A função $\mathcal{I}$ na definição é chamada de *função imagem*. O conjunto $\mathcal{U}$ é chamado de *suporte da imagem*,

e o conjunto de valores de $\mathcal{I}$, que é um subconjunto de $\mathcal{C}$, é chamado de conjunto de *valores da imagem* [Gomes e Velho 1996]. Quando $\mathcal{C}$ é um espaço de cor de dimensão 1, dizemos que a imagem é *monocromática*.

A representação mais comum de uma imagem espacial consiste em tomar um subconjunto discreto $\mathcal{U}' \subset \mathcal{U}$ do domínio da imagem, uma espaço de cor $\mathcal{C}$ associado a um dispositivo gráfico e representar a imagem pela amostragem da função imagem $\mathcal{I}$ no conjunto $\mathcal{U}'$. Cada ponto (x_i, y_i) do subconjunto discreto $\mathcal{U}'$ é chamado de elemento da imagem ou *pixel*. Para a representação em computador, devemos também trabalhar com modelos de imagem onde a função imagem $\mathcal{I}$ toma valores em um subconjunto discreto do espaço de cor $\mathcal{C}$. Esse processo de discretização de uma imagem é chamado de *quantização*.

O caso mais utilizado de discretização espacial de uma imagem consiste em tomar o domínio como sendo um retângulo e discretizar esse retângulo usando os pontos de um reticulado bidimensional. Dessa forma, a imagem pode ser representada de forma matricial por uma matriz A de ordem m linhas e n colunas tal que $A = (a_{ij} = (\mathcal{I}(x_i, y_j))$. Cada elemento a_{ij}, $i = 1, \ldots, m$ e $j = 1, \ldots n$ da matriz representa o valor da função imagem $\mathcal{I}$ no ponto de coordenadas (x_i, y_j) do reticulado, sendo pois, um vetor do espaço de cor, representando a cor do *pixel* na coordenada (i, j). Nesse contexto, chamamos de *resolução de cor* ao número de *bits* utilizado para armazenar o vetor de cor a_{ij} de cada *pixel* da imagem. Se cada ponto possui três valores associados e cada valor precisa de oito *bits* para ser representado, então cada *pixel* dessa imagem pode ser representado com 24 *bits* e a imagem é dita de 24 *bits*.

Diversas decomposições de uma imagem podem ser feitas de acordo com cada aplicação. Neste capítulo, falamos em:

- **Decomposição em canais de cor**. Quando separamos a imagem em suas cores básicas representadas no espaço de cores $\mathcal{C}' \subset \mathcal{C}$. Por exemplo, se o espaço de cores utilizado é um espaço conhecido como RGB, temos os componentes vermelho (***R**ed*), verde (***G**reen*), e azul (***B**lue*);
- **Decomposição *wavelet***. Quando decompomos a imagem em diversas escalas e orientações segundo a transformada *wavelet*;
- **Decomposição em planos de *bits***. Quando decompomos a imagem em seus planos de *bits*. Por exemplo, após a decomposição da imagem de 24 *bits* em seus três canais de cores (R,G,B), podemos ainda, fazer uma decomposição por planos de *bits*. Neste caso, cada canal de cor possui 8 *bits* e possui 8 planos de *bits* por canal de cor.

A.2. Aprendizado de Máquina

Aprendizado de máquina é uma área da Inteligência Artificial concentrada no desenvolvimento de técnicas que permitem que computadores sejam capazes de aprender com a experiência [Mitchell 1997]. Alguns problemas que utilizam aprendizado de máquina são: reconhecimento de caracteres, reconhecimento da fala, predição de ataques cardíacos e detecção de fraudes em cartões de créditos [Mitchell 1997, Friedman et al. 2001].

Na solução desses problemas, podemos ter classificadores fixos ou baseados em aprendizado, que, por sua vez, pode ser supervisionado ou não-supervisionado [Friedman et al. 2001].

Neste sentido, podemos ver um classificador, matematicamente, como um mapeamento a partir de um espaço de características X para um conjunto discreto de rótulos (*labels*) Y. Mais especificamente, em Inteligência Artificial, um classificador de padrões é um tipo de motor de inferência que implementa estratégias eficientes para computar relações de classificação entre pares de conceitos ou para computar relações entre um conceito e um conjunto de instâncias [Duda et al. 2000].

Classificadores supervisionados como os utilizados pela maioria das técnicas descritas neste capítulo, consistem em técnicas em que procuramos estimar uma função f de classificação a partir de um conjunto de treinamento. O conjunto de treinamento consiste de pares de valores de entrada X, e sua saída desejada Y [Friedman et al. 2001]. Valores observados no conjunto X são denotados por x_i, isto é, x_i é a i-ésima observação em X. O número de variáveis que constituem cada uma das entradas em X é p. Desta forma, X é formado por N vetores de entrada, chamados vetores de características, e cada vetor de entrada é composto por p graus de liberdade (dimensões e/ou variáveis).

A saída da função f pode ser um valor contínuo (regressão), ou pode predizer a etiqueta (*label*) de um objeto de entrada (classificação). A tarefa do aprendizado é predizer o valor da função para qualquer objeto de entrada que seja válido após ter sido suficientemente treinado com um conjunto de exemplos. Alguns exemplos de classificadores supervisionados são *Support Vector Machines*, *Linear Discriminant Analysis*, *Boosting* [Bishop 2006].

Um outro grupo de técnicas de aprendizado, não utilizam exemplos de treinamento e são conhecidos como técnicas para aprendizado não-supervisionado. Esta forma de aprendizado, na maioria das vezes, trata o seu conjunto de entrada como um conjunto de variáveis aleatórias. Um modelo de distribuição conjunta (*joint distribution model*) é então construído para a representação dos dados. Desta forma, o objetivo deste aprendizado é avaliar como os dados estão organizados e agrupados [Friedman et al. 2001]. Técnicas de *Maximização de Esperança* [Baeza-Yates 2003], por exemplo, podem ser utilizadas para aprendizado não-supervisionado.

A.3. Definições Complementares

Cross-Correlation

Correlação cruzada (*Cross-Correlation*) é uma medida de similaridade entre dois sinais como uma função de um pequeno deslocamento aplicado a um dos sinais. É também conhecida como o produto interno com deslocamento de um dos sinais. Para funções discretas, é definida como

$$(f \star g)[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m]g[n+m] \tag{36}$$

Por exemplo, suponha que tenhamos duas funções f e g que diferem por um pequeno deslocamento no eixo x. Podemos utilizar a correlação cruzada para identificar o quanto g precisa ser deslocada no eixo x para torná-la idêntica a f. Dependendo da aplicação, no cálculo da correlação cruzada, podemos utilizar uma etapa de normalização.

Peak to Correlation Energy

Peak to Correlation Energy (PCE) é uma medida para estimar o pico em uma superfície. Por exemplo, pode ser utilizada para calcular o pico em uma superfície de valores produzidos pelo cálculo de uma correlação cruzada entre dois sinais. PCE é definida (em termos da correlação cruzada, por exemplo) como

$$PCE = \frac{NCC[u_{peak}, v_{peak}]^2}{\frac{1}{mn-|\mathcal{N}_{peak}|}\sum_{(u,v)\notin\mathcal{N}_{peak}} NCC[u,v]^2}, \tag{37}$$

onde $\mathcal{N}_{peak}$ é uma pequena vizinhança ao redor de um pico, m e n são os comprimentos do sinais.

Referências bibliográficas

[Acbibas et al. 2003] Acbibas, I., Memon, N. e Sankur, B. (2003). Steganalysis using image quality metrics. *IEEE Transactions on Image Processing (TIP)*, 12(2):221–229.

[Avcibas et al. 2004] Avcibas, I., Bayaram, S., Memon, N., Ramkumar, M. e Sankur, B. (2004). A classifier design for detecting image manipulations. In *Intl. Conf. on Image Processing (ICIP)*, pp. 2645–2648, Singapore.

[Baeza-Yates 2003] Baeza-Yates, R. (2003). *Clustering and Information Retrieval*. Kluwer Academic Publishers, 1 edição.

[Bayaram et al. 2005a] Bayaram, S., Avcibas, I., Sankur, B. e Memon, N. (2005a). Image manipulation detection with binary similarity measures. In *European Signal Processing Conf. (EUSIPCO)*, pp. 752–755, Antalya, Turkey.

[Bayaram et al. 2006] Bayaram, S., Avcibas, I., Sankur, B. e Memon, N. (2006). Image manipulation detection. *Journal of Electronic Imaging (JEI)*, 15(4):1–17.

[Bayaram et al. 2005b] Bayaram, S., Sencar, H., Memon, N. e Avcibas, I. (2005b). Source camera identification based on CFA interpolation. In *Intl. Conf. on Image Processing (ICIP)*, Genova, Italy.

[Bayram et al. 2005] Bayram, S., Sencar, H. e Memon, N. (2005). Source camera identification based on CFA interpolation. In *Intl. Conf. on Image Processing*, Genova, Italy. IEEE.

[Bayram et al. 2006] Bayram, S., Sencar, H. e Memon, N. (2006). Improvements on source camera-model identiciation based on CFA interpolation. In *WG 11.9 Int. Conf. on Digital Forensics*, Orlando, USA. IFIP.

[Bishop 2006] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, 1 edição.

[Castro e Morandi 1987] Castro, E. D. e Morandi, C. (1987). Registration of translated and rotated images using finite fourier transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:700–703.

[Celiktutan et al. 2005] Celiktutan, O., Avcibas, I., Sankur, B. e Memon, N. (2005). Source cell-phone identification. In *Intl. Conf. on Advanced Computing and Communication (ADCOM)*.

[Choe Sang-Hun 2006] Choe Sang-Hun (2006). Disgraced cloning expert convicted in south korea. *The New York Times*. `http://www.nytimes.com/2009/10/27/world/asia/27clone.html`.

[Choi et al. 2006] Choi, K. S., Lam, E. e Wong, K. (2006). Automatic source camera identification using the intrinsic lens radial distortion. *Optics Express*, 14(24):11551–11565.

[Clifford J. Levy 2008] Clifford J. Levy (2008). It isn't magic: Putin opponents vanish from tv. *The New York Times*. `http://www.nytimes.com/2008/06/03/world/europe/03russia.html?_r=1&fta=y`.

[Cormen et al. 2001] Cormen, T., Leiserson, C., Rivest, R. e Stein, C. (2001). *Introduction to Algorithms*. MIT Press, 2 edição.

[Cox et al. 2007] Cox, I., Miller, M., Bloom, J., Fridrich, J. e Kalker, T. (2007). *Digital Watermarking and Steganography*. Morgan Kaufmann, 2 edição.

[Dehnie et al. 2006] Dehnie, S., Sencar, T. e Memon, N. (2006). Identification of computer generated and digital camera images for digital image forensics,. In *Intl. Conf. on Image Processing (ICIP)*, Atlanta, USA.

[Dirik et al. 2008] Dirik, A. E., Sencar, H. T. e Memon, N. (2008). Digital single lens reflex camera identification from traces of sensor dust. *IEEE Transactions on Information Forensics and Security (TIFS)*, 3(3):539–552.

[Dirik et al. 2007] Dirik, E., Sencar, H. e Memon, N. (2007). Source camera identification based on sensor dust characteristics. In *IEEE Intl. Workshop on Signal Processing Applications for Public Security and Forensics (SAFE)*, pp. 1–6, Washington DC, USA.

[Duda et al. 2000] Duda, R. O., Hart, P. E. e Stork, D. G. (2000). *Pattern Classification*. Wiley-Interscience, 2.

[Farid 2007] Farid, H. (2007). *Deception: Methods, Motives, Contexts and Consequences*, capítulo Digital Doctoring: can we trust photographs? Stanford University Press.

[Farid 2009] Farid, H. (2009). The Lee Harvey Oswald backyard photos: real

or fake? *Perception*, 38(11):1731–1734.

[Folha de São Paulo 2010] Folha de São Paulo (2010). Globo apaga nome de banco no "JN". `http://www1.folha.uol.com.br/folha/ilustrada/ult90u678272.shtml`. 12 de janeiro.

[Fridrich et al. 2003] Fridrich, J., Soukal, D. e Lukas, J. (2003). Detection of copy-move forgery in digital images. In *Digital Forensic Research Workshop (DFRWS)*, Cleveland, USA.

[Friedman et al. 2001] Friedman, J., Hastie, T. e Tibshirani, R. (2001). *The Elements of Statistical Learning*. Springer, 1 edição.

[Geradts et al. 2001] Geradts, Z., Bijhold, J., Kieft, M., Kurusawa, K., Kuroki, K. e Saitoh, N. (2001). Methods for identification of images acquired with digital cameras. In *Enabling Technologies for Law Enforcement and Security*, volume 4232, –. SPIE.

[Gloe et al. 2007a] Gloe, T., Franz, E. e Winkler, A. (2007a). Forensics for flatbed scanners. In *SPIE Intl. Conf. on Security, Steganography, Watermarking of Multimedia Contents*, pp. 65051–I.

[Gloe et al. 2007b] Gloe, T., Kirchner, M., Winkler, A. e Bohme, R. (2007b). Can we trust digital image forensics? In *ACM Multimedia (ACMMM)*, pp. 78–86, Augsburg, Germany.

[Goldenstein e Rocha 2009] Goldenstein, S. e Rocha, A. (2009). High-profile forensic analysis of images. In *Intl. Conf. on Imaging for Crime Detection and Prevention (ICDP)*, pp. 1–6.

[Goljan et al. 2008] Goljan, M., Fridrich, J. e Lukas, J. (2008). Camera identification from printed images. In *SPIE Electronic Imaging, Forensics, Security, Steganography, and Watermarking of Multimedia Contents*, pp. OI–1–OI–12.

[Gomes e Velho 1996] Gomes, J. e Velho, L. (1996). *Computação Gráfica: Imagem*. IMPA-SBM, 1.

[Gonzalez e Woods 2007] Gonzalez, R. e Woods, R. (2007). *Digital Image Processing*. Prentice-Hall, 3 edição.

[Gou et al. 2007] Gou, H., Swaminathan, A. e Wu, M. (2007). Robust scanner identification based on noise features. In *SPIE Security, Steganography, and Watermarking of Multimedia Contents (SSWMC)*, San Jose, USA.

[Grossberg e Nayar 2010] Grossberg, M. e Nayar, S. (2010). Database of Response Functions (DoRF). Available at `http://www.cs.columbia.edu/CAVE/software/softlib/dorf.php`.

[He et al. 2006] He, J., Lin, Z., Wang, L. e Tang, X. (2006). Detecting doctored jpeg images via dct coefficient analysis. In *European Conf. on Computer Vision (ECCV)*, pp. 423–435.

[Johnson 2007] Johnson, M. K. (2007). *Lighting and Optical Tools for Image Forensics*. Phd thesis, Dep. of Computer Science - Dartmouth College, Ha-

nover, USA.

[Johnson e Farid 2005] Johnson, M. K. e Farid, H. (2005). Exposing digital forgeries by detecting inconsistencies in lighting. In *ACM Multimedia and Security Workshop*, New York, USA.

[Johnson e Farid 2007a] Johnson, M. K. e Farid, H. (2007a). Exposing digital forgeries in complex lighting environments. *IEEE Transactions on Information Forensics and Security (TIFS)*, 2(3):450–461.

[Johnson e Farid 2007b] Johnson, M. K. e Farid, H. (2007b). Exposing digital forgeries through specular highlights on the eye. In *Intl. Workshop in Information Hiding (IHW)*, Saint Malo, France.

[Kavanagh 2006] Kavanagh, E. (2006). Editorial expression of concern. *Science*, 314:592–594.

[Kee e Farid 2010] Kee, E. e Farid, H. (2010). Digital image authentication from thumbnails. In *SPIE Symposium on Electronic Imaging*, San Jose, USA.

[Khanna et al. 2007] Khanna, N., Mikkilineni, A. K., Chiu, G. T. C., Allebach, J. P. e Delp, E. J. (2007). Scanner identification using sensor pattern noise. In *SPIE Security, Steganography, and Watermarking of Multimedia Contents (SSWMC)*, volume 6505, pp. 1–11.

[Khanna et al. 2009] Khanna, N., Mikkilineni, A. K. e Delp, E. J. (2009). Scanner identification using feature-based processing and analysis. *IEEE Transactions on Information Forensics and Security (TIFS)*, 4(1):123–139.

[Kharrazi et al. 2004] Kharrazi, M., Sencar, H. e Memon, N. (2004). Blind source camera identification. In *Intl. Conf. on Image Processing (ICIP)*, Singapore.

[Kossoy 2006] Kossoy, B. (2006). *Hercule Florence – A descoberta isolada da fotografia no Brasil*. Edusp, 1 edição.

[Kurosawa et al. 1999] Kurosawa, K., Kuroki, K. e Saitoh, N. (1999). Ccd fingerprint method. In *Intl. Conf. on Image Processing*, Kobe, Japan. IEEE.

[Li et al. 2004] Li, Y., Sun, J., Tang, C.-K. e Shum, H.-Y. (2004). Lazy snapping. *ACM Transactions on Graphics (ToG)*, 23(3):303–308.

[Liang et al. 2001] Liang, L., Liu, C., Xu, Y. Q., Guo, B. e Shum, H. (2001). Real-time texture synthesis by patch-based sampling. *ACM Transactions on Graphics (ToG)*, 20(3):127–150.

[Lin et al. 2004] Lin, S., Gu, J., Yamazaki, S. e Shum, H. Y. (2004). Radimetric calibration from a single image. In *Intl. Conf. on Computer Vision and Pattern Recognition*, pp. 938–945, Washington, USA. IEEE.

[Lin et al. 2005] Lin, Z., Wang, R., Tang, X. e Shum, H.-Y. (2005). Detecting doctored images using camera response normality and consistency. In *Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, New York, USA.

[Liu et al. 2009] Liu, J., Sun, J. e Shum, H.-Y. (2009). Paint selection. *ACM*

Transactions on Graphics (ToG), 28(3):69:1–69:8.

[Long e Huang 2006] Long, Y. e Huang, Y. (2006). Image based source camera identification using demosaicing. In *Intl. Workshop on Multimedia Signal Processing*, Victoria, Canada. IEEE.

[Lukas et al. 2006] Lukas, J., Fridrich, J. e Goljan, M. (2006). Digital camera identification from sensor noise sensor. *IEEE Transactions on Information Forensics and Security (TIFS)*, 1(2):205–214.

[Lukas et al. 2007] Lukas, J., Fridrich, J. e Goljan, M. (2007). Detecting digital image forgeries using sensor pattern noise. In *SPIE Photonics West*.

[Lyu 2005] Lyu, S. (2005). *Natural Image Statistics for Digital Image Forensics*. Phd thesis, Dep. of Computer Science - Dartmouth College, Hanover, USA.

[Lyu e Farid 2002] Lyu, S. e Farid, H. (2002). Detecting hidden messages using higher-order statistics and support vector machines. In *Intl. Workshop in Information Hiding (IHW)*, pp. 340–354.

[Lyu e Farid 2004] Lyu, S. e Farid, H. (2004). Steganalysis using color wavelet statistics and one-class support vector machines. In *Symposium on Electronic Imaging*.

[Lyu e Farid 2005] Lyu, S. e Farid, H. (2005). How realistic is photorealistic? *IEEE Transactions on Signal Processing (TSP)*, 53(2):845–850.

[Maurer 1992] Maurer, U. (1992). A universal statistical test for random bit generators. *Intl. Journal of Cryptology*, 5(2):89–105.

[Columbia DVMM Research Lab. 2004] Columbia DVMM Research Lab. (2004). Columbia image splicing detection evaluation data set. Available at `http://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/AuthSplicedDataSet.htm`.

[Mitchell 1997] Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, 1 edição.

[Nemer et al. 2001] Nemer, E., Goubran, R. e Mahmoud, S. (2001). Robust voice activity detection using higher-order statistics in the LPC residual domain. *IEEE Transactions on Speech and Audio Processing*, 9(3):217–231.

[Ng e Chang 2004] Ng, T.-T. e Chang, S.-F. (2004). Blind detection of photomontage using higher order statistics. In *Intl. Symposium on Circuits and Systems (ISCAS)*, pp. 688–691, Vancouver, Canada.

[Ng et al. 2005] Ng, T.-T., Chang, S.-F. e Tsui, M.-P. (2005). Physics-motivated features for distinguishing photographic images and computer graphics. In *ACM Multimedia (ACMMM)*, pp. 239–248, Singapore.

[Nillius e Eklundh 2001] Nillius, P. e Eklundh, J.-O. (2001). Automatic estimation of the projected light source direction. In *Intl. Conf. on Computer Vision and Pattern Recognition*, pp. 1076–1082, Hawaii, US. IEEE.

[Ostrovsky et al. 2005] Ostrovsky, Y., Cavanagh, P. e Sinha, P. (2005). Percei-

ving illumination inconsistencies in scenes. *Perception*, 34(11):1301–1314.

[Parrish e Noonan 2009] Parrish, D. e Noonan, B. (2009). Image manipulation as research misconduct. *Sci Eng Ethics (2009)*, 15:161–167.

[Pearson 2005] Pearson, H. (2005). Image manipulation: CSI: Cell biology. *Nature*, 434:952–953.

[Popescu 2004] Popescu, A. C. (2004). *Statistical Tools for Digital Image Forensics*. Phd thesis, Dep. of Computer Science - Dartmouth College, Hanover, USA.

[Popescu e Farid 2004a] Popescu, A. C. e Farid, H. (2004a). Exposing digital forgeries by detecting duplicated image regions. Relatório Técnico TR 2004-515, Dep. of Computer Science - Dartmouth College, Hanover, USA.

[Popescu e Farid 2004b] Popescu, A. C. e Farid, H. (2004b). Statistical tools for digital forensics. In *Intl. Workshop in Information Hiding (IHW)*, Toronto, Canada.

[Popescu e Farid 2005a] Popescu, A. C. e Farid, H. (2005a). Exposing digital forgeries by detecting traces of re-sampling. *IEEE Transactions on Signal Processing (TSP)*, 53(2):758–767.

[Popescu e Farid 2005b] Popescu, A. C. e Farid, H. (2005b). Exposing digital forgeries in color filter array interpolated images. *IEEE Transactions on Signal Processing (TSP)*, 53(10):3948–3959.

[Rocha e Goldenstein 2006] Rocha, A. e Goldenstein, S. (2006). Progressive randomization for steganalysis. In *Intl. Workshop on Multimedia and Signal Processing (MMSP)*, pp. 314–319.

[Rocha e Goldenstein 2007] Rocha, A. e Goldenstein, S. (2007). Pr: More than meets the eye. In *Intl. Conf. on Computer Vision (ICCV)*, pp. 1–8.

[Rocha e Goldenstein 2010] Rocha, A. e Goldenstein, S. (2010). Progressive randomization: Seeing the unseen. *Computer Vision and Image Understanding (CVIU)*, 114(3):349–362.

[Rocha et al. 2011] Rocha, A., Scheirer, W., Boult, T. E. e Goldenstein, S. (2011). Vision of the unseen: Current trends and challenges in digital image and video forensics. *ACM Computing Surveys (CSUR)*.

[Sacchi et al. 2007] Sacchi, D. L. M., Agnoli, F. e Loftus, E. F. (2007). Changing history: Doctored photographs affect memory for past public events. *Applied Cognitive Psychology*, 21(8):249–273.

[Sencar e Memon 2008] Sencar, T. e Memon, N. (2008). *Overview of State-of-the-art in Digital Image Forensics*, capítulo Statistical Science and Interdisciplinary Research. World Scientific Press.

[Shi et al. 2007] Shi, Y. Q., Chen, C. e Chen, W. (2007). A natural image model approach to splicing detection. In *ACM Multimedia and Security Workshop*, pp. 51–62, Dallas, USA.

[Sun et al. 2004] Sun, J., Jia, J., Tang, C.-K. e Shum, H.-Y. (2004). Poisson matting. *ACM Transactions on Graphics (ToG)*, 23(3):315–321.

[Sun et al. 2005] Sun, J., Yuan, L., Jia, J. e Shum, H.-Y. (2005). Image completion with structure propagation. *ACM Transactions on Graphics (ToG)*, 24(3):861–868.

[Sutcu et al. 2007] Sutcu, Y., Bayaram, S., Sencar, H. e Memon, N. (2007). Improvements on sensor noise based source camera identification. In *Intl. Conf. on Multimedia and Expo (ICME)*, Beijing, China.

[Swaminathan et al. 2006] Swaminathan, A., Wu, M. e Liu, K. R. (2006). Non-instrusive forensics analysis of visual sensors using output images. In *Intl Conf. on Image Processing*, Atlanta, USA. IEEE.

[Tsai e Wu 2006] Tsai, M. e Wu, G. (2006). Using image features to identify camera sources. In *Intl. Conf. on Acoustics, Speech, and Signal Processing*, Toulouse, France. IEEE.

[Tyson 2001] Tyson, J. (2001). How scanners work. `http://com-puter.howstuffworks.com/scanner.htm`.

[UOL Notícias 2009] UOL Notícias (2009). Globo 'lima' hotel de reportagem. `http://noticias.uol.com.br/ooops/ultnot/2009/11/24/ult2548u809.jhtm`. 24 de novembro.

[Vaidyanathan 1987] Vaidyanathan, P. P. (1987). Quadrature mirror filter banks, m-band extensions and perfect reconstruction techniques. *IEEE Signal Processing Magazine*, 4(3):4–20.

[Wang e Farid 2007] Wang, W. e Farid, H. (2007). Exposing digital forgeries in video by detecting duplication. In *ACM Multimedia and Security Workshop*, Dallas, USA.

[Westfeld e Pfitzmann 1999] Westfeld, A. e Pfitzmann, A. (1999). Attacks on steganographic systems. In *Intl. Workshop in Information Hiding (IHW)*, pp. 61–76.

[Yedidia et al. 2003] Yedidia, J. S., Freeman, W. T. e Weiss, Y. (2003). *Exploring Artificial Intelligence in the New Millennium*, capítulo Understanding Belief Propagation and Its Generalizations, pp. 239–236. Science & Technology Books.

[Yu et al. 2008] Yu, H., Ng, T.-T. e Sun, Q. (2008). Recaptured photo detection using specularity distribution. In *Intl. Conf. on Image Processing*, San Diego, California. IEEE.