IEEE
Signal Processing
MAGAZINE

ARTIGOS REEDITADOS DA IEEE SIGNAL PROCESSING MAGAZINE – 2011

IEEE Signal Processing Society

IEEE

Renove a sua afiliação à IEEE para 2012 e torne-se também sócio da Signal Processing Society

Os sócios da Signal Processing Society (SPS) recebem os seguintes benefícios:

- ✔ Uma assinatura da *Signal Processing Magazine* (SPM) do IEEE
- ✔ Acesso eletrônico a 7 publicações da SPS
- ✔ O *Content Gazette*, contendo o índice das 7 publicações
- ✔ Informativo eletrônico Insight *Signal Processing*
- ✔ Desconto na taxa de inscrição para reuniões da Sociedade
- ✔ Direito a bolsa-auxílio para viagens relacionadas a conferências

IEEE Signal Processing Society®

SÓCIO NÍVEL 1
US$ 20/ Estudantes US$ 10

Inclui:
- 6 edições digitais da SPM por ano
- 12 edições do informativo eletrônico Insight SP por ano

SÓCIO NÍVEL 2
US$ 35/ Estudantes US$ 18

Inclui:
Todos os benefícios do nível 1

MAIS
- 6 edições impressas da SPM
- Acesso eletrônico a 7 publicações da SPS
- 6 edições impressas do Content Gazette por ano

Quando for renovar a sua associação neste outono, visite: **www.ieee.org/renew**

IEEE

Li Deng,
Editor-Chefe and
E Linda Cherry, Gerente de Publicações,
IEEE Signal Processing Society

Impacto Global da Signal Processing Magazine

Todos os anos, o Journal Citation Reports (JCR) da Thomson ISI analisa o impacto das publicações científicas, determinando com que frequência os artigos dessas publicações são citados em pesquisas posteriores.

Pelo segundo ano em seguida, a Signal Processing Magazine da IEEE ocupa o primeiro lugar entre as 247 publicações da categoria de engenharia elétrica e eletrônica no mundo inteiro e o primeiro lugar entre as 127 publicações da IEEE.

Nossa meta é publicar artigos de alta qualidade que causem impacto. No relatório da ISI publicado em 2010, o fator de impacto era 5,86, um aumento considerável em relação a 2009, quando o impacto da revista foi de 4,91. O impacto social [da tecnologia de] processamento de sinais e a sua grande variedade de aplicações são a razão principal do impacto de nossa revista. A influência da SPM e da IEEE vai longe. As informações científicas da IEEE estão por trás de muitas inovações tecnológicas. Os artigos científicos e técnicos da IEEE são citados três vezes mais do que os de qualquer outra publicação acadêmica ou comercial.

Uma característica importante de nossa revista é a ampla cobertura das mais diversas áreas técnicas, mas sempre com acuidade e rigor matemático. A meta é ajudar a organizar tópicos abrangentes [da área] de processamento de sinais, defendendo, ao mesmo tempo, a unificação das metodologias existentes. Muitas técnicas de processamento de sinais são eficazes, não somente numa única área de aplicação, mas, frequentemente também em outras áreas, apesar da necessidade de se usar com discernimento uma certa intuição em algumas áreas específicas. Uma boa parte de nossa revista tem o objetivo de apresentar artigos instrutivos com a interação de diferentes áreas de aplicação com metodologias comuns e tratamento rigoroso. Todas as nossas edições especiais são preparadas por equipes editoriais de primeira linha.

Nossos editores têm trabalhado arduamente para expandir o alcance da SPM. A edição digital da revista é enviada por e-mail a leitores do mundo inteiro e dá aos nossos membros um acesso rápido e fácil a pesquisas de ponta e tendências inovadoras do setor. Nossas traduções possibilitarão a um número ainda maior de estudantes e praticantes acessar informações científicas inovadoras.

Como membros da comunidade de processamento de sinais, vamos abraçar a vitalidade de nossa área, celebrar o impacto global do nosso trabalho e compartilhar a honra de ocupar o primeiro lugar no número de citações feitas, o que comprova a qualidade do nosso trabalho e a influência social de nosso campo.

[SP]

Li Deng, Editor Chefe e
Linda Cherry, Gerente de Publicações da
Signal Processing Society do IEEE

Lançamento da edição brasileira da Signal Processing Magazine (SPM)

Em setembro último, cumprindo a estratégia global da IEEE, a Signal Processing Magazine do IEEE (SPM) lançou a primeira edição de artigos já publicados, traduzidos para o idioma chinês simplificado. A edição virtual foi publicada no website da SPS e cópias impressas, juntamente com uma pesquisa de opinião, foram distribuídas aos participantes da ICIP, realizada em Hong Kong entre 26 e 29 de setembro de 2010. As respostas à pesquisa foram bastante positivas, com diversas sugestões para aperfeiçoamento das futuras traduções. Em agosto deste ano, lançamos outra versão chinesa, que foi distribuída juntamente com a revista Spectrum na China e em Taiwan.

Com esta edição, estamos estendendo nosso trabalho de tradução para o Brasil, um país importante para a comunidade de engenharia. O Brasil é um dos principais países em termos de crescimento econômico. Nessa edição brasileira, incluímos uma matéria sobre aplicações de videovigilância e três matérias didáticas escolhidas cuidadosamente. Essa tradução reflete as últimas tendências do segmento e traz matérias educativas sobre a teoria, métodos e aplicações conhecidas do processamento de sinais.

Esta edição traduzida que você está recebendo foi enviada juntamente com a revista Spectrum a todos os membros da IEEE no Brasil e em Portugal.

Para tornar a comunicação ainda mais eficiente, envolvemos parceiros estratégicos do setor e de instituições educacionais. Esta edição será publicada por ocasião do Workshop Internacional da IEEE sobre Análise Forense de Informações e Segurança (WIFS) 2011, que será realizado em Foz do Iguaçu, Brasil, entre 16 e 19 de novembro de 2011.

Apesar da popularidade do inglês entre a comunidade científica, ainda existem barreiras relacionadas ao idioma. Uma descoberta muito interessante da nossa pesquisa destaca a importância do nosso trabalho de tradução de forma esclarecedora. Ou seja, muitos estudantes e profissionais da área de processamento de sinais acharam as nossas matérias traduzidas úteis, não por não entenderem bem o inglês escrito, mas por desejarem escrever melhor em inglês. O avanço em suas carreiras exige a publicação de seu trabalho de pesquisa em inglês. Poder acessar rapidamente uma versão em inglês das matérias da SPM os ajudará a escrever melhor, em inglês, os seus artigos sobre processamento de sinais.

Nosso trabalho de tradução foi guiado pelo princípio do "global = global + local", no qual uma visão global é personalizada para refletir o gosto e a cultura local. Enquanto você estiver usando e lendo esta revista, mantenha esse princípio em mente e, sempre que possível, envie-nos a sua opinião sobre quais alterações poderiam ser feitas nas edições futuras para que elas possam ser mais úteis no seu trabalho.

Agradecemos ao Prof. Dr. Rodrigo Capobianco Guido por sua ajuda na edição dos textos.

[SP]

GLOSSARY OF ACADEMIC TERMS

English	Portuguese
Bayesian	bayesiano/ bayesiana
clustering	agrupamento, *clustering*
compressed sensing	percepção comprimida
computer code	código de computador
computer vision	visão computacional
data	dados
digital	digital
factor	fator
global minimum	mínimo global
hierarchical	hierárquico/ hierárquica
image	imagem
iterative	iterativo/ iterativa
labeling	rotulagem
least square	quadrado mínimo
module	módulo
Monte Carlo	Monte Carlo
noise	ruído
operator	operador/ operadora
optimization	otimização
Physical layer	camada física
probability	probabilidade
program	programa
read-only memory	ROM
real time	tempo real
robustness	robustez
sampling	amostragem
sensor	sensor
signal	sinal
transistor	transistor
variable	variável
variance	variância
video	vídeo

Ron Schneiderman

Trends in Video Surveillance Give DSP an Apps Boost

They're everywhere. Literally millions of video surveillance cameras tracking our every move just about everywhere in the world. And you can't even see most of them.

We're all familiar with video cameras placed in convenience and food stores and other retail stores and catching people sneaking into "secure" office and apartment buildings. We're also familiar with the use of video technology in monitoring at airports.

One of the most publicized cases of the use of video surveillance was the failed car bombing in New York City's Times Square in May, which was recorded by multiple video cameras in the area.

The Mineta San Jose International Airport upgraded its video surveillance system earlier this year, from a primarily analog to an IP-based network system that enables security personnel to query up to six cameras at a time and synchronize them. If an alarm sounds, the cameras are programmed to automatically focus on one area.

More recently, police in Seattle have come under scrutiny after an officer was caught on camera punching a 17-year-old girl in the face during a traffic stop.

Video camera-equipped unmanned aircraft are used in the war zones of Afghanistan and Iraq to collect intelligence and to monitor the southern borders of the United States.

And in the realm of advances in video analytics, the New Meadowlands Stadium in New Jersey, the home of the New York Giants football team, security officials will have access to special bracelets for children. If children get lost, the stadium's video surveillance system can almost immediately locate them on a TV screen.

Indeed, advances in technology are turning video surveillance into a growing opportunity for the digital signal processing (DSP) community.

SEVERAL TECH TRENDS

Several trends are currently working in favor of advancing the state of the art of video surveillance equipment.

The big trend is the rapid shift from analog to higher-quality, higher-resolution digital video. Analog cameras top out at a maximum resolution of 720 × 480 at 30 frames per second, but most analog security systems don't even operate at that level. (Analog cameras currently represent more than 85% of the installed base of video surveillance cameras.) End users are also demanding that video surveillance cameras be upgraded to high definition, which must be digital. They also want the flexibility to add more cameras on the fly, even wirelessly, and monitor and control surveillance from just about anywhere. (Motorola believes that law enforcement vehicles will someday be able to record four separate license plate numbers from vehicles moving in any direction and then instantly match them against a database to determine if any of those vehicles—or their owners—are subject to outstanding warrants or other violations.)

And then there's the adoption of analytics, enabling video to be analyzed by machines rather than humans.

Open standards, including the Open Network Video Interface Forum (ONVIF) and the Physical Security Interoperability Alliance (PSIA), will have a significant and positive impact on the market and help accelerate the transition to network video surveillance. The H.264 high-definition video compression standard is already helping to speed the adoption of Internet protocol (IP) video surveillance.

"The market is still dominated by second-generation, or so-called hybrid, systems," says Michael L. Long, security and surveillance segment manager for Analog Devices, Inc. (ADI). "These are mainly analog cameras, or some intermediate systems, that do signal conversion and compression and transmission of compressed video streams for storage."

Long says, "We see more signal processing moving into the camera itself. That's what the market will require over the next five to ten years." (IP network cameras have their own IP address and the computing functions necessary to handle network communication.)

ADI formed its video surveillance business unit late last year, and, to date, most of its products have come directly out of the company's standard product catalog. "Moving forward, our new products will target the video surveillance market, and by the end of 2011, we will be bringing to market a more integrated video system-on-a-chip (SoC) for IP cameras," notes Long.

To pull this off will require improved sensors and video data converters for improved image capture. But even more, Long says, the market will require greatly improved processors to handle, process, and analyze these higher-quality images. To transmit these images in real time over a wireless network, they will have to be compressed with the latest algorithms, such as MPEG4 and H.264.

But Altera Corp. also believes it has a serious shot at this market with its field-programmable gate array (FPGA) technology and recently announced a

Digital Object Identifier 10.1109/MSP.2010.938113

Artigos reeditados da IEEE Signal Processing Magazine – 2011

1053-5888/10/$26.00©2010IEEE

Ron Schneiderman

Tendências em videovigilância alavancam aplicações de PDS (DSP)

Elas estão em todos os lugares. Literalmente milhões de câmeras de videovigilância observando cada passo que damos em todos os lugares do mundo. E não conseguimos nem ver onde estão.

Todos nós conhecemos aquelas câmeras de vídeo instaladas em lojas de conveniência, mercados e outros pontos de varejo, flagrando pessoas entrando sorrateiramente em prédios "seguros" de escritórios e apartamentos. Também estamos acostumados com o uso de tecnologias de vídeo para a vigilância de aeroportos.

Um dos casos mais famosos do uso de videovigilância foi a tentativa fracassada de explosão de um carro na Times Square, em Nova Iorque, no mês de maio, gravado por diversas câmeras instaladas no local.

O Aeroporto Internacional Mineta San Jose atualizou o seu sistema de videovigilância no início do ano, trocando o seu equipamento analógico por um sistema em rede baseado no protocolo IP, que permite ao pessoal da segurança usar seis câmeras ao mesmo tempo e sincronizá-las. As câmaras estão programadas para focalizar uma determinada área se um alarme soar.

Mais recentemente, uma investigação foi iniciada contra a polícia de Seattle, após um policial ser flagrado por uma câmera esbofeteando uma garota de 17 anos durante uma fiscalização de trânsito.

Aviões pilotados remotamente e equipamentos com câmeras são usados para coletar informações em zonas de guerra do Afganistão e Iraque e monitorar as fronteiras do sul dos Estados Unidos.

E no domínio dos avanços no campo da análise inteligente de vídeo (*video analytics*), no estádio New Meadowlands, em Nova Jersey, casa do time de futebol americano Giants, os seguranças, em breve, distribuirão braceletes especiais para crianças. No caso de uma criança se perder, o sistema de videovigilância do estádio conseguirá localizá-la rapidamente numa tela de TV.

Digital Object Identifier 10.1109/MSP.2010.938113

Com certeza, os avanços tecnológicos estão transformando a videovigilância numa grande oportunidade para a comunidade da área de processamento digital de sinais (PDS).

AS TENDÊNCIAS TECNOLÓGICAS

Várias tendências estão impulsionando os avanços na área de desenvolvimento de equipamentos de videovigilância..

A tendência do momento é a rápida mudança das tecnologias analógicas para o vídeo digital de maior qualidade e maior resolução. As câmeras analógicas atingem uma resolução máxima de 720 x 480 a 30 quadros por segundo, mas a maioria dos sistemas de segurança analógicos não conseguem atingir esse nível (As câmeras analógicas representam mais do que 85% da base de câmeras de segurança instaladas). Há também uma demanda dos usuários finais, que desejam câmeras com maior definição e, obrigatoriamente, digitais. Eles também querem ter flexibilidade para acrescentar mais câmeras, inclusive sem fio, tão logo surja a necessidade, e monitorar e controlar a segurança a partir de qualquer lugar. A Motorola, por exemplo, acredita que os carros usados pela polícia conseguirão um dia registrar quatro placas diferentes de veículos movendo-se em qualquer direção e imediatamente compará-las com o seu banco de dados para ve-rificar se algum desses veículos - ou seu proprietário - possui restrições ou multas pendentes.

Existe também a possibilidade de se usar a análise inteligente de vídeos, ou seja, análises de vídeos feitas por máquinas e não por seres humanos.

Padrões abertos, como o *Open Network Video Interface Forum (ONVIF)* e o *Physical Security Interoperability Alliance (PSIA)* trarão um considerável e positivo impacto no mercado e ajudarão a agilizar a transição para a videovigilância em rede. O padrão H.264 de compactação de vídeos de alta definição já está ajudando a acelerar a adoção de sistemas de videovigilância que usam o protocolo IP.

"O mercado ainda é dominado por sistemas de segunda geração ou 'híbridos', como são chamados", disse Michael L. Long, gerente do segmento de segurança e vigilância da Analog Devices, Inc. (ADI). "Trata-se, na maior parte, de câmeras analógicas ou sistemas intermediários que fazem a conversão e compactação de sinais e a transmissão dos vídeos compactados para armazenamento."

De acordo com Long, "Vemos um maior processamento de sinais sendo feito na própria câmera. É isso que o mercado irá demandar nos próximos cinco anos." Tal observação é devida ao fato de que as câmeras em redes IP possuem o seu próprio endereço IP, além das funções necessárias para lidar com comunicação via rede.

A ADI montou a sua unidade de negócios de videovigilância no ano passado e, até hoje, a maioria dos produtos vêm diretamente do catálogo de produtos padrão da empresa. "No futuro, nossos produtos terão como alvo o mercado de videovigilância e, até o final de 2011, lançaremos um sistema de vídeo integrado, do tipo *system-on-a-chip (SoC)* para câmeras IP", afirmou Long.

Para que isso seja possível, precisamos de melhores sensores e conversores de dados de vídeo para uma melhor captura das imagens. Mas, cada vez mais, o mercado demandará processadores bem mais avançados para manusear, processar e analisar essas imagens de alta qualidade. Para transmitir as imagens em tempo real usando redes sem fio, algoritmos avançados, tais como MPEGG4 e H.264 deverão ser usados para fazer a conversão.

A Altera Corporation também acredita que tem boas chances nesse mercado com a sua tecnologia de arranjo de portas programável em campo (*field-programmable gate array- FPGA*). A empresa anunciou recentemente um projeto de referência de alta definição, de sua propriedade intelectual, para câmeras de segurança num único FPGA. Judd Heape, gerente sênior

Evolving video analytic technologies is allowing video surveillance monitors to track the movement of people and assets (such as vehicles) automatically from one video camera to another in real time.

high-definition, intellectual property reference design for surveillance cameras on a single FPGA. Judd Heape, senior strategic marketing manager in Altera's Industrial Business Unit, says the all-in-one solution offers surveillance equipment manufacturers the ability to reduce board space, lower power consumption, increase flexibility, and reduce development time compared to previous architectures using traditional DSPs and application-specific standard products (ASSPs).

Altera has partnered with Apical, which provides the image processing IP. Three other partners—Ocean Logic, Eylitics, and Jointwave—provide H.264 encoders for the FPGAs. Other partners provide graphics rendering cores for on-screen display and storage interface cores for local camera storage and standard interface IP cores.

VIDEO ANALYTICS

One of the new big things in video surveillance is video analytics, which can automatically analyze video for specific data and behavior. According to Texas Instruments (TI) application notes, video analytics servers handle multiple camera inputs, digitize, compress, and stream digital media content over an IP network such as a local area network (LAN), intranet or the Internet, turning an analog video system into a network video system. Users can view live images using Web browsers or application software on any local or remote computer on a network. In actual use, video analytics can be used to count the number of pedestrians entering a door or large area and determine their location. The technology can also help identify suspicious movement of people or assets.

ADI, for example, uses an array of networked surveillance cameras in several of its buildings. "Analytics will allow us to pay attention to something that may be important rather than something that may just be white noise in the system," says ADI's Long.

"What's interesting are some of the facial recognition techniques [being developed] and being able to track someone in a crowd through a facility using multiple cameras," says Altera's Heape. "You can watch them go from floor to floor and office to office. And they're tracked automatically."

"We think you can use DSPs, or a blade server, or do analytic algorithms. But to do them well, you have to put it in the camera. FPGAs can do massive parallel processing and break images into small sections and do analytics on each section separately. We have an IP partner that's doing that," Heape adds.

Still considered a niche technology by some, video analytics is in its early stages of adoption. Nevertheless, IMS Research forecasts the video analytics market will explode over the next few years (see "Global Transition To IP Video Surveillance Sparks Market"), despite some existing issues like false alerts and cost, which reportedly can run to US$500 or more per channel.

Since the attempted terrorist bombing in Times Square in New York, Mayor Michael Bloomberg has renewed efforts to promote his Midtown Manhattan Security Initiative, a network of public and private security cameras to blanket a 30-block area.

Under the mayor's program, the area of coverage by video surveillance cameras would have analytic software that could detect suspicious behavior, such as a vehicle cruising several times around the same block or the identification of an unattended bag. Bloomberg has already announced plans to extend New York City's network of security cameras that can read license plates and eventually include facial-recognition software and other, unspecified features, using US$24 million in homeland security funds.

NEW TOOLS

Several products have been announced in recent months that are designed to help upgrade commercial, industrial, and government surveillance systems.

Responding to increased developer demand for low-cost, fast, and efficient video technology software tools, ADI began offering a free video library for use in real-time video analytics applications. The software modules are fully optimized for ADI's Blackfin processor line.

The free software modules are available in object code or a C source wrapper and provide advanced video design functionality such as video filtering, transforms, color operations, and utilities suitable for a wide range of applications, mainly video surveillance, but also automotive vision systems and industrial vision.

ADI is working the Avnet, a major distributor, to help developers streamline the design process in video surveillance

Com as avançadas tecnologias de análise de vídeos, os monitores de videovigilância seguem o movimento de pessoas e objetos (como veículos) automaticamente de uma câmera de vídeo para outra em tempo real.

de marketing estratégico da unidade de negócios industriais da Altera, disse que a solução integrada oferece aos fabricantes de equipamentos de videovigilância condições de reduzir o espaço de placa e o consumo de energia, aumentar a flexibilidade e diminuir o tempo de desenvolvimento, em comparação às arquiteturas anteriores que usam PDSs tradicionais e produtos-padrão específicos para uma única aplicação (*application-specific standard products - ASSP*).

A Altera formou uma parceria com a Apical, que fornece o IP para processamento das imagens. Outras três parceiras, Ocean Logic, Eylitics e Jointwave, fornecem codificadores H.264 para os FPGAs. Outras empresas parceiras fornecem núcleos (*cores*) de renderização gráfica para display em tela, núcleos de interface para armazenamento local em câmeras e núcleos de interface padrão IP.

ANÁLISE INTELIGENTE DE VÍDEO

Uma das grandes novidades na área de videovigilância é a análise inteligente de vídeos (*video analytics*) que consiste na análise automatizada de vídeos para buscar dados e comportamentos específicos. De acordo com as notas de aplicativo da Texas Instruments, os servidores de análise inteligente de vídeos lidam com vários inputs de câmeras, digitalizam, compactam e transmitem conteúdos de mídia digital em redes IP, tais como as *local area networks* (LAN), Intranet ou Internet, transformando um sistema de vídeo analógico num sistema de vídeo em rede. Os usuários conseguem visualizar as imagens ao vivo por meio de *browsers* da Internet ou aplicativos em qualquer computador local ou remoto de uma rede. Na prática, a análise inteligente de vídeos pode ser usada para contar o número de pedestres que entra por uma porta ou numa grande área e determinar a sua localização. A tecnologia também pode ser usada para ajudar a identificar movimentos suspeitos de pessoas ou objetos.

A ADI, por exemplo, usa um conjunto de câmeras de vigilância em rede em vários de seus prédios. "Com o sistema de análise inteligente conseguimos prestar atenção num detalhe que pode ser importante, em vez de coisas que podem ser somente um ruído do sistema", explica Long, da ADI.

"As técnicas de reconhecimento facial (em desenvolvimento) são bem interessantes, assim como a possibilidade de se identificar alguém numa multidão usando várias câmeras", diz Heape, da Altera. "Podemos observá-los indo de um andar para outro, de um escritório para outro. E isso se dá automaticamente."

"Acreditamos que seja possível usar PDSs ou um servidor *blade*, ou ainda, algoritmos analíticos. Mas para ter sucesso, é necessário embuti-los na câmera. Os FPGAs conseguem executar uma grande quantidade de processamento paralelo e dividir as imagens em pequenas seções, realizando a análise de cada seção separadamente. Trabalhamos com uma empresa parceira para IP e é isso que ela está fazendo," complementa Heape.

Ainda considerada uma tecnologia de nicho por alguns, a análise inteligente de vídeos está em fase inicial de adoção. Mesmo assim, a IMS Research prevê que o mercado irá explodir nos próximos anos (veja *"Transição Global para Videovigilância IP Aquece o Mercado"*), apesar da existência de alguns problemas, como alarmes falsos e o custo, que pode chegar a 500 dólares ou mais por canal.

Desde o atentado terrorista ocorrido na Times Square em Nova Iorque, o prefeito Michael Bloomberg tem se empenhado para promover a sua *Midtown Manhattan Security Initiative*, uma rede de câmeras de segurança públicas e privadas que cobrirão uma área de 30 quarteirões.

Como parte do programa, a área de cobertura por câmeras de segurança contaria com software analítico capaz de detectar comportamentos suspeitos, tais como um veículo passando várias vezes pelo mesmo quarteirão, e identificar sacolas abandonadas. Bloomberg já anunciou planos para expandir a rede de câmeras de segurança de Nova Iorque que conseguem ler placas de carro e, no futuro, incluir software de reconhecimento facial, além de outras funções ainda não especificadas, usando os 24 milhões de dólares liberados pelo Departamento de Segurança Nacional.

NOVAS FERRAMENTAS

Vários produtos lançados nos últimos meses foram desenvolvidos para aprimorar os sistemas de vigilância comercial, industrial e pública.

Respondendo à crescente demanda dos desenvolvedores por ferramentas de tecnologia de vídeo mais baratas, rápidas e eficientes, a ADI passou a oferecer gratuitamente bibliotecas de vídeos para uso em aplicações de análise inteligente de vídeos em tempo real. Os módulos do software são otimizados para uso com a linha de processadores Blackfin da ADI.

Os módulos do software gratuito estão disponíveis em código-objeto ou '*C source wrapper*' e oferecem funções avançadas de desenvolvimento de vídeos, tais como filtragem de vídeos, transformadas, operações com cores e várias utilidades que aceitam uma ampla gama de aplicações, principalmente a videovigilância, mas também sistemas de visão automotiva e visão industrial.

GLOBAL TRANSITION TO IP VIDEO SURVEILLANCE SPARKS MARKET

The economic downturn had an impact on the global video surveillance equipment market in 2009, but the transition from analog to digital and image processing (IP) is growing the market again.

"There are a host of new trends that will keep the industry talking and drive market resurgence," says Alastair Hayfield, research manager at IMS Research.

Hayfield also sees storage as an increasingly big issue in video surveillance. The latest data from IMS Research forecasts the world market for video surveillance storage is that it will exceed US$5.6 billion in 2013.

IT network storage, in particular IP storage area networks (IP SANs), is at the forefront of new enterprise surveillance projects, according to the company's market research. Network storage is to account for more than 30% of world video surveillance storage revenue in 2013.

Although the storage market is currently dominated by digital video recorder (DVR) solutions, IMS Research says network surveillance storage is gaining traction and is more scalable and flexible to the needs of some end users.

The market research firm says that the most forward-looking end users are asking for new technologies, such as network storage or video surveillance as a service (VSaaS), which usually calls for low camera count installations with minimal recording capabilities and little requirement for constant monitoring. The market is current small and nascent. However, if brought to the market in the right way and at the right time, Hayfield says it has potential to disrupt the traditional DVR market.

Still, the DVR market should not be ruled out, as there will continue to be a sizeable need for a locally recorded "plug and play" solution, particularly in low-end applications.

Another plus for the market is that network camera prices are expected to drop in 2010 as a result of increasing competition, lower cost imports, and manufacturer desire to penetrate price-sensitive markets.

Remote video monitoring is forecast to be the fastest growing service within the remote monitoring market as it is experiencing significant growth globally.

"Customers who are able to replace guards with a monitoring system are reportedly seeing return on investment in a relatively short period of time," according to an IBM Research report, which also notes that insurance companies are listing remote video monitoring as an equivalent substitute for guards and are reducing premiums accordingly.

Law enforcement agencies across the United States are increasingly installing video surveillance equipment in police cars. Even with recent funding cuts, the long-term outlook is optimistic. (More than 40% of the 450,000 police vehicles in the United States are already equipped with digital video surveillance gear.)

Beyond the United States, the Middle East is a relatively immature video surveillance market, but it is moving quickly to adopt IP video surveillance. IMS Research warns that companies hoping to exploit this region will have to tailor their marketing strategy to the unique market structure that exists in the Middle East.

China is another major market opportunity. But while the market for analog equipment in China is vast, the shift toward IP surveillance equipment is already well underway with growing interest in networking security cameras. More than half of the respondents to a survey in China indicated that more than 25% of the cameras installed in that country over the next three years will be networked. (Half of that number suggested it would be more than 50%.)

One potential development that could change the global competitive landscape of this product sector over the next few years is mergers and acquisitions, possibly even of some of the key players.

by providing a digital video surveillance kit with an image sensor, Blackfin processor, and software development tools for video capture and analysis.

TI has a single-platform H-264 reference design based on the TMS320DM365 digital video processor with DaVinci technology and the TI TVP5158 multichannel video decoder for faster development.

Also new, from Montreal, Canada-based Octasic, Inc., is a series of multicore DSP devices for video processing that can simultaneously support two channels of 1,080p while consuming less than 3 W of power. The same OCT2224M device can process hundreds of channels simultaneously, allowing for any combination of hit rates and resolutions.

In July, Via:Sys Intelligent Video Analytics and Frankfurt, Germany-based Surveillance Grid Co. announced a partnership to protect North America's commercial solar power industry with a comprehensive video surveillance system armed with state-of-the-art video analytics technology developed by Viv:Sys. Surveillance Grid also will deploy the perimeter security system that is installed throughout Europe's emerging solar industry and is accepted by insurance companies as commercial grade security protection.

In China, Nanjing Topsky Technology Co. Ltd., working with the University of Hong Kong, has developed an intelligent video surveillance device called ThinkSmart that can monitor live video data, detect, track and classify objects, analyze behavior, and send alerts in real time. ThinkSmart's algorithms are customizable for special security requirements. The device, which uses ADI's Blackfin BF561 DSP, works seamlessly with existing security and surveillance systems. It can also be integrated into DVRs.

Toshiba Surveillance and IP Network Video, a business unit of Toshiba America Information Systems, has also introduced an IP network camera (IK-WB30A) featuring an advanced two-megapixel complementary metal–oxide–semiconductor sensor.

GE Intelligent Platforms has also announced the ICS-8580, a rugged video XMC module for video streaming from unmanned vehicles. The module uses two

TRANSIÇÃO GLOBAL PARA VIDEOVIGILÂNCIA IP AQUECE O MERCADO

A retração econômica afetou o mercado global de equipamentos de videovigilância em 2009, mas, com a transição do sistema analógico para o digital e o processamento de imagens (PI), o mercado está crescendo novamente.

"Há uma série de novas tendências que movimentarão o setor e trarão o crescimento de volta ao mercado", disse Alastair Hayfield, gerente de pesquisas da IMS Research.

Hayfield também acredita que o armazenamento é um ponto cada vez mais importante em videovigilância. Os últimos estudos da IMS Research preveem que o mercado mundial de armazenamento em videovigilância ultrapassará 5,6 bilhões de dólares em 2013.

O armazenamento em rede, principalmente em redes de armazenamento que usam o protocolo IP (*IP storage area networks - IP SANs, na sigla em inglês*), encabeçarão os projetos de vigilância corporativa, de acordo com as pesquisas da empresa. O armazenamento em rede deverá representar mais do que 30% das rendas obtidas com armazenamento em videovigilância em 2013.

Mesmo com o mercado dominado por soluções de gravação digital de vídeos (*digital video recorder - DVR*, em inglês), a IMS Research afirma que o armazenamento de vídeos de vigilância em rede está crescendo, sendo mais adaptável e flexível para atender às necessidades de alguns usuários finais.

A empresa de pesquisas de mercado diz que os usuários finais de maior visão estão demandando novas tecnologias, como armazenamento em rede e videovigilância como serviço (*video surveillance as a service - VSaaS*, em inglês), que requer poucas câmeras instaladas com alguma capacidade de gravação e menor necessidade de monitoramento. Esse mercado ainda é reduzido e nascente. No entanto, se oferecido aos consumidores da maneira certa e no tempo certo, a tecnologia poderá abalar o tradicional mercado de DVR.

Ainda assim, não podemos descartar de imediato o mercado de DVR, uma vez que ainda existe uma demanda considerável por soluções de gravação local, do tipo *plug-and-play*, principalmente aplicações de baixo custo.

Um outro fator positivo é que deve haver uma queda dos preços dos sistemas de câmeras em rede em 2010, como resultado da concorrência mais acirrada, dos menores custos de importação e do desejo dos fabricantes de penetrar os mercados mais sensíveis ao preço.

Estima-se que o mercado de monitoria remota por vídeo será o serviço do setor de monitoramento remoto que crescerá mais rapidamente, uma vez que já se registra um crescimento global considerável.

Os clientes que conseguem substituir os seus guardas de segurança por sistemas de monitoramento estão tendo um retorno mais rápido de seus investimentos, conforme apontou um relatório da IBM Research, que também afirma que as seguradoras estão considerando o monitoramento por vídeo um substituto adequado do sistema de guardas e estão reduzindo seus preços de acordo. Os departamentos de polícia dos Estados Unidos estão instalando cada vez mais equipamentos de videovigilância nos carros de suas frotas policiais. Mesmo com os recentes cortes no orçamento, a previsão de longo prazo é otimista. Particularmente, mais do que 40% dos 450.000 veículos policiais nos Estados Unidos já estão equipados com dispositivos de videovigilância digital.

Além dos Estados Unidos, o Oriente Médio, um mercado relativamente imaturo na área de videovigilância, está acelerando a adoção da videovigilância em rede IP. A IMS Research lembra às empresas que desejem explorar os mercados do Oriente Médio que terão que adaptar a sua estratégia de marketing à infraestrutura única de mercado que existe na região.

A China representa outra grande oportunidade de mercado. Apesar da grande extensão do mercado chinês de equipamentos analógicos, já se verifica um movimento na direção de equipamentos de videovigilância IP com crescente interesse na adoção de câmeras de segurança em rede. Mais da metade das pessoas entrevistadas numa pesquisa na China disseram que mais do que 25% das câmeras instaladas no país nos próximos três anos serão câmeras em rede. Metade dessa parcela indicou que o número ficaria acima de 50%.

Um possível fato que poderia mudar o cenário concorrencial global nesse setor nos próximos anos seria a realização de fusões e aquisições, possivelmente até de algumas das principais fabricantes.

A ADI trabalha com a Avnet, uma importante distribuidora, para ajudar os desenvolvedores a aprimorar o processo de desenvolvimento de sistemas de videovigilância, oferecendo um kit de videovigilância digital com um sensor de imagens, um processador Blackfin e ferramentas de desenvolvimento de software para a captura e análise de vídeos.

A Texas Instrument possui um desenho de referência para uma plataforma única H-264 baseada no processador de vídeo di-gital TMS320DM365, com tecnologia DaVinci e decodificador de vídeo com múltiplos canais para um desenvolvimento mais rápido.

A Octasic, empresa sediada em Montreal, Canadá, também lançou uma série de dispositivos *multi-core* de PDS para videovigilância, que suportam, ao mesmo tempo, dois canais de 1080p, com um consumo de menos de 3W de energia. O mesmo dispositivo OCT2224M pode processar centenas de canais simultaneamente, permitindo qualquer combinação de taxas de acerto e resoluções.

Em julho, a Via:Sys Intelligent Video Analytics e a Surveillance Grid Co, sediada em Frankfurt, Alemanha, anunciaram uma parceria para proteger o setor de energia solar norte-americano com um sistema abrangente de videovigilância equipado com avançada tecnologia de análise inteligente de vídeos desenvolvido pela Via:Sys. A Surveillance Grid também instalará o sistema de segurança de perímetro que é usado em todo o setor emergente de energia solar europeu e é aceito pelas seguradoras como proteção de seguro em nível comercial.

Na China, a Nanjing Topsky Technology Co. Ltd., trabalhando em conjunto com a Universidade de Hong Kong, desenvolveu um dispositivo de videovigilância inteligente, chamado Think-Smart, para monitorar dados de vídeos ao vivo; detectar, monitorar e classificar objetos; analisar comportamentos e enviar alertas em tempo real. Os algoritmos do Think-Smart são customizados para atender às necessidades especiais do setor de segurança. O dispositivo, que usa o PDS Blackfin BF561, funciona perfeitamente com os sistemas de segurança e vigilância existentes. Ele também pode ser integrado aos sistemas DVR.

A Toshiba Surveliillance and IP Network Video, uma divisão da Toshiba America Information Systems, também lançou uma câmera de rede IP (IK-WB30A) que vem com um sensor que usa um avançado

Video surveillance cameras such as this one are used widely, mostly for monitoring traffic and security applications. IMS Research projects that China will account for 70% of security camera shipments in 2014, with the United States purchasing 12% of global shipments and the rest of the world the remaining 18%. (Image courtesy of Analog Devices.)

TI TMS320DM6467 DSPs for processing two streams of up to 1,080p H.264 (or JPEG 2000) encoding. Up to four streams of input data can be compressed in parallel.

NOD TO DSPs

While DSPs seem to have competition for driving the technologies that will enhance many of today's and most of tomorrow's video surveillance systems, Prof. Mohan S. Kankanhalli, who is also the vice-dean of the School of Computing at the National University of Singapore, gives the nod to DSPs.

Kankanhalli analysis is based on his studies in two areas: One is multimedia surveillance, where he has been investigating the use of multiple sensors, which could be multiple cameras or different kinds of sensors, such as cameras, audio sensors, infrared cameras, and motion sensors. He also is looking at the coordination and control of multiple active cameras, where the cameras sense the environment, process the sensed data and react to the results—for instance, panning or zooming towards an object of interest or collaboratively trying to track a person so that the person is never lost.

"I believe the requirement for DSPs will only increase with time," says Kankanhalli. "More processing can and will be done at the camera nodes. It could be merely compression or it can be more sophisticated processing. With the increasing usage of IP cameras, the cameras will be linked with each other for form networks."

He points to the trend towards using multiple cameras, either to improve the surveillance quality or to increase the coverage, or both. "Once we have multiple cameras, the issue of data transmission becomes important. Moreover, the possibility of distributed computation comes into the picture."

If the data is transmitted to a central server, the data has to be compressed and some minimal amount of on-camera processing is desirable (like motion/blob detection). And if it is a large camera network, then a large amount of distributed processing is possible by spreading the computation around in the camera nodes, which can be accomplished by a DSP coprocessor.

Kankanhalli says that current research in the area of computational photography will also lead to more computation done at the camera, which he believes will require DSPs.

What's next? Just about everyone is trying to differentiate their product or service in some way. One possibility, says Heape of Altera, is the adoption of more "sensor fusion"-based technology in surveillance products in the future that would combine video cameras with infrared for range finding and other functions. [SP]

Câmeras de videovigilância como esta são bastante usadas, principalmente para monitoramento do trânsito e aplicações de segurança. A IMS Research estima que a China receberá 70% de todas as câmeras de segurança entregues em 2014; os Estados Unidos comprarão 12% do volume mundial e o resto do mundo ficará com os restantes 18%. (Foto cedida por cortesia da Analog Devices.)

semicondutor de óxido metálico complementar de dois megapixels

A GE Intelligent Plaftorms também anunciou o lançamento do seu módulo ICS-8580 XMC para transmissão de vídeo a partir de veículos não dirigidos por pessoas. O módulo usa dois PDSs TMS322DM6467 da Texas Instrument para processar dois fluxos de dados (*streams*) de ate 1.080p com codificação H.264 (ou JPEG 2000). Até quatro fluxos de dados de entrada podem ser compactados simultaneamente.

APROVAÇÃO DOS PDS

Apesar da concorrência enfrentada pelos PDSs na luta para se tornar a tecnologia que aprimorará a maior parte dos sistemas de videovigilância atuais e futuros, o professor Mohan S. Kankanhalli, vice-diretor da Faculdade de Computação da Universidade Federal de Cingapura, aprova o uso dos PDSs.

A analise de Kankanhalli baseia-se em seus estudos em duas áreas: a de vigilância multimídia, na qual o professor investiga o uso de múltiplos sensores, que poderiam ser múltiplas câmeras, ou tipos diferentes de sensores, tais como câmeras, sensores de áudio, câmeras infraverme-lhas e sensores de movimento. Ele também estuda a coordenação e controle de um conjunto de câmeras ativas, que examinam o meio-ambiente, processam os dados examinados e reagem aos resultados, realizando *panning*, ampliando um objeto de interesse ou tentando, em conjunto, acompanhar uma pessoa, sem perdê-la de vista.

"Acredito que a demanda por PDSs só crescerá", disse Kankanhalli. "Poderá haver e haverá mais processamento feito nos nós das câmeras". Isso pode incluir compactação ou um processamento mais sofisticado. Com o crescente uso de câmeras IP, as câmeras serão ligadas umas as outras, formando redes.

Ele aponta para a tendência de uso de diversas câmeras em conjunto, tanto para melhorar a qualidade da vigilância, quanto para aumentar a cobertura ou ambos. "Com o uso de múltiplas câmeras, fica evidente a importância da questão da transmissão de dados". Além disso, devemos considerar a possibilidade de computação distribuída.

Quando os dados são transmitidos a um servidor central, eles precisam ser compactados, sendo desejável haver algum processamento na própria câmera (como detecção de movimentos/ *blobs*). Com uma ampla rede de câmeras, é possível obter uma grande quantidade de processamento distribuído, realizando-se a computação nos nós das diversas câmeras, o que pode ser feito usando-se um coprocessador PDS.

Kankanhalli diz que as pesquisas atuais na área de fotografia computacional também levarão a uma maior computação feita na própria câmera, o que exigirá o uso de PDSs.

O que vem a seguir? A maioria das empresas está tentando diferenciar o seu produto ou serviço de alguma maneira. Uma possibilidade, diz Heape, da Altera, é adotar, no futuro, mais tecnologias baseadas na "fusão de sensores", que combinaria câmeras de vídeo com infravermelho para telemetria e outras funções. [SP]

por Hany Farid

Detecção de Falsificações em Imagens

Uma pesquisa

Sem dúvida vivemos numa época em que somos expostos a uma infinidade de imagens visuais. Tradicionalmente, sempre confiamos na integridade dessas imagens, até que, com o surgimento das tecnologias digitais, essa confiança começou a ser abalada. Dos tablóides à indústria da moda, passando pelos principais canais de mídia, publicações científicas, campanhas políticas e tribunais, até os *e-mails* mal-intencionados que recebemos; todos exibem fotos adulteradas com uma frequência e sofisticação cada vez maior. Nos últimos cinco anos, com o surgimento do campo de perícia forense de mídias digitais, nossa confiança nas imagens digitais tem sido restaurada. Neste artigo, vou falar sobre o que há de mais avançado nessa nova e excitante área.

Uma das maneiras propostas para se verificar a autenticidade de imagens é o uso de marcação digital (*digital watermarking*) (veja [21] e [5], por exemplo, para pesquisas em geral). A desvantagem dessa abordagem é que uma marcação digital deve ser inserida no momento da aquisição da imagem, o que limita esse método a câmeras digitais altamente equipadas. Ao contrário dessa abordagem, técnicas passivas de perícia forense de imagens funcionam na ausência de marcações e assinaturas digitais. Essas técnicas funcionam sob a premissa de que mesmo não havendo pistas visuais deixadas pelas fraudes digitais indicando adulteração, pode haver alteração das características estatísticas da imagem. As ferramentas de perícia forense de imagens podem, *grosso modo*, ser divididas em cinco categorias: 1) técnicas de análise dos pixels, que detectam anomalias estatísticas inseridas no nível dos pixels; 2) técnicas baseadas no formato, que realçam as correlações estatísticas introduzidas por um esquema de compressão com perdas (*lossy compression*); 3) técnicas baseadas na câmera, que exploram artefatos introduzidos pela lente, pelo sensor ou no pós-processamento no *chip*; 4) técnicas baseadas nas características físicas, que explicitamente modelam e detectam anomalias na interação tridimensional entre os objetos físicos, a luz e a câmera; e 5) técnicas baseadas na geometria, que realizam a medição dos objetos no mundo e suas posições em relação à câmera.

Identificador de Objetos Digitais 101109IMSP2008.931079

 1053-5888/09/$25.00©2009IEEE

Examinarei vários exemplos de ferramentas forenses dentro de cada uma das categorias mencionadas. Ao fazê-lo, omitirei, sem dúvida, algumas publicações importantes. Mesmo assim, espero que esta pesquisa ofereça uma amostragem representativa desse emergente campo da detecção de falsificações de imagens digitais.

TÉCNICAS BASEADAS NOS PIXELS

O sistema judicial geralmente confia numa série de análises forenses que incluem, entre outras, a identificação forense (ácido desoxirribonucleico [DNA] ou impressões digitais); odontologia forense (arcada dentária); entomologia forense (insetos) e geologia forense (solo). Nas ciências forenses tradicionais, todo tipo de prova física é analisada. Na esfera digital, a ênfase recai sobre o pixel - a base das imagens digitais. Descrevo, a seguir, quatro técnicas usadas na detecção de vários tipos de adulteração, cada uma das quais examina, direta ou indiretamente, as correlações que surgem no nível do pixel, de acordo com a forma específica de adulteração.

CLONAGEM

Talvez um dos tipos de manipulação de imagens mais comuns é a clonagem ("copiar e colar") de partes da imagem para ocultar uma pessoa ou objeto presente na cena. Quando bem feita, fica difícil detectar a clonagem visualmente. E como as regiões clonadas podem ter qualquer forma e localização, é impossível analisar todos os locais e tamanhos por meios computacionais. Dois algoritmos computacionalmente eficientes foram desenvolvidos para detectar zonas de imagens clonadas ([11], [34]; consulte também [23], [27] e [42]). .

AS TECNOLOGIAS MODERNAS POSSIBILITAM A ALTERAÇÃO E MANIPULAÇÃO DE IMAGENS DIGITAIS DE UMA MANEIRA CONSIDERADA IMPOSSÍVEL 20 ANOS ATRÁS.

Os autores em [11] aplicam uma transformada discreta de cosseno (*Discrete Cosine Transform* - DCT) em blocos da imagem. As regiões duplicadas são detectadas classificando-se lexicograficamente os coeficientes dos blocos da DCT e agrupando-se os blocos semelhantes com o mesmo deslocamento espacial da imagem. Numa abordagem relacionada, os autores em [34] aplicam a análise dos componentes principais (*Principal Component Analysis* - PCA) em pequenos blocos de imagens de tamanho fixo para obter uma representação com dimensões reduzidas. As regiões duplicadas são novamente detectadas por meio de classificação lexicográfica e do agrupamento de todos os blocos de imagem. Tanto a representação por DCT como a por PCA é usada para reduzir a complexidade computacional e garantir que a detecção da clonagem seja eficaz em ambientes onde a imagem esteja sujeita a pequenas variações em função dos ruídos aditivos ou da compressão com perdas.

REAMOSTRAGEM

Para a criação de uma composição de imagens convincente, geralmente é necessário redimensionar, girar ou esticar partes da imagem. Por exemplo, ao se criar uma composição de duas pessoas, talvez seja preciso alterar o tamanho de uma delas para se obter alturas semelhantes. Esse processo requer a reamostragem da imagem original para obter uma nova matriz de amostragem, introduzindo correlações periódicas específicas entre pixels vizinhos. Como geralmente essas correlações não acontecem naturalmente, a sua presença pode ser usada para detectar esse tipo específico de manipulação [36]. Outras abordagens relacionadas são descritas em [43], [22], [31] e [38].

Vejamos um exemplo simples de interpolação de um sinal unidimensional $x(t)$ com comprimento m por um fator de dois, usando interpolação linear para obter $y(t)$. As amostras ímpares do sinal interpolado reproduzem os valores do sinal original: $y(2i-1) = x(i), i = 1, \ldots, m$, enquanto as amostras pares são a média dos valores dos pixels vizinhos do sinal original.

$$y(2i) = 0.5x(i) + 0.5x(i+1). \tag{1}$$

Como cada amostra do sinal original pode ser encontrada no sinal reamostrado, os pixels interpolados podem ser expressos somente em termos de amostras reamostradas.

$$y(2i) = 0.5y(2i-1) + 0.5y(2i+1). \tag{2}$$

Assim sendo, por todo o sinal reamostrado, cada amostra par apresenta exatamente a mesma combinação linear de seus dois vizinhos adjacentes. Nesse caso simples, o sinal reamostrado pode ser detectado observando-se que uma amostra em cada duas é perfeitamente correlacionada às suas vizinhas. Essa correlação não está limitada à interpolação por um fator de dois. Uma ampla série de reamostragens reproduz correlações periódicas semelhantes. Quando a forma específica das correlações da reamostragem é conhecida, consegue-se determinar quais pixels estão correlacionados aos seus vizinhos. Quando sabemos quais pixels estão correlacionados aos seus vizinhos, a forma específica das correlações pode ser facilmente determinada. Mas, na prática, não conhecemos nem um nem outro. O algoritmo de maximização da expectativa (*Expectation Maximization - EM*) é usado para resolver, simultaneamente, cada um desses problemas. O algoritmo EM é iterativo e de dois passos: 1) no passo da expectativa, é estimada a probabilidade de cada pixel estar correlacionado ao seu vizinho e (2) no passo da maximização, é estimada a forma específica da correlação entre os pixels. Supondo-se um modelo de interpolação linear, o passo da expectativa se reduz a um estimador bayesiano e o passo da maximização se reduz a uma estimação dos mínimos quadrados ponderados. Em seguida, usa-se a probabilidade estimada para determinar se uma parte da imagem foi reamostrada.

COMBINAÇÃO (SPLICING)

Uma forma comum de manipulação de fotos é a combinação digital de duas ou mais imagens para formar uma única composição. Quando realizada corretamente, a borda entre as zonas combinadas é visualmente imperceptível. Em [7] e [32], no entanto, os autores demonstram que a combinação perturba as estatísticas de ordem superior de Fourier, que podem ser usadas para detectar que a técnica foi aplicada.

Vamos considerar um sinal unidimensional $x(t)$ e sua transformada de Fourier $X(\omega)$. O espectro de potência $P(\omega) = X(\omega)X^*(\omega)$ é comumente usado para analisar a composição da frequência de um sinal ($*$ denota um conjugado complexo). Indo além do espectro de potência, o bispectro

$$B(\omega_1,\omega_2) = X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2) \tag{3}$$

mede as correlações de ordem superior entre triplas de frequência ω_1, ω_2, e $\omega_1 + \omega_2$. Algumas descontinuidades sutis que surgem com a combinação aparecem com um aumento da magnitude do biespectro e num desvio na fase de biespectro, sendo usadas para detectar se houve combinação de áudio [7] e imagens [32].

ESTATÍSTICAS

Existem, 256^{n^2} imagens possíveis de 8 bits em escala de cinza de tamanho $n \times n$. Mesmo com somente $n = 10$ pixels, existem, surpreendentemente, 10^{240} imagens possíveis (mais do que o número estimado de átomos no universo). Se escolhêssemos uma imagem aleatoriamente entre essa fantástica quantidade de imagens possíveis, seria extremamente improvável obter uma imagem perceptivamente significativa. Com base nessas observações, podemos deduzir que as fotografias contêm propriedades estatísticas específicas. Os autores em [9], [1] e [2] exploram as regularidades estatísticas de imagens naturais para detectar vários tipos de manipulação.

Os autores em [9] computam estatísticas de primeira ordem e de ordem superior a partir da decomposição *wavelet*. Esse tipo de decomposição divide o espaço de frequência em escalas múltiplas e sub-bandas de orientação. O modelo estatístico é composto pelos quatro primeiros momentos estatísticos de cada sub-banda *wavelet* e estatísticas de ordem superior, que capturam as correlações entre as diversas sub-bandas. As imagens são classificadas com base nessas características estatísticas por meio da classificação supervisionada de padrões. Numa abordagem complementar, os autores em [1] constroem um modelo estatístico baseado nas estatísticas de co-ocorrência de planos de bits nas imagens. Especificamente, os primeiros quatro momentos estatísticos são calculados a partir da frequência das concordâncias e discordâncias de bits nos planos de bits. Nove características que incorporam similaridade binária dos *strings* são extraídas dessas medições. Outras oito características são extraídas dos histogramas dessas medições. O algoritmo de busca sequencial flutuante para frente é usado para selecionar as características mais descritivas, que são, por sua vez, usadas num classificador de regressão linear para distinguir imagens autênticas de imagens manipuladas. Em ambos os casos, o modelo estatístico é usado para detectar várias coisas, inclusive manipulações básicas de imagens, tais como redimensionamento e filtragem [1]; diferenciação de imagens fotográficas das imagens geradas em computadores [29] e detecção de imagens ocultas (esteganografia).

NO ENTANTO, O CAMPO DA ANÁLISE FORENSE DE IMAGENS TEM FEITO E CONTINUARÁ FAZENDO COM QUE SEJA CADA VEZ MAIS DIFÍCIL E DEMORADO (MAIS NUNCA IMPOSSÍVEL) CRIAR UMA FRAUDE QUE NÃO SEJA DESCOBERTA.

TÉCNICAS BASEADAS NO FORMATO

A primeira regra em qualquer análise forense é "preservar as provas". Sendo assim, os esquemas de compressão de imagens com perdas, tais como JPEG, podem ser considerados os piores inimigos dos investigadores forenses. Por isso, é irônico saber que as propriedades únicas da compressão com perdas podem ser exploradas nas análises forenses. Descrevo, a seguir, três técnicas forenses que detectam a adulteração de imagens comprimidas, cada uma das quais realça os detalhes do esquema de compressão com perdas do tipo JPEG.

QUANTIZAÇÃO JPEG

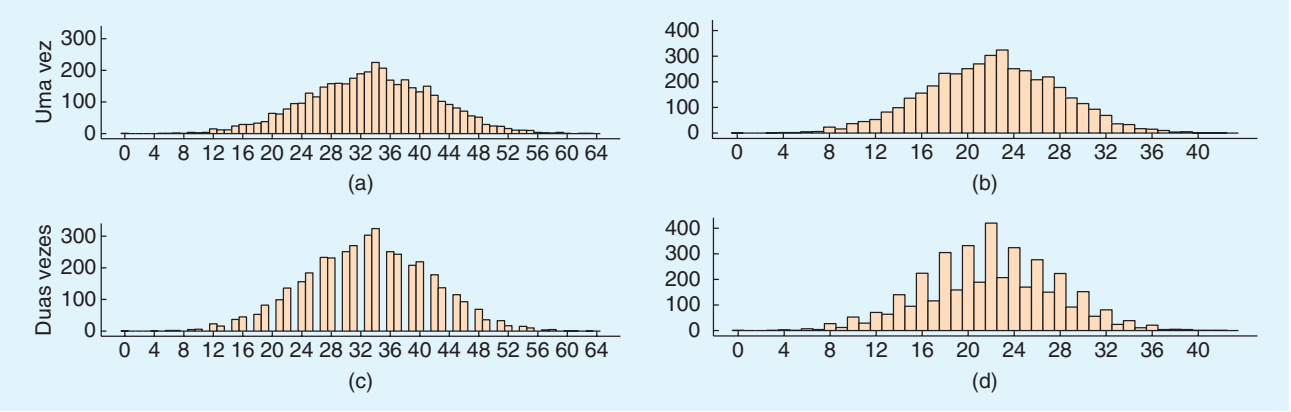
A maioria das câmeras codifica imagens no formato JPEG. Esse esquema de compressão com perdas traz mais flexibilidade em termos da quantidade de compressão obtida. Os fabricantes geralmente configuram seus dispositivos de maneira diferente para que as funções de compressão e a qualidade atendam às suas próprias necessidades e gostos. Conforme descrito em [8] e [14], essas diferenças podem ser usadas para identificar a fonte de aquisição de uma imagem (marca / modelo da câmera).

Dada uma imagem colorida de três canais (RGB), o sistema de compressão JPEG funciona da seguinte forma: primeiramente, a imagem RGB é convertida num espaço de luminância / crominância (YCbCr). Os dois canais de crominância (CbCr) são geralmente subamostrados por um fator de dois em relação ao canal de luminância (Y). Em seguida, cada canal é particionado em blocos de 8×8 pixels. Esses valores são convertidos de inteiros sem sinal para inteiros com sinal (de $[0, 255]$ a $[-128, 127]$, por exemplo). Cada bloco é convertido para o espaço de frequência usando a DCT bidimensional. Dependendo da frequência e do canal escolhidos, cada coeficiente da DCT, c, é quantizado por um valor q: $\lfloor c/q \rfloor$. Esse estágio é a fonte principal da compressão. A quantização total é especificada como uma tabela de 192 valores - um conjunto de valores 8×8 associados a cada frequência para cada um dos três canais (YCbCr). Para baixas taxas de compressão, esses valores tendem na direção de um valor de 1 e aumentam para taxas mais altas de compressão. Com algumas variações, a sequência de passos descrita acima é empregada por codificadores JPEG em câmeras digitais e *softwares* de edição de fotos. A fonte principal de variação desses codificadores é a escolha da tabela de quantização. Sendo assim, um tipo de assinatura é embutido em cada imagem JPEG. As tabelas de quantização podem ser extraídas da imagem JPEG codificada ou estimada às cegas, a partir da imagem, conforme descrito em [6].

Podemos observar que as tabelas de quantização podem variar numa única câmera em função do ajuste da qualidade e, apesar de haver alguma diferença entre as tabelas, existe certa sobreposição em câmeras de diferentes marcas e modelos. Mesmo assim, essa simples observação permite uma forma crua de "balística" de imagens digitais, por meio da qual a fonte de uma imagem pode ser confirmada ou negada.

JPEG DUPLO

Qualquer tipo de manipulação de imagens requer, no mínimo, que uma imagem seja carregada num programa de edição de fotos e salva. Como a maioria das imagens é armazenada no formato JPEG, é provável que tanto as imagens originais quanto as manipuladas estejam armazenadas nesse formato. Nesse caso, a imagem manipulada é comprimida duas vezes. Devido à compressão com perdas, característica do formato JPEG, essa compressão dupla insere artefatos específicos que não estão presentes em imagens comprimidas uma só vez (supondo-se que a imagem não tenha sido recortada antes da segunda compressão). Portanto, a presença desses artefatos pode ser vista como prova de que houve manipulação [25], [35]. Observe que a compressão JPEG dupla, em si, não prova que houve adulteração ilícita. Por exemplo, é possível salvar uma imagem inadvertidamente após visualizá-la.



[FIG1] Na fileira de cima, temos histogramas de sinais quantizados uma vez com o (a) Passo 2 e (b) Passo 3. Na fileira de baixo, temos histogramas de sinais quantizados duas vezes com (c) Passo 3, seguido pelo Passo 2, e (d) Passo 2, seguido pelo Passo 3. Observe os artefatos periódicos nos histogramas de sinais quantizados duas vezes.

Conforme descrito na seção anterior, a quantização dos coeficientes da DCT, c , é a principal maneira de se realizar a compressão, denotada como $q_a(c) = \lfloor c/a \rfloor$, onde a é o passo de quantização (um número inteiro positivo). A desquantização traz os valores quantizados à sua faixa original: $q_a^{-1}(c) = ac$. Devemos notar que a quantização não é irreversível e que a desquantização não é uma função inversa da quantização. A quantização dupla, que resulta da compressão dupla, é dada por: $q_{ab}(c) = \lfloor \lfloor c/b \rfloor b/a \rfloor$, onde a e b são os passos de quantização. A quantização dupla pode ser representada como uma sequência de três passos: 1) quantização com passo b, seguida de 2) desquantização com passo b, seguida de 3) quantização com passo a. Consideremos agora um conjunto de coeficientes distribuídos normalmente na faixa $[0, 127]$. Para ilustrar a natureza dos artefatos inseridos com a quantização dupla, consideremos quatro quantizações diferentes desses coeficientes. A linha superior da Figura 1 mostra os histogramas dos coeficientes quantizados com os passos 2 e 3. A linha inferior da Figura mostra os histogramas dos coeficientes duplamente quantizados com o passo 3 seguido do passo 2 e o passo 2 seguido do passo 3. Quando o tamanho do passo diminui [Figura 1(c)], alguns *bins* do histograma ficam vazios, pois a primeira quantização coloca as amostras do sinal original em 42 *bins*, enquanto a segunda quantização os redistribui entre 64 *bins*. Quando o tamanho do passo aumenta [Figura 1(d)], alguns *bins* contêm mais amostras do que os *bins* vizinhos, pois os *bins* pares recebem amostras de quatro *bins* do histograma original, enquanto os *bins* ímpares recebem amostras de apenas dois. Em ambos os casos de quantização dupla, podemos notar a periodicidade dos artefatos introduzidos nos histogramas. É essa periodicidade que os autores em [35] exploraram para detectar uma compressão JPEG dupla. O trabalho de [13] ampliou essa abordagem para detectar traços localizados de dupla compressão.

FORMAÇÃO DE BLOCOS EM JPEG

Conforme descrito nas seções anteriores, a base da compressão JPEG é a transformada DCT em bloco. Como cada bloco de imagem de 8×8 pixels é individualmente transformado e quantizado, alguns artefatos aparecem nas bordas de blocos vizinhos na forma de arestas (*edges*) horizontais e verticais. Quando uma imagem é manipulada, esses artefatos podem ser perturbados.

Em [28], os autores caracterizam esses artefatos usando as diferenças nos valores dos pixels dentro dos blocos e em suas bordas. Essas diferenças tendem a ser menores dentro dos blocos em relação às bordas. Quando uma imagem é recortada e recomprimida, pode haver a introdução de um novo conjunto de artefatos não alinhados às bordas originais. As diferenças dos valores dos pixels dentro dos blocos e entre eles são computadas a partir de vizinhanças de 4- pixels, que são espacialmente deslocadas em relação umas às outras por um valor fixo, onde uma vizinhança fica totalmente dentro de um bloco JPEG e a outra fica adjacente ou sobreposta a outro bloco JPEG. Um histograma dessas diferenças é calculado a partir de todas as diferenças dos blocos de imagens 8×8 não sobrepostos. Uma matriz de artefatos em blocos (*blocking artifact matrix - BAM*) 8×8 é calculada como a diferença média entre esses histogramas. Para imagens não comprimidas, essa matriz é randômica, enquanto que, em imagens comprimidas, a matriz tem um padrão específico. Quando uma imagem é recortada e recomprimida, esse padrão é perturbado. A classificação supervisionada de padrões é empregada para distinguir uma BAM autêntica de uma não-autêntica.

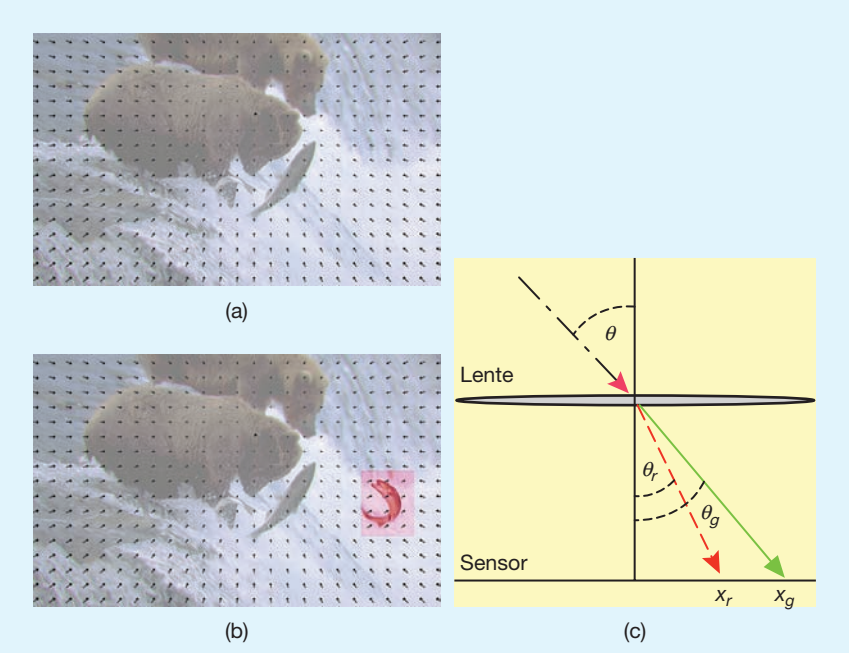
Em [41], os autores descrevem como detectar manipulações mais localizadas com base nas inconsistências apresentadas pelos artefatos em blocos. Escolhendo uma região da imagem supostamente autêntica, o nível de quantização é inicialmente estimado para cada uma das 64 frequências da DCT. As inconsistências entre os coeficientes D da DCT e o valor estimado da quantização Q são calculadas como:

$$B = \sum_{k=1}^{64} \left| D(k) - Q(k)\text{round}\left(\frac{D(k)}{Q(k)}\right)\right|. \qquad (4)$$

As variações dos valores B na imagem são usadas para detectar regiões manipuladas.

TÉCNICAS BASEADAS NA CÂMERA

As ranhuras existentes nos canos de armas de fogo fazem com que os projéteis girem rapidamente, melhorando a precisão e o alcance do tiro. Essas ranhuras causam marcas distintas nos projéteis deflagrados, podendo ser usadas para ligar o projétil a uma pistola específica. Seguindo o mesmo princípio, várias técnicas forenses, desenvolvidas para a análise de imagens, especifi-



[FIG2] Aberração cromática: (a) Um campo vetorial aparece sobreposto na imagem original mostrando o deslocamento de um pixel entre os canais vermelho e verde.(b) O peixe, copiado de outra imagem, foi adicionado a esta imagem. Sua aberração cromática é inconsistente com o padrão geral. (c) A luz policromática entra na lente num ângulo θ e emerge num ângulo que depende do comprimento de onda. Como resultado, diferentes comprimentos de luz, dois dos quais são representados pelos raios vermelho (pontilhado) e um verde (sólido), aparecerão em pontos diferentes, x_r e x_g, ,originando a aberração cromática.

camente modelam os artefatos introduzidos durante os vários estágios de processamento da imagem. Descreverei quatro técnicas usadas para modelar e estimar diversos artefatos de câmera. As inconsistências encontradas nesses artefatos podem, por sua vez, ser usadas como prova de adulteração.

ABERRAÇÃO CROMÁTICA

Num sistema ideal de aquisição de imagens, a luz passa pela lente e foca num ponto único do sensor. Porém, os sistemas óticos fogem desse padrão ideal, pois não focam perfeitamente todos os comprimentos de onda possíveis da luz. Mais especificamente, a aberração cromática lateral se manifesta como uma mudança espacial dos pontos onde a luz com diferentes comprimentos de onda atingem o sensor. Em [16], os autores mostram que essa aberração lateral pode ser definida como uma expansão ou contração dos canais de cores entre si. A Figura 2(a), por exemplo, mostra uma imagem que exibe um campo de vetores que representa o desalinhamento do canal vermelho em relação ao canal verde. A Figura 2(b) mostra a mesma imagem após o peixe ter sido incluído. Podemos observar, neste caso, que a aberração lateral nessa região adulterada é inconsistente com a aberração global. Os autores em [16] descrevem como estimar a aberração cromática lateral para detectar esse tipo de manipulação.

Na ótica clássica, a refração da luz nos limites entre dois meios é descrita pela lei de Snell: $n\sin(\theta) = n_f \sin(\theta_f)$, onde θ é o ângulo de incidência; θ_f é o ângulo de refração e n e n_f são os índices refrativos do meio por onde a luz passa. O índice refrativo do vidro, n_f, depende do comprimento de onda da luz que o atravessa. Essa dependência faz com que a luz policromática se divida de acordo com o comprimento de onda no momento em que sai da lente e atinge o sensor. A Figura 2(c), por exemplo, mostra esquematicamente como ocorre a divisão da luz com comprimento de onda curto (raio verde sólido) e com o comprimento de onda longo (raio vermelho pontilhado). A posição dos raios vermelho e verde no sensor é denotada como (x_r, y_r) e (x_g, y_g). Quando ocorre a aberração cromática, essas posições podem ser modeladas como:

$$x_r = \alpha(x_g - x_0) + x_0$$
$$\text{e} \qquad = \alpha(y_g - y_0) + y_0, \qquad (5)$$

onde α é o valor escalar e (x_0, y_0) é o centro da distorção. A estimação desses parâmetros de modelamento é mostrada como um problema de registro da imagem. Como a aberração causa um desalinhamento entre os canais de cores, os parâmetros de modelamento são estimados por meio da maximização do alinhamento dos canais de cores. Mais especificamente, as informações mútuas entre os canais de cores vermelha e verde são maximizadas (uma estimação semelhante é realizada para determinar a distorção entre os canais azul e verde). Em seguida, as estimações locais da aberração cromática são comparadas com a aberração total estimada, possibilitando detectar se houve adulteração.

CONJUNTO DE FILTROS DE CORES

Uma imagem digital colorida é composta por três canais que contêm amostras de diferentes faixas do espectro de cor, como, por exemplo, vermelha, verde e azul. A maioria das câmeras digitais, no entanto, é equipada com um único sensor CCD ou CMOS e captura imagens coloridas usando um conjunto de filtros de cores (*color filter array - CFA*). A maioria dos CFAs usa filtros de três cores (vermelha, verde e azul) colocados sobre cada elemento sensor. Como somente uma única cor é gravada em cada pixel, as outras duas cores devem ser estimadas com base nas amostras adjacentes para se obter uma imagem colorida de três canais. A determinação das cores ausentes é conhecida como interpolação CFA ou demosaico. Os métodos mais simples de demosaico são aqueles baseados em *kernels*, que agem em cada canal independentemente (por exemplo, interpolação bilienar ou bicúbica). Algoritmos mais sofisticados interpolam as arestas de maneira diferente das áreas uniformes, evitando, assim, o borramento (*blurring*) das características salientes da imagem. Independentemente da implementação específica, a interpolação CFA introduz correlações estatísticas entre um subconjunto de pixels em cada canal de cor. Como os filtros de cor de um CFA seguem geralmente um padrão periódico, essas correlações também são periódicas. Ao mesmo

tempo, é improvável que os pixels originalmente gravados exibam as mesmas correlações periódicas. Assim, essas correlações podem ser usadas como um tipo de assinatura digital.

NOS ÚLTIMOS CINCO ANOS, COM O SURGIMENTO DO CAMPO DE PERÍCIA FORENSE DE MÍDIAS DIGITAIS, NOSSA CONFIANÇA NAS IMAGENS DIGITAIS TEM SIDO RESTAURADA.

Quando conhecemos a forma específica das correlações periódicas, podemos determinar facilmente quais pixels estão correlacionados aos seus vizinhos. Por outro lado, quando sabemos quais pixels estão correlacionados aos seus vizinhos, a forma específica das correlações pode ser facilmente determinada. Na prática, não sabemos nem um nem outro. Em [37], os autores descrevem como determinar simultaneamente a forma das correlações e quais pixels apresentam ou não interpolação CFA (veja também [3] e [4]). Particularmente, os autores usaram o algoritmo EM, que é um algoritmo iterativo de dois passos: 1) no passo da expectativa, é estimada a probabilidade de cada pixel estar correlacionado ao seu vizinho e 2) no passo da maximização, é estimada a forma específica da correlação entre os pixels. Ao modelarmos as correlações CFA usando um modelo linear simples, o passo da expectativa se reduz a um estimador bayesiano e o passo da maximização se reduz a uma estimação dos mínimos quadrados ponderados. Numa imagem autêntica, espera-se um padrão periódico de pixels que sejam altamente correlacionados aos seus vizinhos; portanto, qualquer desvio desse padrão indica a existência de adulteração localizada ou geral.

RESPOSTA DA CÂMERA

Como a maioria dos sensores de câmeras digitais é praticamente linear, deve haver uma relação linear entre a quantidade de luz medida por cada elemento sensor e o valor final dos pixels correspondentes. A maioria das câmeras, no entanto, aplica uma não-linearidade pontual para realçar a imagem final. Os autores em [24] descrevem como estimar esse mapeamento, chamado função de resposta, a partir de uma só imagem. Em seguida, as diferenças na função de resposta por toda a imagem são usadas para determinar se houve adulteração (uma abordagem relacionada é descrita em [14]).

Considere uma aresta onde os pixels abaixo da aresta são de uma cor constante C_1 e os pixels acima da aresta são de uma cor diferente C_2. Se a resposta da câmera for linear, os pixels intermediários ao longo da aresta devem ser uma combinação linear das cores vizinhas. O desvio dos valores desses pixels intermediários da resposta linear esperada é usado para estimar a função de resposta da câmera. A função inversa de resposta da câmera, que traz as cores dos pixels de volta à relação linear, é estimada usando-se um estimador *maximum a posteriori* (MAP). Para estabilizar o estimador, as arestas são selecionadas de maneira que as áreas de cada lado da aresta sejam semelhantes, as variâncias em cada lado da aresta sejam pequenas, a diferença entre C_1 e C_2 seja grande e os pixels ao longo da aresta fiquem entre C_1 e C_2. Também são impostas restrições à função de resposta estimada da câmera: a função deve ser monotonicamente aumentada com no máximo um ponto de inflexão, devendo ser semelhante para cada um dos canais de cor. Como a função de resposta da câmera pode ser estimada localmente, as variações importantes dessa função por toda a imagem podem ser usadas para detectar adulterações.

RUÍDO DO SENSOR

Conforme a imagem digital se move do sensor da câmera para a memória do computador, ela passa por uma série de passos de processamento, inclusive quantização, balanceamento de branco, demosaico, correção da cor, correção gama, filtragem e, geralmente, compressão JPEG. Esse processamento insere uma assinatura distinta na imagem. Os autores em [12] modelam esse processamento usando um modelo genérico de ruído aditivo e usam as estatística do ruído estimado para a análise forense de imagens. Em [40], os autores modelam o processamento na câmera com uma série de operações de processamento e uma segunda filtragem. Em seguida, os parâmetros desse processamento em câmera são usados para determinar se uma imagem passou por algum tipo de processamento posterior.

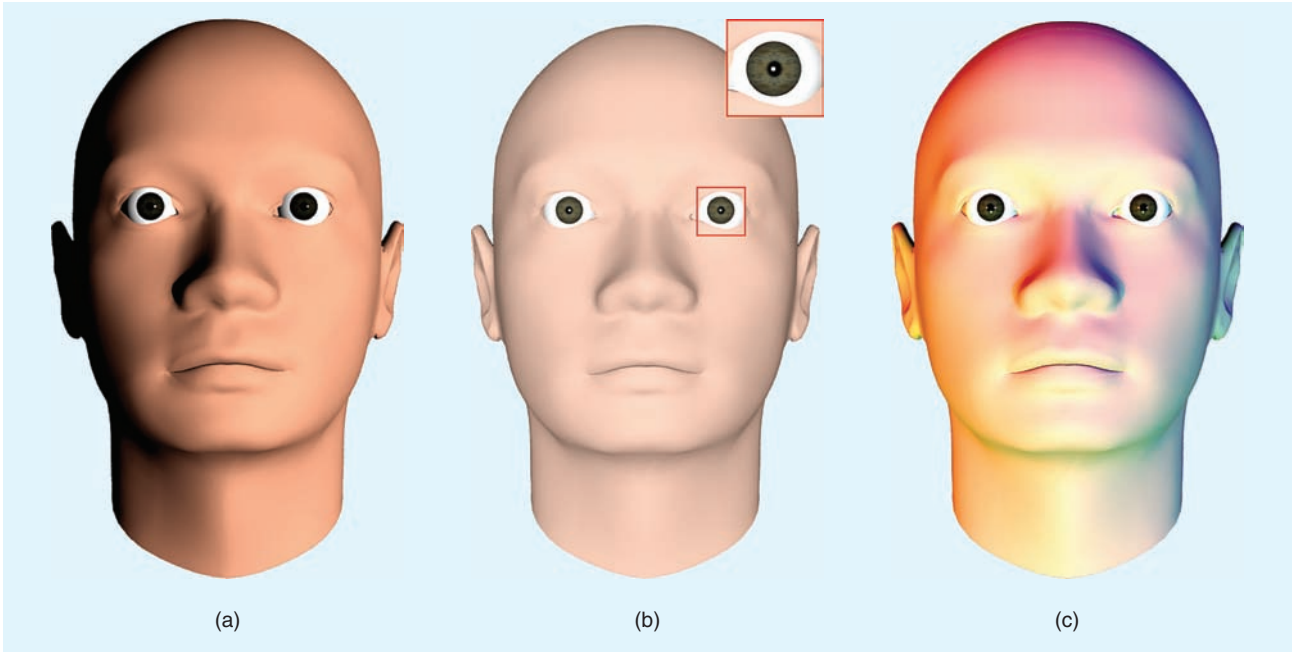
Os autores em [10] modelam o processamento em câmera com um modelo de ruído aditivo e multiplicativo. Os parâmetros do modelo de ruído são estimados a partir da câmera original ou de uma série de imagens originadas na câmera conhecida. As correlações entre o ruído estimado da câmera e o ruído das imagens extraídas são, então, usadas para verificar a autenticidade da imagem. O efeito do processamento em câmera é modelado da seguinte maneira:

$$I(x,y) = I_0(x,y) + \gamma I_0(x,y)K(x,y) + N(x,y), \qquad (6)$$

onde $I_0(\cdot)$ é a imagem sem ruídos, γ é a constante multiplicativa, $K(\cdot)$ é o ruído multiplicativo, chamado de ruído de não-uniformidade da resposta fotônica (*photo-response non-uniformity noise - PRNU*), e $N(\cdot)$ é um termo do ruído aditivo. Como o PRNU varia por toda a imagem, ele pode ser usado para detectar inconsistências locais ou globais da imagem. O PRNU é estimado a partir de uma série de imagens autênticas usando técnicas de remoção de ruídos baseadas nas transformadas wavelets. Como regra geral, pelo menos 50 imagens são necessárias para obter estimações precisas do PRNU. Menos imagens podem ser usadas se forem imagens de baixa frequência (por exemplo, imagens do céu). A autenticação é realizada por meio de uma correlação de blocos (*blockwise correlation*) entre o PRNU estimado e a imagem cuja autenticidade esteja sendo investigada. Comprovou-se também que o PRNU é único para um sensor específico [26]. Como tal, o PRNU também pode ser usado para a "balística" da câmera, identificando a câmera específica de onde uma imagem foi gerada.

ANÁLISE FÍSICA

Consideremos a criação de uma imagem falsa mostrando duas celebridades supostamente envolvidas romanticamente, andando numa praia ao por do sol. Essa imagem poderia ser criada por meio de uma composição, juntando imagens individuais de cada uma das celebridades. Sendo esse o caso, geralmente é difícil fazer corresponder a luminosidade sob a qual cada pessoa foi originalmente fotografada. Descreverei três técnicas para estimar as diferentes propriedades da iluminação existente quando uma pessoa ou objeto foi fotografado. Essas diferenças da iluminação numa imagem podem então ser usadas como prova de adulteração.



[FIG3] A direção de uma fonte de luz individual pode ser determinada a partir (a) do gradiente da luz pelo rosto e (b) a partir da posição da especularidade (ponto branco) do olho. Outros ambientes com iluminação mais complexa, com várias luzes coloridas (C) podem ser modelados como funções contínuas por pedaços na esfera.

DIREÇÃO DA LUZ (2-D)

Como o lado direito do rosto na Figura 3(a) está mais iluminado do que o lado esquerdo, podemos deduzir que a fonte de luz está posicionada à direita. Esta observação pode ser formalizada elaborando-se premissas para simplificação: a quantidade de luz que está atingindo a superfície é proporcional à normal à superfície e à direção da luz. Conhecendo-se as normais às superfícies tridimensionais (3-D), pode-se, portanto, determinar a direção da fonte de luz. Como as normais às superfícies 3-D não podem ser determinadas com base numa única imagem, os autores em [15] consideram somente as normais às superfícies bidimensionais (2-D) na borda de ocultamento do objeto. Em compensação, eles estimam dois dos três componentes da direção da fonte de luz. Apesar de haver ainda uma ambiguidade na direção estimada da luz, esses dois componentes são úteis no contexto forense.

Visando simplificar a estimação da direção da fonte de luz, supomos que a superfície de interesse seja lambertiana (a superfície reflete a luz isotropicamente), tenha um valor de reflectância constante e seja iluminada por uma fonte de luz pontual a uma distância infinita. Com base nestas premissas, a intensidade da imagem pode ser expressa como $I(x,y) = R(\vec{N}(x,y)\cdot\vec{L}) + A$, onde R é o valor de reflectância constante; $\vec{L}$ é um vetor tridimensional que aponta na direção da fonte de luz; $\vec{N}(x,y)$ é um vetor tridimensional que representa a normal à superfície no ponto (x,y), e A um termo constante de luz ambiente. Como somente a direção da fonte de luz é de interesse, podemos considerar que o termo de reflectância R, tem valor unitário. Na borda de ocultamento de uma superfície, o componente z da normal à superfície é zero. Além disso, os componentes x e y da normal à superfície podem ser estimados diretamente a partir da imagem. Com base nesta premissa adicional, a intensidade da imagem é expressa agora como:

$$I(x,y) = \vec{N}(x,y)\cdot\vec{L} + A \;=\; \begin{pmatrix} N_x(x,y) & N_y(x,y) \end{pmatrix}\begin{pmatrix} L_x \\ L_y \end{pmatrix} + A. \tag{7}$$

Podemos observar que, nessa fórmula, o componente z tanto da normal à superfície quanto da direção da luz é ignorado, pois $N_z(x,y) = 0$. Com pelo menos três pontos com a mesma reflectância R, e as normais à superfície $\vec{N}$, distintas, a direção da fonte de luz e o termo ambiente podem ser resolvidos usando-se a estimação padrão dos mínimos quadrados. Uma função de erro quadrático, incorporando o modelo de imagem de (7), é expresso por (8), conforme mostrado no rodapé da página.

Essa função de erro quadrático é minimizada usando-se a estimação padrão dos mínimos quadrados para obter $\vec{v} = (M^T M)^{-1} M^T \vec{b}$. Esse processo pode ser repetido para diferentes objetos e pessoas presentes na imagem para verificar se a iluminação é consistente.

$$E(\vec{L},A) \;=\; \left|\left| \begin{pmatrix} N_x(x_1,y_1) & N_y(x_1,y_1) & 1 \\ N_x(x_2,y_2) & N_y(x_2,y_2) & 1 \\ \vdots & \vdots & \vdots \\ N_x(x_p,y_p) & N_y(x_p,y_p) & 1 \end{pmatrix} \cdot \begin{pmatrix} L_x \\ L_y \\ A \end{pmatrix} - \begin{pmatrix} I(x_1,y_1) \\ I(x_2,y_2) \\ \vdots \\ I(x_p,y_p) \end{pmatrix} \right|\right|^2 \;=\; ||M\vec{v} - \vec{b}||^2. \tag{8}$$

DIREÇÃO DA LUZ (3-D)

A estimação da direção da fonte de luz na seção anterior foi limitada a 2-D, pois, geralmente, é difícil determinar normais à superfícies tridimensionais a partir de uma única imagem. Em [20], os autores descrevem como estimar a direção 3-D até uma fonte de luz a partir da reflexão da luz no olho humano [veja Figura 3(b)]. As normais à superfícies 3-D necessárias são determinadas usando um modelo 3-D do olho humano.

A Figura 4 mostra a geometria básica da imagem onde a reflexão da luz é visível no olho. Essa reflexão é chamada de realce especular. Neste diagrama, os três vetores $\vec{L}$, $\vec{N}$, e $\vec{R}$ correspondem à direção até a luz; à normal à superfície no ponto onde o realce é formado e à direção na qual o realce será visto. A lei da reflexão diz que um raio de luz reflete numa superfície num ângulo de reflexão θ_r igual ao ângulo de incidência θ_i, sendo esses ângulos medidos em relação à normal à superfície $\vec{N}$. Considerando-se vetores unitários, a direção do raio refletido $\vec{R}$ pode ser descrita em termos da direção da luz $\vec{L}$ e a normal à superfície $\vec{N}$:

$$\vec{R} = \vec{L} + 2(\cos(\theta_i)\vec{N} - \vec{L}) = 2\cos(\theta_i)\vec{N} - \vec{L}. \quad (9)$$

Considerando-se um refletor perfeito ($\vec{V} = \vec{R}$), a restrição acima resulta em:

$$\vec{L} = 2\cos(\theta_i)\vec{N} - \vec{V} = 2(\vec{V}^T\vec{N})\vec{N} - \vec{V}. \quad (10)$$

Portanto, a direção da luz $\vec{L}$ pode ser estimada a partir da normal à superfície $\vec{N}$ e a direção de visualização $\vec{V}$ no realce especular. Essa direção estimada da luz pode ser comparada entre as diversas pessoas presentes numa imagem ou a direção estimada da luz usando a técnica descrita na seção anterior.

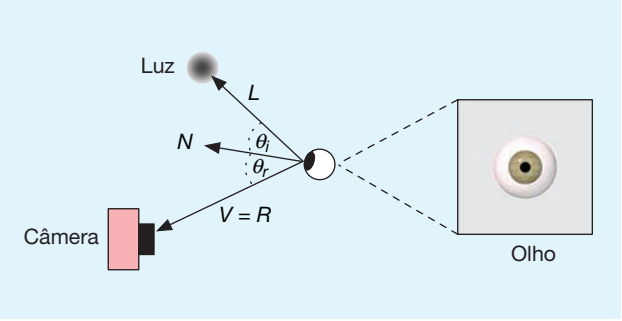
AMBIENTE DE LUZ

Nas duas seções anteriores, consideramos um modelo simplificado de iluminação que consiste em uma única fonte de luz dominante. No entanto, na prática, a iluminação de uma cena pode ser complexa: pode haver um grande número de fontes de luz em várias posições diferentes, criando diversos ambientes de luz [veja a Figura 3(c)]. Em [19], os autores descrevem como estimar uma representação de baixos parâmetros para esses ambientes de iluminação complexa.

Primeiramente, os autores aplicam a observação [39], que diz que a aparência de uma superfície lambertiana pode ser aproximada por:

$$E(\vec{N}) \approx \sum_{n=0}^{2} \sum_{m=-n}^{n} \hat{r}_n l_{n,m} Y_{n,m}(\vec{N}), \quad (11)$$

onde $E(\vec{N})$ é a quantidade de luz que atinge uma superfície (irradiância) com normal à superfície $\vec{N}$, $\hat{r}_n$ são constantes conhecidas, $Y_{n,m}(\cdot)$ são as funções harmônicas esféricas e $l_{n,m}$ são os pesos lineares desconhecidos dessas funções. As harmônicas esféricas formam uma base ortonormal de funções contínuas por partes na esfera e são análogas às transformadas de Fourier na reta ou no plano. Podemos observar que essa expressão é linear nos nove coeficientes do ambiente de iluminação; de $l_{0,0}$ a $l_{2,2}$ podendo, portanto, ser estimada usando-se o método dos mínimos quadrados. No entanto, essa solução requer normais à superfície 3-D em pelo menos nove pontos na superfície de um objeto. Quando não há diversas imagens nem geometria conhecida, fica difícil obter isso a partir de uma única imagem. A observação principal em [19] é que, considerando-se apenas a borda oculta de um objeto, (11) é simplificada para:



[FIG4] Formação do realce especular no olho (pequeno ponto branco na íris). A posição do realce é determinada pela normal à superfície $\vec{N}$ e as direções relativas até a fonte de luz $\vec{L}$ e o observador $\vec{V}$.

$$\begin{aligned} E(\vec{N}) = {} & l_{1,-1}\frac{2\pi}{3}Y_{1,-1}(\vec{N}) + l_{1,1}\frac{2\pi}{3}Y_{1,1}(\vec{N}) \\ & + l_{2,-2}\frac{\pi}{4}Y_{2,-2}(\vec{N}) + l_{2,2}\frac{\pi}{4}Y_{2,2}(\vec{N}) \\ & + l_{0,0}\frac{\pi}{2\sqrt{\pi}} - l_{2,0}\frac{\pi}{16}\sqrt{\frac{5}{\pi}}, \end{aligned} \quad (12)$$

onde, mais criticamente, as funções $Y_{i,j}(\cdot)$ dependem apenas dos componentes x e y da normal à superfície $\vec{N}$. Ou seja, os cinco coeficientes de iluminação podem ser estimados somente a partir de normais à superfícies 2-D. Além disso, a função (12) ainda é linear em seus (atuais) cinco coeficientes do ambiente de iluminação, que podem ser estimados usando o método dos mínimos quadrados. A adição de um termo de regularização melhora ainda mais a estabilidade da estimação. Os coeficientes estimados podem ser comparados para detectar inconsistências de luminosidade numa imagem.

TÉCNICAS BASEADAS NA GEOMETRIA

PONTO PRINCIPAL

Em imagens autênticas, o ponto principal (projeção do centro da câmera no plano da imagem) fica próximo ao centro da imagem. Quando uma pessoa ou objeto é movido na imagem, o ponto principal se move proporcionalmente. As diferenças do ponto principal estimado na imagem podem, portanto, ser usadas como prova de que houve adulteração. Em [18], os autores descrevem como estimar o ponto principal da câmera a partir de uma imagem de um par de olhos (ou seja, dois círculos) ou outras formas geométricas planas. Eles mostraram como uma translação no plano da imagem é equivalente a uma mudança do ponto principal. As inconsistências do ponto principal na imagem podem, então, ser usadas como prova de que houve adulteração.

[FIG5] (a) Imagem original. (b) Um *close-up* da placa do carro, praticamente ilegível. (c) O resultado da retificação planar seguida da equalização do histograma.

O limbo, junção entre a córnea e a esclera, é bem modelado por um círculo. Consideremos agora a projeção de um par de olhos (círculos), supondo-se que sejam coplanares. Nesse caso, a transformação das coordenadas do mundo para as coordenadas da imagem pode ser modelada com uma matriz de transformação projetiva planar H 3×3 H: $\vec{x} = H\vec{X}$, onde os pontos $\vec{X}$ do mundo e os pontos $\vec{x}$ da imagem são representados por vetores homogêneos bidimensionais. A transformação H pode ser estimada a partir da geometria conhecida dos olhos de uma pessoa e fatorada num produto de matrizes que incorpora os parâmetros intrínsecos e extrínsecos da câmera.

$$H = \lambda \begin{pmatrix} f & 0 & c_1 \\ 0 & f & c_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \vec{r}_1 & \vec{r}_2 & \vec{t} \end{pmatrix}, \qquad (13)$$

onde λ é o fator de escala; a matriz à esquerda é a matriz intrínseca (onde f o comprimento focal e (c_1, c_2) é o ponto principal) e a matriz mais à esquerda incorpora a transformação de corpo rígido (rotação/translação) entre as coordenadas do mundo e da câmera. Uma vez fatorada, a matriz traz a estimativa desejada do ponto principal. Quando, durante a criação de uma composição de imagens de duas ou mais pessoas, a posição de uma pessoa foi mudada na imagem original, os pontos principais estimados para cada pessoa serão inconsistentes, evidenciando uma adulteração.

MEDIÇÕES MÉTRICAS

A Figura 5 mostra a imagem de um carro com a placa praticamente ilegível. A Figura também mostra [Figura 5(c)] o resultado da transformação da placa, como se ela estivesse sendo vista de frente. Essa imagem retificada revela claramente o número da placa. Em [17], os autores analisam várias ferramentas de geometria projetiva que permitem a retificação de superfícies planares e, sob certas condições, a medição de superfícies planares do mundo real. São descritas três técnicas de retificação de superfícies planares transformadas em imagens sob projeção em perspectiva. Cada método requer somente uma imagem. O primeiro método explora o conhecimento sobre polígonos de formas conhecidas (por exemplo, sinais de rua, placas de carro, frases em outdoors). O segundo método requer conhecimento de dois ou mais pontos de fuga num plano e, por exemplo, um par de ângulos conhecidos do plano. O terceiro método requer dois ou mais círculos coplanares (por exemplo, as rodas de um carro). Em cada caso, é estimada a transformação do mundo para a imagem, permitindo a remoção das distorções planares e a realização de medições métricas no plano.

O FUTURO

As tecnologias modernas possibilitam a alteração e manipulação de vários tipos de mídias digitais usando técnicas impossíveis 20 anos atrás. As tecnologias do futuro, com certeza, nos permitirão manipular imagens digitais de uma forma difícil de imaginar hoje em dia. E com o aperfeiçoamento contínuo dessas tecnologias, torna-se cada vez mais importante para o campo de análise forense de mídias digitais acompanhar esses avanços.

Não há dúvida de que, com o desenvolvimento de novas tecnologias para a detecção de fraudes fotográficas, novas técnicas surgirão para criar falsificações mais difíceis de detectar. Enquanto é possível "enganar" algumas ferramentas de análise forense, o usuário médio terá à frente algumas ferramentas difíceis de ludibriar. Por exemplo, uma vez perturbada, a interpolação do conjunto de filtros de cores (*color filter array*) pode ser regenerada simplesmente colocando-se a imagem em sua matriz original e reinterpolando cada canal de cor. Por outro lado, a correção de uma iluminação inconsistente não é trivial num *software* padrão de edição de fotos. Como é o caso na guerra entre *spam* e *antispam* e entre vírus e antivírus, a batalha entre os falsificadores e os analistas forenses é, de certa forma, inevitável. O campo de análise forense de imagens, no entanto, tem conseguido e ainda vai conseguir tornar mais difícil e mais dispendioso criar falsificações que não possam ser detectadas.

AGRADECIMENTOS

Este trabalho foi realizado por meio de doação da *Adobe Systems* e da Microsoft; bolsa de pesquisa da *National Science Foundation* (CNS-0708209) e do *Institute of Security Technology Studies* do *Dartmouth College*, com subsídio do *Bureau of Justice Assistance* (2005-DD-BX-1091) e do Departamento Americano de *Homeland Security* (2006-CS-001-000001). Os pontos de vista e opiniões expressas neste documento são dos autores e não representam a posição ou política oficial do Departamento Americano de Justiça, do Departamento Americano de *Homeland Security* nem de qualquer outro patrocinador.

AUTHOR

Hany Farid (farid@cs.dartmouth.edu) obteve o título de Bacharel em Ciência da Computação e Matemática Aplicada da Universidade de Rochester em 1989. Em 1997, ele se formou Doutor em Ciência da Computação pela Universidade da Pensilvânia. Tendo trabalhando dois anos com ciências cognitivas e do cérebro no *Massachusetts Institute of Technology*, após o seu doutorado, em 1999, passou a fazer parte do corpo docente de Dartmouth. Hany é Professor Emérito William H. Neukom em Ciências da Computação e diretor do Neukom *Institute for Computational Science*. Ele recebeu o Prêmio NSF CAREER e bolsas de pesquisa da Sloan Foundation e da *Guggenheim Foundation*.

REFERENCES

[1] S. Bayram, I. Avcibas, B. Sankur, and N. Memon, "Image manipulation detection with binary similarity measures," in *Proc. European Signal Processing Conf.*, Turkey, 2005.

[2] S. Bayram, I. Avcibas, B. Sankur, and N. Memon, "Image manipulation detection," *J. Electron. Imaging*, vol. 15, no. 4, p. 41102, 2006.

[3] S. Bayram, H. T. Sencar, and N. Memon, "Source camera identification based on CFA interpolation," in *IEEE Int. Conf. Image Processing*, Genova, Italy, 2005, pp. 69–72.

[4] S. Bayram, H. T. Sencar, and N. Memon, "Improvements on source camera model identification based on CFA interpolation," in *Proc. IFIP WG 11.9 Int. Conf. Digital Forensics*, Orlando, FL, 2006, pp. 24–27

[5] I. J. Cox, M. L. Miller, and J. A. Bloom, Digital Watermarking. San Francisco, CA: Morgan Kaufmann, 2002.

[6] Z. Fan and R. L. de Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 230–235, 2003.

[7] H. Farid, "Detecting digital forgeries using bispectral analysis," *AI Lab, Massachusetts Institute of Technology*, Tech. Rep. AIM-1657, 1999.

[8] H. Farid, "Digital image ballistics from JPEG quantization," Dept. Comput. Sci., Dartmouth College, Tech. Rep. TR2006-583, 2006.

[9] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," in *Proc. IEEE Workshop on Statistical Analysis in Computer Vision (in conjunction with CVPR)*, Madison, WI, 2003.

[10] J. Fridrich, M. Chen, and M. Goljan, "Imaging sensor noise as digital x-ray for revealing forgeries," in *Proc. 9th Int. Workshop on Information Hiding,* Sant Malo, France, 2007, pp. 342–358.

[11] J. Fridrich, D. Soukal, and J. Lukás, "Detection of copy move forgery in digital images," in *Proc. Digital Forensic Research Workshop*, Aug. 2003.

[12] H. Gou, A. Swaminathan, and M. Wu, "Noise features for image tampering detection and steganalysis," in *Proc. IEEE Int. Conf. Image Processing*, San Antonio, TX, 2007, vol. 6, pp. 97–100.

[13] J. He, Z. Lin, L. Wang, and X. Tang, "Detecting doctored JPEG images via DCT coefficient analysis," in *Proc. European Conf. Computer Vision*, Graz, Austria, 2006, pp. 423–435.

[14] Y.-F. Hsu and S.-F. Chang, "Image splicing detection using camera response function consistency and automatic segmentation," in *Proc. Int. Conf. Multimedia and Expo*, Beijing, China, 2007.

[15] M. K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *Proc. ACM Multimedia and Security Workshop*, New York, NY, 2005, pp. 1–10.

[16] M. K. Johnson and H. Farid, "Exposing digital forgeries through chromatic aberration," in *Proc. ACM Multimedia and Security Workshop*, Geneva, Switzerland, 2006, pp. 48–55.

[17] M. K. Johnson and H. Farid, "Metric measurements on a plane from a single image," Dept. Comput. Sci., Dartmouth College, Tech. Rep. TR2006-579, 2006.

[18] M. K. Johnson and H. Farid, "Detecting photographic composites of people," in *Proc. 6th Int. Workshop on Digital Watermarking*, Guangzhou, China, 2007.

[19] M. K. Johnson and H. Farid, "Exposing digital forgeries in complex lighting environments," *IEEE Trans. Inform. Forensics Security*, vol. 3, no. 2, pp. 450–461, 2007.

[20] M. K. Johnson and H. Farid, "Exposing digital forgeries through specular highlights on the eye," in *Proc. 9th Int. Workshop on Information Hiding*, Saint Malo, France, 2007, pp. 311–325.

[21] S. Katzenbeisser and F. A. P. Petitcolas, Information Techniques for Steganography and Digital Watermarking. Norwood, MA: Artec House, 2000.

[22] M. Kirchner, "Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue," in *ACM Multimedia and Security Workshop*, 2008, pp. 11–20.

[23] G. Li, Q. Wu, D. Tu, and S. Sun, "A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD," in *IEEE Int. Conf. Multimedia and Expo*, Beijing, China, 2007, pp. 1750–1753.

[24] Z. Lin, R. Wang, X. Tang, and H-V Shum, "Detecting doctored images using camera response normality and consistency," in *Proc. Computer Vision and Pattern Recognition*, San Diego, CA, 2005.

[25] J. Lukas and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," in *Proc. Digital Forensic Research Workshop*, Cleveland, OH, Aug. 2003.

[26] J. Lukás, J. Fridrich, and M. Goljan, "Digital camera identification from sensor noise," *IEEE Trans. Inform. Forensics Security*, vol. 1, no. 2, pp. 205–214, 2006.

[27] W. Luo, J. Huang, and G. Qiu, "Robust detection of region-duplication forgery in digital images," in *Proc. Int. Conf. on Pattern Recognition*, Washington, D.C., 2006, pp. 746–749.

[28] W. Luo, Z. Qu, J. Huang, and G. Qiu, "A novel method for detecting cropped and recompressed image block," in *Proc. IEEE Conf. Acoustics, Speech and Signal Processing*, Honolulu, HI, 2007, pp. 217–220.

[29] S. Lyu and H. Farid, "How realistic is photorealistic?" *IEEE Trans. Signal Processing*, vol. 53, no. 2, pp. 845–850, 2005.

[30] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE Trans. Inform. Forensics Security*, vol. 1, no. 1, pp. 111–119, 2006.

[31] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation," *IEEE Trans. Inform. Forensics Security*, vol. 3, no. 3, pp. 529–538, 2008.

[32] T.-T. Ng and S.-F. Chang, "A model for image splicing," in *Proc. IEEE Int. Conf. Image Processing*, Singapore, 2004, vol. 2, pp. 1169–1172.

[33] P. Nillius and J.-O. Eklundh, "Automatic estimation of the projected light source direction," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2001, pp. 1076–1083.

[34] A.C. Popescu and H. Farid, "Exposing digital forgeries by detecting duplicated image regions," Dept. Comput. Sci., Dartmouth College, Tech. Rep. TR2004-515, 2004.

[35] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Proc. 6th Int. Workshop on Information Hiding*, Toronto, Canada, 2004, pp. 128–147.

[36] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of re-sampling," *IEEE Trans. Signal Processing*, vol. 53, no. 2, pp. 758–767, 2005.

[37] A. C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Trans. Signal Processing*, vol. 53, no. 10, pp. 3948–3959, 2005.

[38] S. Prasad and K. R. Ramakrishnan, "On resampling detection and its application to image tampering," in *Proc. IEEE Int. Conf. Multimedia and Exposition*, Toronto, Canada, 2006, pp. 1325–1328.

[39] R. Ramamoorthi and P. Hanrahan, "On the relationship between radiance and irradiance: Determining the illumination from images of a convex Lambertian object," *J. Opt. Soc. Amer. A*, vol. 18, no. 10, pp. 2448–2559, 2001.

[40] A. Swaminathan, M. Wu, and K. J. R. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Trans. Inform. Forensics Security*, vol. 3, no. 1, pp. 101–117, 2008.

[41] S. Ye, Q. Sun, and E. C. Chang, "Detecting digital image forgeries by measuring inconsistencies of blocking artifact," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Beijing, China, 2007, pp. 12–15.

[42] B. Mahdian and S. Saic, "Detection of copy move forgery using a method based on blur movement invariants," *Forensic Sci. Int.*, vol. 171, pp. 180–189, 2007.

[43] A.C. Gallagher, "Detection of linear and cubic interpolation in jpeg compressed images," in *Proc. 2nd Canadian Conf. Computer and Robot Vision.*, Victoria, British Columbia, Canada, vol. 171, 2005, pp. 65–72.

[44] H. Farid, "Digital ballistics from jpeg quantization: A followup study," Dept. Comp. Sci., Dartmouth College, Tech. Rep. TR2008-638, 2008. [SP]

Jessica Fridrich

Digital Image Forensics

Introdução aos métodos de estimação e detecção da impressão digital de sensores

© BRAND X PICTURES

Este tutorial descreve como a não-uniformidade da resposta fotônica (PRNU, na sigla em inglês) de sensores de imagens pode ser usada em várias tarefas importantes de perícia digital, tais como identificação de dispositivos, correlação com um dispositivo específico, recuperação do histórico do processamento e detecção de falsificações digitais. A PRNU, uma propriedade intrínseca de todos os sensores de imagens digitais, é causada por pequenas variações na capacidade de cada pixel de converter fótons em elétrons. Consequentemente, todos os sensores geram um padrão semelhante a um ruído em todas as imagens que produzem. Esse padrão, que representa a "impressão digital" do sensor, é, basicamente, uma marca d'água digital estocástica de espectro espalhado não intencional, que permanece após o processamento por compressão com perdas ou por filtragem. Este tutorial explica como estimar essa impressão digital a partir de imagens capturadas pela câmera e detectá-la posteriormente numa determinada imagem para estabelecer a origem e integridade da imagem. Várias tarefas forenses são formuladas como um problema de teste de hipóteses de dois canais, que é resolvido por meio do teste da razão de máxima verossimilhança. O desempenho dos métodos forenses apresentados é ilustrado por meio de exemplos, para uma melhor compreensão por parte dos leitores.

Existem dois tipos de sensores de imagens normalmente instalados em câmeras digitais, *camcorders e scanners*: o dispositivo de carga acoplada (*charge-coupled device* - CCD) e o semicondutor metal-óxido complementar (*complementary metal-oxide semiconductor* - CMOS). Ambos são compostos por vários detectores de fótons, também chamados de pixels. Os pixels são feitos de silício e capturam a luz por meio da conversão dos fótons em elétrons usando o efeito fotoelétrico. A carga acumulada é transferida para fora do sensor, amplificada e convertida em sinal digital num conversor A/D, sendo ainda processada antes do armazenamento dos dados em algum formato de imagem, como

Digital Object Identifier 10.1109/MSP.2008.931078

 1053-5888/09/$25.00©2009IEEE

JPEG (*Joint Photographic Experts Group*).

Os pixels são geralmente retangulares com vários mícrons de largura. A quantidade de elétrons gerados pela luz incidente sobre cada pixel depende das dimensões físicas da área fotossensível do pixel e da homogeneidade do silício. As dimensões físicas dos pixels variam levemente devido às imperfeições inseridas pelo processo de fabricação. Além disso, a falta de homogeneidade do silício causa variações da eficiência quântica (capacidade de converter fótons em elétrons) entre os pixels. As diferenças entre os pixels podem ser capturadas por meio de uma matriz **K** com as mesmas dimensões do sensor. Quando o sensor é iluminado com uma intensidade luminosa idealmente uniforme Y, na ausência de outras fontes de ruído, o sensor registra um sinal $Y+Y\mathbf{K}$ parecido com um ruído. O termo $Y\mathbf{K}$ é geralmente referido como PRNU.

QUANDO DETECTADA, A PRESENÇA DE UMA IMPRESSÃO DIGITAL NUMA IMAGEM NOS AJUDA A DETERMINAR SE A IMAGEM FOI PRODUZIDA POR UMA CÂMERA ESPECÍFICA.

A matriz **K** é responsável pela maior parte do que chamamos de impressão digital da câmera (*camera fingerprint*). A impressão digital pode ser estimada experimentalmente, tirando-se, por exemplo, várias fotos de uma superfície uniformemente iluminada e calculando a média das imagens para isolar o componente sistemático de todas as imagens. Ao mesmo tempo, o cálculo da média elimina componentes de ruídos randômicos, tais como ruído quântico (*shot noise*), que consiste em variações no número de fótons que atingem o pixel, variações essas resultantes das propriedades quânticas da luz; ruído de leitura, que é o ruído randômico inserido durante a leitura do sensor; e outros. Mais detalhes sobre outras fontes de ruídos que afetam a aquisição de imagens estão à disposição dos leitores em [1], [2]. A Figura 1 mostra uma seção ampliada de uma impressão digital de uma câmera Canon G2 de quatro megapixels, obtida por meio do cálculo da média de 120 imagens de 8 bits em tons de cinza com escala de cinza média de 128 ao longo de cada imagem. Os pontos brilhantes são pixels que normalmente geram mais elétrons, enquanto os pontos escuros são pixels cuja resposta é geralmente mais baixa. A variância nos valores dos pixels ao longo da imagem calculada (antes do ajuste da sua faixa para a visualização) foi de 0,5, ou 51dB. Apesar de a robustez da impressão digital depender, em grande parte, do modelo da câmera, a impressão digital do sensor é geralmente um sinal bem fraco.

A Figura 2 mostra a magnitude da transformada de Fourier de uma das linhas de pixels na imagem calculada. O sinal é parecido com um ruído branco com banda de atenuação de alta frequência.

Além da PRNU, a impressão digital da câmera contém todos os defeitos sistemáticos do sensor, inclusive os pixels supersaturados (*hot pixels*) e pixels com pouca saturação (*dead pixels*), sendo ambos pixels que produzem sinais altos ou baixos independentemente da iluminação; a chamada corrente escura (um tipo de ruído que a câmera teria se sua objetiva fosse coberta). Mas o componente mais importante da impressão digital é a PRNU. O termo $Y\mathbf{K}$ da PRNU quase não aparece nas áreas escuras onde $Y \approx 0$. Além disso, áreas totalmente saturadas da imagem, onde os pixels foram cheios até a sua capacidade máxima, produzindo um sinal constante, não carregam nenhum traço de PRNU nem outros tipos de ruídos.

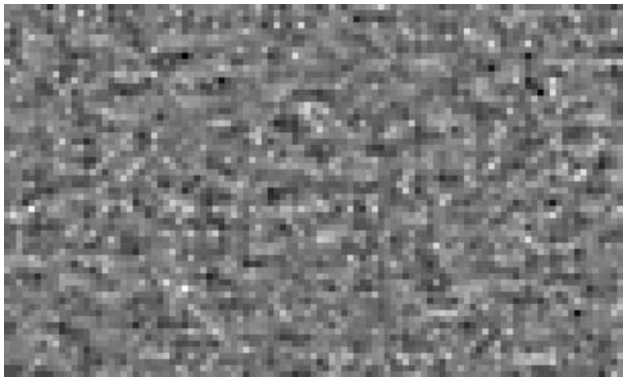
Podemos notar que, basicamente, todos os sensores de imagens (CCD, CMOS, JFET e CMOS-Foveon-X3) são construídos a partir de semicondutores e que as técnicas usadas para sua fabricação são parecidas. Portanto, esses sensores devem exibir "impressões digitais" com propriedades parecidas.

Apesar de ser estocástico por natureza, o termo da PRNU é relativamente estável durante a vida do sensor. Portanto, o fator **K** é uma medida forense bastante útil, sendo responsável por uma impressão digital única do sensor com as seguintes e importantes propriedades:

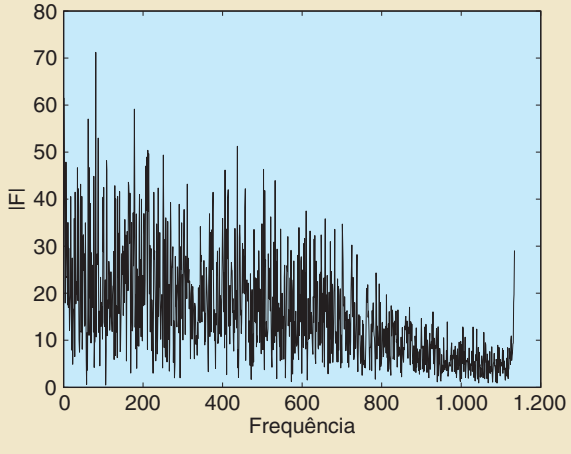
1) **Dimensionalidade:** A impressão digital é estocástica por natureza e possui uma grande quantidade de informações, fazendo com que seja única para cada sensor.
2) **Universalidade:** Todos os sensores de imagens exibem a PRNU.
3) **Generalidade:** A impressão digital está presente em todas as fotos, independentemente do sistema ótico da câmera, de suas configurações ou da cena, sendo as imagens totalmente escuras uma exceção.
4) **Estabilidade:** Ela é estável ao longo do tempo e sob uma ampla gama de condições ambientais (temperatura, umidade, etc.).
5) **Robustez:** Ela permanece mesmo quando há compressão com perdas, filtragem, correção gama e outros tipos de processamento.

A impressão digital pode ser usada em várias tarefas forenses:

- Com a detecção da presença de uma impressão digital específica na imagem, é possível identificar com segurança o dispositivo de aquisição (provando, por exemplo, que uma



[FIG1] Parte ampliada da impressão digital do sensor de uma Canon G2. A faixa dinâmica foi dimensionada para o intervalo [0, 255] para possibilitar a visualização.



[FIG2] Ampliação da transformada de Fourier de uma linha da impressão digital do sensor.

imagem foi capturada por uma câmera específica) ou provar que duas imagens foram produzidas pelo mesmo dispositivo (correlação com um dispositivo específico). A presença de uma impressão digital numa imagem também indica que a imagem investigada é natural e não resultado de uma renderização por computador.

- A verificação da ausência de uma impressão digital em regiões individuais da imagem, possibilita descobrir partes substituídas da imagem. Essa tarefa é parte da verificação da integridade.
- Por meio da detecção da robustez ou forma da impressão digital, é possível reconstruir parte do histórico de processamento. Por exemplo, podemos usar a impressão digital como um modelo para estimarmos formas de processamento geométrico, tais como redimensionamento, recorte e rotação. Operações não-geométricas também influenciam a robustez da impressão digital presente na imagem e podem, portando, ser detectadas.
- As características espectrais e espaciais da impressão digital podem ser usadas para identificar o modelo da câmera ou distinguir entre uma imagem escaneada e uma produzida por uma câmera digital (a imagem escaneada exibe anisotropia espacial).

ALÉM DE CÂMERAS DIGITAIS, ESSA TECNOLOGIA PODE SER USADA EM CAMCORDERS E SCANNERS.

Explicaremos neste tutorial os métodos usados para estimar a impressão digital e detectá-la em imagens. O material é baseado na teoria de estimação e detecção de sinais.

Este artigo está organizado da maneira descrita a seguir. Em primeiro lugar, descrevemos um modelo simplificado de saída de sinal, o qual usamos para derivar um estimador de máxima verossimilhança para a impressão digital. Ao mesmo tempo, enfatizamos a necessidade de se realizar um pré-processamento do sinal estimado para remover certos padrões sistemáticos, que poderiam gerar mais falsos positivos durante a identificação do dispositivo e mais falhas na detecção durante o uso da impressão digital para verificar a integridade da imagem. Usando o modelo do sensor novamente, a tarefa de detecção da PRNU é formulada como um problema de dois canais que é resolvido por meio do teste da razão de verossimilhança no conjunto de Neyman-Pearson. Primeiramente, derivamos o detector para identificar o dispositivo e, em seguida, adaptamos esse detector para verificar a correlação com um dispositivo específico (device linking) e a correspondência da impressão digital (fingerprint matching). A seguir, mostramos como verificar a integridade por meio da impressão digital, detectando-a em blocos individuais da imagem.

Neste artigo, fontes em negrito denotam vetores (ou matrizes) do comprimento especificado no texto; por exemplo, $\mathbf{X}$ e $\mathbf{Y}$ são vetores de comprimento n, e $\mathbf{X}[i]$ denota o i-ésimo componente de $\mathbf{X}$. Algumas vezes, indexamos os pixels de uma imagem usando um índice bidimensional formado pelo índice da linha e o da coluna. Salvo quando definido de outra maneira, todas as operações entre vetores e matrizes, tais como multiplicação, razão, potenciação, etc., são operações elemento a elemento. O produto escalar dos vetores é denotado como $\mathbf{X} \odot \mathbf{Y} = \sum_{i=1}^{n} \mathbf{X}[i]\mathbf{Y}[i]$ com $||\mathbf{X}|| = \sqrt{\mathbf{X} \odot \mathbf{X}}$ o módulo L_2 de $\mathbf{X}$. Denotando a média de amostra com uma barra, a correlação normalizada é:

$$\mathrm{corr}(\mathbf{X}, \mathbf{Y}) = \frac{(\mathbf{X} - \bar{\mathbf{X}}) \odot (\mathbf{Y} - \bar{\mathbf{Y}})}{||\mathbf{X} - \bar{\mathbf{X}}|| \cdot ||\mathbf{Y} - \bar{\mathbf{Y}}||}.$$

ESTIMAÇÃO DA IMPRESSÃO DIGITAL DO SENSOR

A PRNU é inserida na imagem durante a aquisição antes de o sinal ser quantizado ou processado de outra maneira.

MODELO DA SAÍDA DE DADOS DO SENSOR

Apesar de o processo de aquisição de imagens digitais ser bem complexo e variar bastante de câmera para câmera, alguns elementos básicos estão presentes na maioria das câmeras. A luz que vem do sistema ótico da câmera é projetada sobre a grade de pixels do sensor de imagens. A carga gerada pela interação dos fótons com o silício é amplificada e quantizada. Em seguida, o sinal de cada canal de cor é ajustado para ganho (*aumentado*) para atingir o balanceamento de branco adequado. Como a maioria dos sensores não consegue registrar cores, os pixels geralmente são equipados com um filtro de cores, que deixam somente a luz de uma cor específica (vermelho, verde ou azul) entrar no pixel. O conjunto de filtros é chamado de matriz de filtros de cores (*color filter array - CFA*). Para se obter uma imagem colorida, é feita a interpolação ou demosaico do sinal. Finalmente, é feita a correção da cor e a correção gama das cores para que sejam exibidas corretamente no monitor do computador. As câmeras podem empregar também técnicas de filtragem, como remoção de ruídos (*denoising*) e aumento da nitidez (*sharpening*). Ao final dessa cadeia de processamento, a imagem é armazenada no formato JPEG ou outro formato, às vezes por meio de quantização.

Vamos denotar para $\mathbf{I}[i]$ o sinal quantizado registrado no pixel i, $i = 1, \ldots, m \times n$, antes do demosaico. Aqui, $m \times n$ são dimensões da imagem. Seja $\mathbf{Y}[i]$ a intensidade da luz incidente no pixel i. Diminuímos os índices dos pixels para uma melhor leitura e usamos a seguinte forma vetorial do modelo de saída de sinal do sensor:

$$\mathbf{I} = g^{\gamma} \cdot [(\mathbf{1} + \mathbf{K})\mathbf{Y} + \Omega]^{\gamma} + \mathbf{Q}. \quad (1)$$

Lembramos aos leitores que todas as operações em (1) (e em todas as partes deste tutorial) são operações elemento a elemento. Em (1), g é o fator de ganho (diferente para cada canal de cor) e γ é o fator de correção gama (geralmente, $\gamma \approx 0{,}45$). A matriz $\mathbf{K}$ é um sinal parecido com um ruído de média zero, responsável pela PRNU (a impressão digital do sensor). A combinação das outras fontes de ruído, tais como corrente escura, ruído quântico e ruído de leitura (2), é denotada por Ω. $\mathbf{Q}$ é a distorção combinada causada pela quantização e/ou compressão JPEG.

Em algumas partes não escuras da imagem, o termo dominante entre chaves em (1) é a intensidade da luz $\mathbf{Y}$. Fatorando essa expressão e mantendo os dois primeiros termos na expansão em série de Taylor de $(1 + x)^{\gamma} = 1 + \gamma x + O(x^2)$ em $x = 0$, temos:

$$\begin{aligned} \mathbf{I} &= (g\mathbf{Y})^{\gamma} \cdot [\mathbf{1} + \mathbf{K} + \Omega/\mathbf{Y}]^{\gamma} + \mathbf{Q} \\ &\doteq (g\mathbf{Y})^{\gamma} \cdot (\mathbf{1} + \gamma\mathbf{K} + \gamma\Omega/\mathbf{Y}) + \mathbf{Q} = \mathbf{I}^{(0)} + \mathbf{I}^{(0)}\mathbf{K} + \Theta. \end{aligned} \quad (2)$$

Em (2), denotamos $\mathbf{I}^{(0)} = (g\mathbf{Y})^{\gamma}$ a saída ideal de sinal do sensor na ausência de ruídos e imperfeições. Note que $\mathbf{I}^{(0)}\mathbf{K}$ é o termo da PRNU e $\Theta = \gamma\mathbf{I}^{(0)}\Omega/\mathbf{Y} + \mathbf{Q}$ é o ruído de modelagem. Na última expressão em (2), o fator escalar γ foi absorvido pelo fator $\mathbf{K}$ da PRNU para simplificar a notação.

ESTIMAÇÃO DA IMPRESSÃO DIGITAL DO SENSOR

Nesta seção, o modelo acima de saída de dados do sensor é usado para derivar um estimador do fator $\mathbf{K}$ da PRNU. Os autores em [3] e [4] apresentam uma boa introdução à estimação e detecção de sinais.

A razão sinal-ruído (SNR, na sigla em inglês) entre o sinal de interesse $\mathbf{I}^{(0)}\mathbf{K}$ K e os dados observados $\mathbf{I}$ pode ser melhorada suprimindo-se a imagem sem ruído $\mathbf{I}^{(0)}$ por meio da subtração, dos dois lados de (2), uma versão de $\mathbf{I}$ sem ruído, $\hat{\mathbf{I}}^{(0)} = F(\mathbf{I})$, obtida com o uso de um filtro de remoção de ruídos F (consulte, no apêndice, a descrição do filtro usado nos experimentos exibidos neste tutorial).

A AUSÊNCIA DA IMPRESSÃO DIGITAL NUMA PEQUENA REGIÃO DA IMAGEM É EVIDÊNCIA DE MANIPULAÇÃO.

$$\begin{aligned}\mathbf{W} &= \mathbf{I} - \hat{\mathbf{I}}^{(0)} = \mathbf{IK} + \mathbf{I}^{(0)} - \hat{\mathbf{I}}^{(0)} + (\mathbf{I}^{(0)} - \mathbf{I})\mathbf{K} + \Theta \\ &= \mathbf{IK} + \Xi\,. \end{aligned} \tag{3}$$

É mais fácil estimar o termo da PRNU a partir de $\mathbf{W}$ do que de $\mathbf{I}$, pois o filtro suprime o conteúdo da imagem. Denotamos por Ξ a soma de Θ e dois termos adicionais introduzidos pelo filtro de remoção de ruído.

Vamos supor que tenhamos um banco de dados com $d \geq 1$ imagens, $\mathbf{I}_1, \ldots, \mathbf{I}_d$, capturadas pela câmera. Para cada pixel i, a sequência $\Xi_1[i], \ldots, \Xi_d[i]$ é modelada como um ruído branco gaussiano (WGN, na sigla em inglês) com variância σ^2. O termo do ruído não é tecnicamente independente do sinal $\mathbf{IK}$ da PRNU devido ao termo $(\mathbf{I}^{(0)} - \mathbf{I})\mathbf{K}$. No entanto, como a energia desse termo é pequena, comparada à de $\mathbf{IK}$, é razoável supor que Ξ seja independente de $\mathbf{IK}$.

Considerando-se (3), podemos escrever para cada $k = 1, \ldots, d$

$$\frac{\mathbf{W}_k}{\mathbf{I}_k} = \mathbf{K} + \frac{\Xi_k}{\mathbf{I}_k}, \quad \mathbf{W}_k = \mathbf{I}_k - \hat{\mathbf{I}}_k^{(0)}, \hat{\mathbf{I}}_k^{(0)} = F(\mathbf{I}_k)\,. \tag{4}$$

Sob a nossa premissa acerca do termo do ruído, a log-verossimilhança de $\mathbf{W}_k/\mathbf{I}_k$ dado $\mathbf{K}$, é

$$L(\mathbf{K}) = -\frac{d}{2}\sum_{k=1}^{d}\log(2\pi\sigma^2/(\mathbf{I}_k)^2) - \sum_{k=1}^{d}\frac{(\mathbf{W}_k/\mathbf{I}_k - \mathbf{K})^2}{2\sigma^2/(\mathbf{I}_k)^2}. \tag{5}$$

Tomando-se derivadas parciais de (5) para os elementos individuais de $\mathbf{K}$ e resolvendo para $\mathbf{K}$, obtemos a estimativa de máxima verossimilhança $\hat{\mathbf{K}}$

$$\frac{\partial L(\mathbf{K})}{\partial \mathbf{K}} = \sum_{k=1}^{d}\frac{\mathbf{W}_k/\mathbf{I}_k - \mathbf{K}}{\sigma^2/(\mathbf{I}_k)^2} = 0 \Rightarrow \hat{\mathbf{K}} = \frac{\sum_{k=1}^{d}\mathbf{W}_k\mathbf{I}_k}{\sum_{k=1}^{d}(\mathbf{I}_k)^2}. \tag{6}$$

O limite inferior de Cramér-Rao (CRLB, na sigla em inglês) nos dá o limite em $\hat{\mathbf{K}}$

$$\frac{\partial^2 L(\mathbf{K})}{\partial \mathbf{K}^2} = -\frac{\sum_{k=1}^{d}(\mathbf{I}_k)^2}{\sigma^2} \Rightarrow \mathrm{var}(\hat{\mathbf{K}}) \geq \frac{1}{-E\left[\frac{\partial^2 L(\mathbf{K})}{\partial \mathbf{K}^2}\right]} = \frac{\sigma^2}{\sum_{k=1}^{d}(\mathbf{I}_k)^2}. \tag{7}$$

Como o modelo de sensor (3) é linear, o CRLB indica que o estimador de máxima verossimilhança é não viezado e com variância mínima, e sua variância $\mathrm{var}(\hat{\mathbf{K}}) \sim 1/d$. vemos que as melhores imagens para a estimação de $\mathbf{K}$ são aquelas com alta luminância (mas não saturadas) e σ^2 pequeno (o que significa conteúdo suave). Se a câmera investigada estivesse em nosso poder, imagens sem foco de um céu nublado e claro seriam ideais. Na prática, podemos obter boas estimativas da impressão digital a partir de 20 a 50 imagens naturais, dependendo da câmera. Quando usamos imagens do céu em vez de imagens naturais, metade das imagens (aproximadamente) seria suficiente para obtermos uma estimativa com a mesma precisão.

O fator $\hat{\mathbf{K}}$ estimado contém todos os componentes sistematicamente presentes em todas as imagens, inclusive os artefatos inseridos com a interpolação de cores, compressão JPEG, transferência de sinais no sensor [5] e modelo do sensor. Enquanto existe uma PRNU única para cada sensor, os outros artefatos são compartilhados por várias câmeras do mesmo modelo ou estilo de sensor. Consequentemente, os fatores estimados da PRNU de duas câmeras diferentes podem estar levemente correlacionados, o que aumenta, de forma indesejável, a taxa de falsos positivos. Felizmente, os artefatos se manifestam, principalmente, como sinais periódicos em médias de linhas e colunas de $\hat{\mathbf{K}}$ podendo ser eliminados por meio da simples subtração das médias de cada linha ou coluna. Para um $\hat{\mathbf{K}}$ estimado da PRNU com m linhas e n colunas, o processamento é descrito por meio do seguinte pseudocódigo.

$r_i = 1/n\sum_{j=1}^{n}\hat{\mathbf{K}}[i,j]$
por $i = 1$ a m $\{\hat{\mathbf{K}}'[i,j] = \hat{\mathbf{K}}[i,j] - r_i$ para $j = 1, \ldots, n\}$
$c_j = 1/m\sum_{i=1}^{m}\hat{\mathbf{K}}'[i,j]$
por $j = 1$ a n $\{\hat{\mathbf{K}}''[i,j] = \hat{\mathbf{K}}'[i,j] - c_j$ para $i = 1, \ldots, m$.

A diferença $\hat{\mathbf{K}} - \hat{\mathbf{K}}''$ é conhecida como o padrão linear (veja a Figura 3), sendo, por si mesma, uma entidade forense útil, podendo ser usada para verificar a correspondência entre a impressão digital de uma câmera e o modelo ou marca da câmera. Como este tutorial não cobre esse tópico, recomendamos ao leitor consultar [6] para mais detalhes.

Para que o texto não fique pesado devido à grande quantidade de símbolos, no resto deste tutorial denotaremos a impressão digital processada $\hat{\mathbf{K}}''$ com o mesmo símbolo $\hat{\mathbf{K}}$.

Encerramos esta seção falando sobre imagens coloridas. Nesse caso, o fator da PRNU pode ser estimado para cada canal de cor separadamente, obtendo-se, portanto, três impressões digitais das mesmas dimensões $\hat{\mathbf{K}}_R$, $\hat{\mathbf{K}}_G$, e $\hat{\mathbf{K}}_B$. Como essas três impressões digitais são altamente correlacionadas devido ao processamento feito na câmera, em todos os métodos forenses descritos neste tutorial, antes de analisar a imagem colorida que está sendo investigada, precisamos convertê-la para tons de cinza e combinar as três impressões digitais para

formar uma única impressão digital usando a conversão normal de RGB para tons de cinza.

$$\hat{\mathbf{K}} = 0.3\hat{\mathbf{K}}_R + 0.6\hat{\mathbf{K}}_G + 0.1\hat{\mathbf{K}}_B. \tag{8}$$

IDENTIFICAÇÃO DA CÂMERA POR MEIO DA IMPRESSÃO DIGITAL DO SENSOR

Nesta seção, introduzimos a metodologia geral usada para determinar a origem de imagens ou vídeos por meio da impressão digital. Iniciamos com o que consideramos a situação mais frequente na prática - a identificação da câmera a partir de imagens. Nossa tarefa é descobrir se a imagem investigada foi capturada por uma determinada câmera. Para isso, verificamos se o residual de ruído da imagem contém a impressão digital da câmera. Em preparação às duas próximas tarefas forenses, intimamente relacionadas, formulamos o problema de teste da hipótese para identificação da câmera num cenário que seja geral o suficiente para cobrir as tarefas restantes, que são a correlação com um dispositivo específico (*device linking*) e a correspondência da impressão digital (*fingerprint matching*). Para a correlação com um dispositivo específico, duas imagens são testadas para determinar se vieram da mesma câmera (a câmera, em si, pode não ser conhecida). Para verificar a correspondência das duas impressões digitais estimadas determinamos a correspondência entre dois clipes de vídeo, pois os quadros individuais de cada clipe podem ser usados como uma sequência de imagens, da qual pode se obter uma estimativa da impressão digital da filmadora (nesse caso também, pode ser que a câmera digital ou filmadora não seja conhecida).

IDENTIFICAÇÃO DO DISPOSITIVO

Vamos considerar um cenário em que a imagem investigada tenha passado por uma transformação geométrica, como redimensionamento (scaling) ou rotação. Vamos supor que, antes da transformação geométrica, a imagem estava em tons de cinza, representada por uma matriz $m \times n$ com $\mathbf{I}[i, j]$, $i = 1, \ldots, m, j = 1, \ldots, n$. Denotemos agora como $\mathbf{u}$ o vetor (desconhecido) de parâmetros que descrevem a transformação geométrica $T_\mathbf{u}$. Por exemplo, $\mathbf{u}$ pode ser uma razão de redimensionamento ou um vetor bidimensional composto por um parâmetro de redimensionamento e um ângulo de rotação desconhecido. Para a identificação do dispositivo, desejamos determinar se a imagem transformada:

$$\mathbf{Z} = T_\mathbf{u}(\mathbf{I})$$

foi capturada por uma câmera com uma estimativa $\hat{\mathbf{K}}$ da impressão digital conhecida. Vamos supor que a transformação geométrica seja redutora (com redução da taxa de amostragem, por exemplo), sendo, portanto, mais vantajoso verificar a correspondência da transformada inversa $T_\mathbf{u}^{-1}(\mathbf{Z})$ com a impressão digital em vez da correspondência de $\mathbf{Z}$ com uma versão reduzida de $\hat{\mathbf{K}}$.

Formulamos agora o problema de detecção de uma maneira um pouco mais geral para cobrir todas as três tarefas forenses mencionadas acima dentro de uma única estrutura. A detecção da impressão digital é o seguinte problema de teste da hipótese em dois canais:

$$\begin{aligned} \mathrm{H}_0&: \mathbf{K}_1 \neq \mathbf{K}_2 \\ \mathrm{H}_1&: \mathbf{K}_1 = \mathbf{K}_2 \end{aligned}, \tag{9}$$

onde

$$\begin{aligned} \mathbf{W}_1 &= \mathbf{I}_1\mathbf{K}_1 + \Xi_1 \\ T_\mathbf{u}^{-1}(\mathbf{W}_2) &= T_\mathbf{u}^{-1}(\mathbf{Z})\mathbf{K}_2 + \Xi_2. \end{aligned} \tag{10}$$

Em (10), todos os sinais são observados, exceto os termos Ξ_1 e Ξ_2 do ruído e as impressões digitais $\mathbf{K}_1$ e $\mathbf{K}_2$. Especificamente para o problema de identificação do dispositivo, $\mathbf{I}_1 \equiv 1$, $\mathbf{W}_1 = \hat{\mathbf{K}}$ estimado na seção anterior e Ξ_1 é o erro de estimação da PRNU. $\mathbf{K}_2$ **é a PRNU da câmera que capturou a imagem,** $\mathbf{W}_2$ é o residual de ruído transformado geometricamente e Ξ_2 é o termo do ruído. Em geral, $\mathbf{u}$ é um parâmetro desconhecido. Notemos que, como $T_\mathbf{u}^{-1}(\mathbf{W}_2)$ e $\mathbf{W}_1$ pode ter dimensões diferentes, a fórmula (10) envolve um deslocamento espacial (spatial shift) desconhecido entre os dois sinais, $\boldsymbol{s}$.

Modelando os termos Ξ_1 e Ξ_2 do ruído como ruído branco gaussiano com variâncias σ_1^2 e σ_2^2, conhecidas, o teste da razão de verossimilhança generalizada para esse problema de dois canais foi derivado em (7). Os dados estatísticos do teste são uma soma de três termos, duas quantidades do tipo energia e um termo de correlação cruzada:

A complexidade envolvida na análise dessas três expressões é proporcional ao quadrado do número de pixels $(m \times n)^2$, o que torna impraticável o uso desse detector. Portanto, simplificamos o detector, obtendo uma correlação cruzada normalizada (NCC, na sigla em inglês), que pode ser avaliada usando-se a transformada de Fourier. Sob H_1, o máximo em (11), mostrada na parte inferior da página, deve-se principalmente à contribuição do termo de correlação cruzada, $C(\mathbf{u}, \mathbf{s})$, que exibe um forte pico dos valores apropriados da transformação geométrica. Portanto, um detector sub-ótimo muito mais rápido é a NCC entre $\mathbf{X}$ e $\mathbf{Y}$ maximizados sobre todos os deslocamentos s_1, s_2, e $\mathbf{u}$

$$\begin{aligned} t &= \max_{\mathbf{u},\mathbf{s}} \{E_1(\mathbf{u}, \mathbf{s}) + E_2(\mathbf{u}, \mathbf{s}) + C(\mathbf{u}, \mathbf{s})\}, \\ E_1(\mathbf{u}, \mathbf{s}) &= \sum_{i,j} \frac{\mathbf{I}_1^2[i,j](\mathbf{W}_1[i+s_1, j+s_2])^2}{\sigma_1^2\mathbf{I}_1^2[i,j] + \sigma_1^4\sigma_2^{-2}(T_\mathbf{u}^{-1}(\mathbf{Z})[i+s_1, j+s_2])^2} \\ E_2(\mathbf{u}, \mathbf{s}) &= \sum_{i,j} \frac{(T_\mathbf{u}^{-1}(\mathbf{Z})[i+s_1, j+s_2])^2(T_\mathbf{u}^{-1}(\mathbf{W}_2)[i+s_1, j+s_2])^2}{\sigma_2^2(T_\mathbf{u}^{-1}(\mathbf{Z})[i+s_1, j+s_2])^2 + \sigma_2^4\sigma_1^{-2}\mathbf{I}_1^2[i,j]} \\ C(\mathbf{u}, \mathbf{s}) &= \sum_{i,j} \frac{\mathbf{I}_1[i,j]\mathbf{W}_1[i,j](T_\mathbf{u}^{-1}(\mathbf{Z})[i+s_1, j+s_2])(T_\mathbf{u}^{-1}(\mathbf{W}_2)[i+s_1, j+s_2])}{\sigma_2^2\mathbf{I}_1^2[i,j] + \sigma_1^2(T_\mathbf{u}^{-1}(\mathbf{Z})[i+s_1, j+s_2])^2} \end{aligned} \tag{11}$$

$$\mathrm{NCC}[s_1, s_2; \mathbf{u}] = \frac{\sum_{k=1}^{m}\sum_{l=1}^{n}(\mathbf{X}[k,l]-\bar{\mathbf{X}})(\mathbf{Y}[k+s_1, l+s_2]-\bar{\mathbf{Y}})}{\|\mathbf{X}-\bar{\mathbf{X}}\|\|\mathbf{Y}-\bar{\mathbf{Y}}\|}, \tag{12}$$

que visualizamos como uma matriz $m \times n$ parametrizada por $\mathbf{u}$, onde

$$\mathbf{X} = \frac{\mathbf{I}_1\mathbf{W}_1}{\sqrt{\sigma_2^2\mathbf{I}_1^2 + \sigma_1^2(T_{\mathbf{u}}^{-1}(\mathbf{Z}))^2}}, \mathbf{Y} = \frac{T_{\mathbf{u}}^{-1}(\mathbf{Z})T_{\mathbf{u}}^{-1}(\mathbf{W}_2)}{\sqrt{\sigma_2^2\mathbf{I}_1^2 + \sigma_1^2(T_{\mathbf{u}}^{-1}(\mathbf{Z}))^2}}. \tag{13}$$

Uma estatística de detecção mais estável, cujo significado ficará aparente na análise de erros descrita posteriormente nesta seção e a qual recomendamos que seja usada para todas as tarefas de identificação de câmeras, é a métrica de estimação da correlação de pico (PCE, na sigla em inglês), definida como

$$\mathrm{PCE}(\mathbf{u}) = \frac{\mathrm{NCC}[\mathbf{s}_{\mathrm{peak}}; \mathbf{u}]^2}{\frac{1}{mn - |\mathcal{N}|}\sum_{\mathbf{s}, \mathbf{s}\notin\mathcal{N}}\mathrm{NCC}[\mathbf{s}; \mathbf{u}]^2}, \tag{14}$$

onde para cada $\mathbf{u}$ fixo, $\mathcal{N}$ é a pequena região ao redor do valor de pico da NCC $\mathbf{s}_{\mathrm{peak}}$ ao longo dos deslocamentos s_1, s_2.

Para a identificação do dispositivo a partir de uma única imagem, o ruído de estimação Ξ_1 da impressão digital é muito mais fraco do que Ξ_2 do residual de ruído da imagem investigada. Portanto, $\sigma_1^2 = \mathrm{var}(\Xi_1) \ll \mathrm{var}(\Xi_2) = \sigma_2^2$ e (12) podem ser mais simplificadas para se obter uma NCC entre

$$\mathbf{X} = \mathbf{W}_1 = \hat{\mathbf{K}} \text{ and } \mathbf{Y} = T_{\mathbf{u}}^{-1}(\mathbf{Z})T_{\mathbf{u}}^{-1}(\mathbf{W}_2).$$

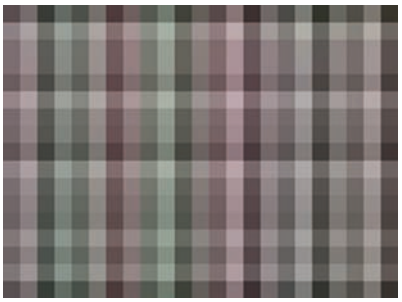
Lembre-se de que $\mathbf{I}_1 = 1$ para a identificação de um dispositivo quando a impressão digital é conhecida.

Na prática, o valor máximo da PCE pode ser encontrado por meio de uma busca numa grade obtida com a discretização da faixa de $\mathbf{u}$. Infelizmente, como as estatísticas são do tipo ruído para valores incorretos de $\mathbf{u}$ e só exibem um forte pico numa pequena vizinhança do valor correto de $\mathbf{u}$, não podemos usar métodos de gradientes, sendo forçados a aplicar uma busca de grade bem onerosa. A grade tem de ser suficientemente densa para que não deixemos de perceber o pico. Como exemplo, explicamos mais detalhadamente como realizar a busca quando $\mathbf{u} = r$ for uma razão de redimensionamento desconhecida. Para mais detalhes, recomendamos ao leitor consultar [9].

Supondo-se que a imagem investigada tenha as dimensões $M \times N$, buscamos o parâmetro de redimensionamento com valores discretos $r_i \le 1$, $i = 0, 1, \ldots, R$, de $r_0 = 1$ (sem redimensionamento, só recorte [*cropping*]), até $r_{\min} = \max\{M/m, N/n\} < 1$

$$r_i = \frac{1}{1 + 0.005i}, \; i = 0, 1, 2, \ldots. \tag{15}$$

Para um parâmetro de redimensionamento fixo r_i, a correlação cruzada (12) não precisa ser computada para todos os deslocamentos $\mathbf{s}$ mas somente para aqueles que movem a imagem com taxa de amostragem aumentada $T_{r_i}^{-1}(\mathbf{Z})$ *dentro* das dimensões de $\hat{\mathbf{K}}$ pois tais deslocamentos podem gerados por meio de recorte (*cropping*). Supondo-se que as dimensões da imagem com taxa de amostragem aumentada $T_{r_i}^{-1}(\mathbf{Z})$ sejam $M/r_i \times N/r_i$, temos a seguinte faixa para o deslocamento espacial $\mathbf{s} = (s_1, s_2)$:



[FIG3] **Detalhe do padrão linear da Canon S40**

$$0 \le s_1 \le m - M/r_i \text{ e } 0 \le s_2 \le n - N/r_i. \tag{16}$$

O pico das duas NCC bidimensionais ao longo dos deslocamentos espaciais $\mathbf{s}$ é avaliado para cada r_i usando-se a PCE(r_i) (14). Se $\max_i \mathrm{PCE}(r_i) > \tau$, decidimos que ocorreu H_1 (a câmera corresponde à imagem). Além disso, o valor do parâmetro de redimensionamento no qual a PCE atinge esse máximo determina a razão de redimensionamento r_{peak}. A localização do pico $\mathbf{s}_{\mathrm{peak}}$ na correlação cruzada normalizada determina os parâmetros de recorte (*cropping*). Portanto, como um subproduto desse algoritmo, podemos determinar o histórico de processamento da imagem investigada (veja a Figura 4). A impressão digital pode, assim, ser encarada como um modelo (*template*) semelhante aos modelos usados na marcação digital (*digital watermarking*). Ela também pode ser usada para engenharia reversa do processamento feito na câmera, inclusive zoom digital [9].

Em qualquer aplicação forense, é importante manter baixa a quantidade de falsos positivos. Em tarefas de identificação de câmeras, isso significa que a probabilidade P_{FA} de uma câmera que não capturou a imagem ser erroneamente identificada como a câmera que o fez deve ser mantida abaixo de um limiar (threshold) definido pelo usuário (método de Neyman-Pearson). Portanto, precisamos obter a relação entre P_{FA} e o limiar na PCE. Devemos notar que o limiar dependerá do tamanho do espaço de busca, que, por sua vez, é determinado pelas dimensões da imagem investigada.

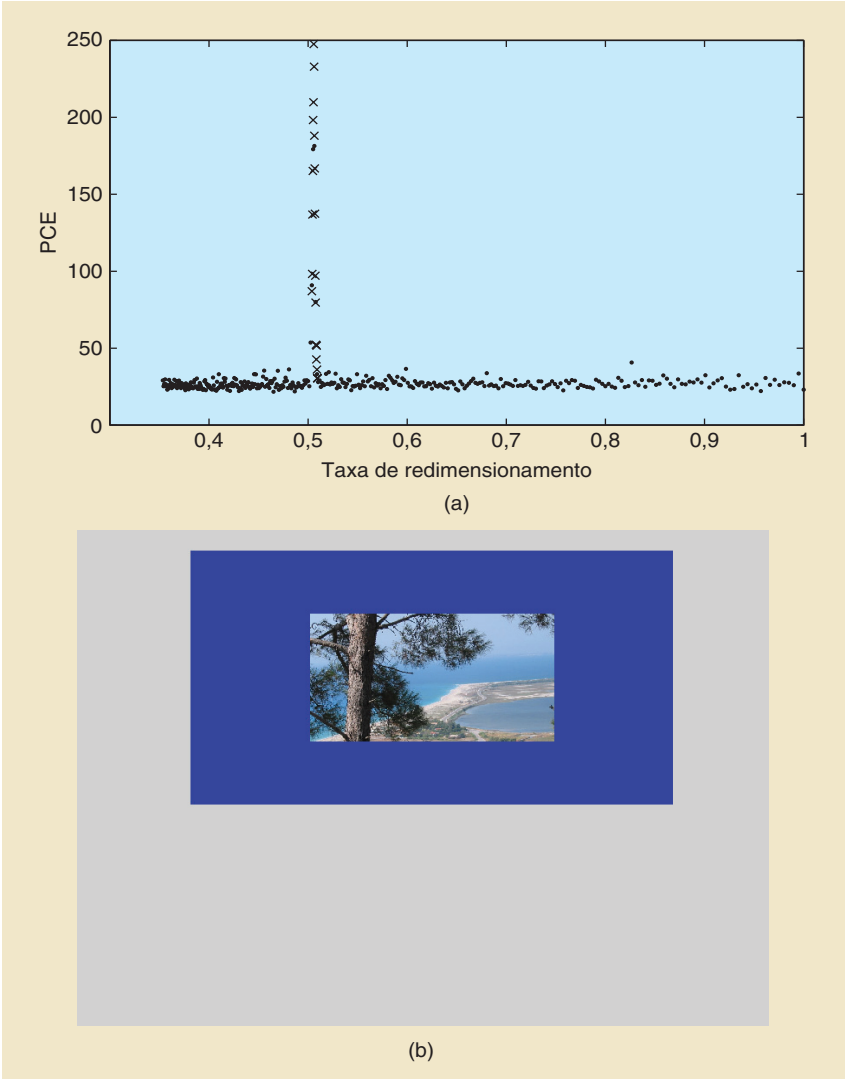
Sob a hipótese H_0 para uma razão de redimensionamento fixa, os valores da correlação cruzada normalizada NCC[s; r_i] como função de $\mathbf{s}$ são bem modelados [9] como um ruído branco gaussiano $\zeta^{(i)} \sim N(0, \sigma_i^2)$ com variância que pode depender de i. Estimando-se a variância do modelo gaussiano usando-se a variância de amostragem $\hat{\sigma}_i^2$ da NCC[s; r_i] sobre $\mathbf{s}$ após excluir uma pequena região central $\mathcal{N}$ ao redor do pico

$$\hat{\sigma}_i^2 = \frac{1}{mn - |\mathcal{N}|}\sum_{\mathbf{s}, \mathbf{s}\notin\mathcal{N}}\mathrm{NCC}[\mathbf{s}; r_i]^2, \tag{17}$$

calculamos agora a probabilidade p_i de que $\zeta^{(i)}$ por acaso atinja o valor de pico NCC[$\mathbf{s}_{\mathrm{peak}}$; r_{peak}] ou maior:

$$p_i = \int_{\mathrm{NCC}[\mathbf{s}_{\mathrm{peak}}; r_{\mathrm{peak}}]}^{\infty} \frac{1}{\sqrt{2\pi}\hat{\sigma}_i} e^{\frac{x^2}{2\hat{\sigma}_i^2}} dx = \int_{\hat{\sigma}_{\mathrm{peak}}\sqrt{\mathrm{PCE}_{\mathrm{peak}}}}^{\infty} \frac{1}{\sqrt{2\pi}\hat{\sigma}_i} e^{\frac{x}{2\hat{\sigma}_i^2}} dx$$
$$= Q\left(\frac{\hat{\sigma}_{\mathrm{peak}}}{\hat{\sigma}_i}\sqrt{\mathrm{PCE}_{\mathrm{peak}}}\right),$$

onde $Q(x) = 1 - \Phi(x)$ com $\Phi(x)$ denotando a função de distri-



[FIG4] (a) Pico detectado na PCE(r_i). (b) Representação visual dos parâmetros de recorte e redimensionamento r_{peak}, s_{peak}. O quadro cinza exibe o tamanho original da imagem e o quadro azul exibe o tamanho da imagem após o recorte e antes do redimensionamento.

buição cumulativa de uma variável normal padrão $N(0, 1)$ e $\text{PCE}_{\text{peak}} = \text{PCE}(r_{\text{peak}})$. Conforme explicado acima, durante a busca do vetor de corte s, precisamos somente buscar na faixa (16), o que significa que estamos tomando amostras máximas $k_i = (m - M/r_i + 1) \times (n - N/r_i + 1)$ de $\zeta^{(i)}$. Portanto, a probabilidade de o valor máximo de $\zeta^{(i)}$ não exceder $\text{NCC}[\mathbf{s}_{\text{peak}}; r_{\text{peak}}]$ é $(1 - p_i)^{k_i}$. Após R passos da busca, a probabilidade de haver um falso positivo é

$$P_{\text{FA}} = 1 - \prod_{i=1}^{R} (1 - p_i)^{k_i}. \tag{18}$$

Já que podemos interromper a busca após a PCE atingir um certo limiar, temos $r_i \le r_{\text{peak}}$. Como $\hat{\sigma}_i^2$ é não-decrescente em i, $\hat{\sigma}_{\text{peak}}/\hat{\sigma}_i \ge 1$. Como $Q(x)$ é decrescente, temos $p_i \le Q(\sqrt{\text{PCE}_{\text{peak}}}) = p$. Portanto, como $k_i \le mn$, obtemos o limiar superior em P_{FA}

$$P_{\text{FA}} \le 1 - (1 - p)^{k_{\max}}, \tag{19}$$

onde $k_{\max} = \sum_{i=0}^{R-1} k_i$ é o número máximo de valores dos parâmetros r e $\mathbf{s}$ sobre os quais o máximo de (11) pode ser tomado. A Equação (19), juntamente com $p = Q(\sqrt{\tau})$, determina o limiar da PCE PCE, $\tau = \tau(P_{\text{FA}}, M, N, m, n)$.

Isso encerra a formulação e solução técnica do algoritmo de identificação da câmera a partir de uma única imagem quando se conhece a impressão digital da câmera. Para dar ao leitor uma idéia de quão confiável esse algoritmo é, incluímos, mais adiante, alguns experimentos com imagens reais. Esse algoritmo pode ser usado com pequenas modificações para as outras tarefas forenses formuladas no início desta seção, ou seja, a correlação com um dispositivo específico (*device linking*) e correspondência da impressão digital (*fingerprint matching*).

CORRELAÇÃO COM UM DISPOSITIVO ESPECÍFICO

O detector derivado na seção anterior pode ser prontamente usado, com algumas modificações, para se fazer a correlação com um dispositivo específico ou determinar se duas imagens, $\mathbf{I}_1$ e $\mathbf{Z}$, foram capturadas pela mesma câmera [11]. Pode ser que, nesse caso, nem a câmera nem a impressão digital seja conhecida.

O problema da correlação com um dispositivo específico corresponde exatamente à formulação de dois canais (9) e (10) com o detector GLRT (11). Sua versão mais rápida e sub-ótima é a PCE (14) obtida a partir do valor máximo da $\text{NCC}[s_1, s_2; \mathbf{u}]$ sobre todos os $s_1, s_2; \mathbf{u}$ [veja (12) e (13)]. Ao contrário do problema de identificação da câmera, agora, a força de ambos os termos Ξ_1 e Ξ_2, 2 é comparável e precisa ser estimada a partir de observações. Felizmente, como o termo IK da PRNU é muito mais fraco do que o ruído de modelagem Ξ, as estimativas razoáveis das variâncias de ruído são simplesmente $\hat{\sigma}_1^2 = \text{var}(\mathbf{W}_1)$, $\hat{\sigma}_2^2 = \text{var}(\mathbf{W}_2)$.

Além disso, ao contrário do problema de identificação da câmera, a busca por um redimensionamento desconhecido deve, agora, ser estendida para dimensões $r_i > 1$ (aumento da taxa de amostragem), pois o efeito combinado do recorte e redimensionamento, desconhecido em ambas as imagens, impede que possamos determinar qual das imagens foi reduzida em relação à outra. A análise de erro continua da seção anterior.

Devido à limitação de espaço, não incluímos a verificação experimental do algoritmo de correlação com um dispositivo específico. Ao invés disso, recomendamos aos leitores consultar [11].

CORRESPONDÊNCIA DE IMPRESSÕES DIGITAIS

O último caso, correspondência de impressões digitais (fingerprint matching), refere-se a situações nas quais precisamos decidir se duas estimativas de duas impressões digitais potencialmente diferentes são idênticas. Isso ocorre, por exemplo, na correlação de clipes de vídeo,

pois a impressão digital pode ser estimada a partir de todos os quadros que formam o clipe [12].

O detector derivado na seção de identificação do dispositivo também pode ser usado nesse caso. Ele pode ser ainda mais simplificado, pois, para encontrarmos a correspondência entre impressões digitais, temos $\mathbf{I}_1 = \mathbf{Z} = 1$ sendo que (12) se torna simplesmente a correlação cruzada normalizada entre $\mathbf{X} = \hat{\mathbf{K}}_1$ e $\mathbf{Y} = T_{\mathrm{u}}^{-1}(\hat{\mathbf{K}}_2)$.

Para a verificação experimental do algoritmo de correspondência entre impressões digitais em clipes de vídeo, recomendamos aos leitores consultar [12].

ESTE TUTORIAL APRESENTA VÁRIOS MÉTODOS DE ANÁLISE FORENSE DE DOCUMENTOS DIGITAIS BASEADOS NO FATO DE QUE CADA SENSOR DE IMAGENS PRODUZ UMA IMPRESSÃO DIGITAL ÚNICA, NA FORMA DE RUÍDO, EM TODAS AS IMAGENS QUE GERA.

DETECCÃO DE FRAUDE USANDO A IMPRESSÃO DIGITAL DA CÂMERA

Um objetivo diferente, mas não menos importante, do uso da impressão digital do sensor é verificar a integridade de imagens. Alguns tipos de adulterações podem ser identificados por meio da detecção da presença da impressão digital em áreas menores. A premissa é que se a região foi copiada de outra parte da imagem (ou de uma imagem diferente), ela não conterá a impressão digital correta. Pode ser que algumas mudanças de conteúdo da imagem preservem a PRNU, não sendo detectadas por meio desta abordagem. Um bom exemplo seria a mudança de cor de uma mancha para que se pareça com uma mancha de sangue.

O algoritmo de detecção de adulteração testa a presença da impressão digital em cada $B \times B$ bloco deslizante (sliding block) separadamente, juntando, em seguida, todas as decisões locais. Para manter a análise simples, vamos supor que a imagem investigada não tenha passado por nenhum tipo de processamento geométrico. Para cada bloco $\mathcal{B}_b$, o problema de detecção é formulado como um teste de hipótese

$$\begin{aligned} \mathrm{H}_0&: \mathbf{W}_b = \Xi_b \\ \mathrm{H}_1&: \mathbf{W}_b = \mathbf{I}_b\hat{\mathbf{K}}_b + \Xi_b\,. \end{aligned} \tag{20}$$

Aqui, $\mathbf{W}_b$ é o residual de ruído do bloco; $\hat{\mathbf{K}}_b$ é o bloco correspondente da impressão digital; $\mathbf{I}_b$ é a intensidade do bloco e Ξ_b é o ruído de modelagem, o qual supomos ser um ruído branco gaussiano com variância desconhecida $\boldsymbol{\sigma}_{\Xi}^2$. O teste de máxima verossimilhança é a correlação normalizada

$$\rho_b = \mathrm{corr}(\mathbf{I}_b\hat{\mathbf{K}}_b, \mathbf{W}_b). \tag{21}$$

Na detecção de fraudes, talvez queiramos controlar os dois tipos de erros: a falha na identificação do bloco adulterado e a identificação errônea de uma região como adulterada. Para isso, precisamos estimar a distribuição dos dados estatísticos do teste sob as duas hipóteses.

A densidade da probabilidade sob H_0, $p(x|\mathrm{H}_0)$, pode ser determinada por meio da correlação do sinal conhecido $\mathbf{I}_b\hat{\mathbf{K}}_b$ com os residuais de ruído das outras câmeras. A distribuição de ρ_b sob H_1, $p(x|\mathrm{H}_1)$, é muito mais difícil de se obter, pois há uma forte influência do conteúdo do bloco sobre ela. Blocos escuros terão valores menores de correlação devido ao caráter multiplicativo da PRNU. A impressão digital também pode estar ausente em áreas planas devido a uma forte compressão JPEG ou saturação. Finalmente, as áreas texturizadas terão valores de correlação menores devido ao ruído de modelagem mais forte. O problema pode ser resolvido com a construção de um preditor de correlação que nos diga qual seria o valor da estatística de teste ρ_b e sua distribuição se o bloco b não tivesse sido adulterado e se tivesse realmente sido originado pela câmera.

O preditor é um mapeamento que deve ser construído para cada câmera. O mapeamento atribui uma estimativa da correlação ρ_b para cada trio $(i_b,\ f_b,\ t_b)$, onde os elementos individuais do trio significam uma medição da intensidade, saturação e textura do bloco b. O mapeamento pode ser construído, por exemplo, usando-se técnicas de regressão ou de aprendizado de máquina, treinando-as num banco de dados de blocos de imagens a partir de imagens capturadas pela câmera. O tamanho do bloco não pode ser muito pequeno, pois, nesse caso, a correlação ρ_b teria uma variância muito grande. Por outro lado, blocos grandes comprometeriam o funcionamento do algoritmo de detecção de adulteração. Blocos de 64×64 ou 128×128 pixels funcionam bem para a maioria das câmeras.

Uma medida razoável da intensidade é a intensidade média no bloco

$$i_b = \frac{1}{|\mathcal{B}_b|}\sum_{i\in\mathcal{B}_b}\mathbf{I}[i]\,. \tag{22}$$

Tomamos, como medida de planicidade (*flatness*) o número relativo de pixels i no bloco cujo desvio padrão de intensidade da amostra $\sigma_1[i]$ da vizinhança local 3×3 de i esteja abaixo de um certo limiar

$$f_b = \frac{1}{|\mathcal{B}_b|}|\{i \in \mathcal{B}_b\,|\sigma_1[i] < c\mathbf{I}[i]\}|\,, \tag{23}$$

onde $c \approx 0.03$ (para a câmera Canon G2). Os melhores valores de c variam de acordo com o modelo da câmera.

Uma boa medida da textura deve, de alguma maneira, avaliar a quantidade de arestas no bloco. Entre várias opções disponíveis, damos o seguinte exemplo:

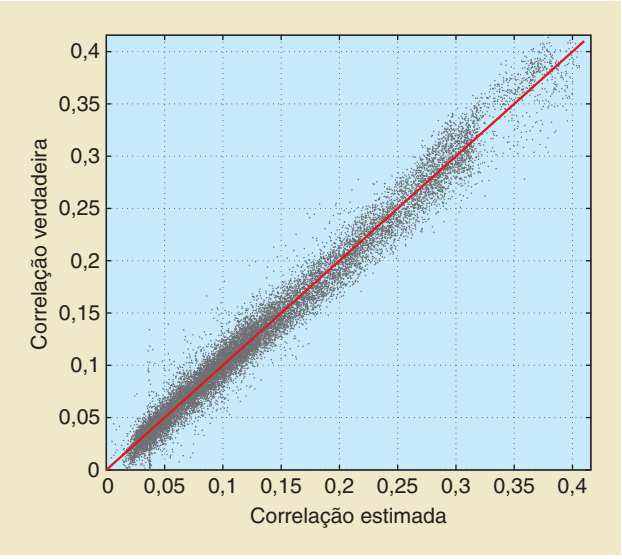
$$t_b = \frac{1}{|\mathcal{B}_b|}\sum_{i\in\mathcal{B}_b}\frac{1}{1+\mathrm{var}_5(\mathbf{F}[i])}, \tag{24}$$

onde $\mathrm{var}_5(\mathbf{F}[i])$ são a variância das amostras calculada a partir de uma vizinhança 5×5 de pixels i para obter uma versão do bloco filtrado com um filtro passa-altas, $\mathbf{F}[i]$, uma variância, por exemplo, obtida por meio do uso de um mapa de arestas ou um residual de ruído no domínio de uma transformada.

Como podemos obter centenas de blocos a partir de uma única imagem, só precisamos de uma pequena quantidade de imagens para treinar (construir) o preditor. Os dados usados para a construção podem também ser usados para estimar a distribuição dos erros do preditor ν_b

$$\rho_b = \hat{\rho}_b + \nu_b\,, \tag{25}$$

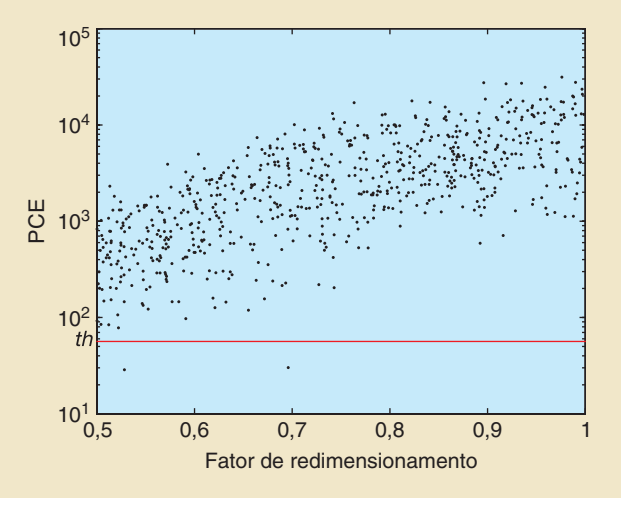
onde $\hat{\rho}_b$ é o valor estimado da correlação.



[FIG5] Diagrama disperso da correlação entre ρ_b e $\hat{\rho}_b$ para 30.000 blocos de128×128 pixels de 300 imagens TIFF produzidas por uma Cannon G2.

A Figura 5 mostra o desempenho do preditor construído por meio de uma regressão polinomial de segunda ordem para uma câmera Canon G2. Digamos que, para um dado bloco investigado, apliquemos o preditor e obtenhamos o valor estimado $\hat{\rho}_b$. A distribuição $p(x|\mathrm{H}_1)$ é obtida aplicando-se uma função de densidade da probabilidade paramétrica (*pdf*, na sigla em inglês) a todos os pontos da Figura 7, cuja correlação estimada esteja numa pequena vizinhança de $\hat{\rho}_b$, $(\hat{\rho}_b - \varepsilon, \hat{\rho}_b + \varepsilon)$. Um modelo suficientemente flexível de *pdf*, que permite caudas finas e grossas é o modelo gaussiano generalizado com *pdf* $\alpha/(2\sigma\Gamma(1/\alpha))e^{-(|x-\mu|/\sigma)^{\alpha}}$ com variância $\sigma^2\Gamma(3/\alpha)/\Gamma(1/\alpha)$, média μ, e parâmetro de formato α.

Continuamos agora com a descrição do algoritmo de detecção de adulteração usando a impressão digital do sensor. O algoritmo realiza o deslizamento de um bloco ao longo da imagem e a avaliação da estatística de teste ρ_b para cada bloco b.



[FIG6] $\mathrm{PCE}_{\mathrm{peak}}$ como uma função da razão de redimensionamento para 720 imagens correlacionadas à câmera. O limite de detecção, sublinhado com uma linha horizontal, corresponde a $P_{\mathrm{FA}} = 10^{-5}$.

O limiar de decisão t para a estatística de teste ρ_b foi estabelecido para obtermos a probabilidade de identificação errônea de um bloco adulterado como não-adulterado $\Pr(\rho_b > t|\,\mathrm{H}_0) = 0.01$.

O bloco b é marcado como possivelmente adulterado se $\rho_b < t$, mas essa decisão é atribuída somente ao pixel central i do bloco. Por meio desse processo, para uma imagem de $m \times n$ obtemos um conjunto binário $(m–B+1) \times (n–B+1)$ com $\mathbf{Z}[i] = \rho_b < t$ indicando os pixels possivelmente adulterados com $\mathbf{Z}[i] = 1$.

O critério de Neyman-Pearson acima decide "adulterado" sempre que $\rho_b < t$ mesmo quando ρ_b pb puder ser "mais compatível" com $p(x|\mathrm{H}_1)$, o que é mais provável de acontecer quando ρ_b for pequeno, como em blocos altamente texturizados. Com o fim de controlar a quantidade de pixels erroneamente apontados como adulterados, calculamos, para cada pixel i, a probabilidade de identificação errônea de um pixel como adulterado.

$$p[i] = \int_{-\infty}^{t} p(x|\mathrm{H}_1)dx\,. \tag{26}$$

O pixel i é marcado como não adulterado (estabelecemos que $\mathbf{Z}[i] = 0$) se $p[i] > \beta$, onde β ? é um limiar definido pelo usuário (nos experimentos mostrados neste tutorial, $\beta = 0.01$). O mapa binário $\mathbf{Z}$ resultante identifica as regiões adulteradas em sua forma sem processamento (*raw*). O mapa final $\mathbf{Z}$ é obtido com o pós-processamento de $\mathbf{Z}$.

O tamanho do bloco impõe um limite mais baixo no tamanho das regiões adulteradas que o algoritmo consegue identificar. Removemos de $\mathbf{Z}$ todas as regiões adulteradas simplesmente conectadas que contenham menos do que 64×64 pixels. O mapa final das regiões adulteradas é obtido dilatando-se $\mathbf{Z}$ com um kernel quadrado de 20×20. O objetivo desse passo é compensar pelo fato de que a decisão sobre o bloco inteiro é atribuída somente ao seu pixel central, fazendo com que, talvez, deixemos de identificar partes da região adulterada.

VERIFICAÇÃO EXPERIMENTAL

Nesta seção, demonstramos como os métodos forenses propostos nas duas seções anteriores podem ser implantados na prática. Incluímos também alguns exemplos de resultados experimentais para dar aos leitores uma idéia de como esses métodos funcionam com imagens reais. Recomendamos aos leitores que consultem [9] e [13] para obterem testes mais completos e [11] e [12] para uma verificação experimental da correlação com o dispositivo específico e da correspondência de impressões digitais em clipes de vídeo. A identificação de câmeras para imagens impressas é discutida em [10].

IDENTIFICAÇÃO DA CÂMERA

Uma câmera Canon G2 com um sensor CCD de quatro megapixels foi usada em todos os experimentos desta seção. A impressão digital da câmera foi estimada para cada canal de cor separadamente usando-se um estimador de máxima verossimilhança (6) de 30 imagens de céu azul capturadas no formato TIFF. As impressões digitais estimadas foram pré-processadas usando o procedimento de coluna e linha com média zero, explicado acima na seção de estimação da impressão digital do sensor, para remover padrões residuais não específicos do sensor. Esse passo é muito importante, pois esses artefatos podem causar uma interferência indesejada em certos fatores de redimensionamen-

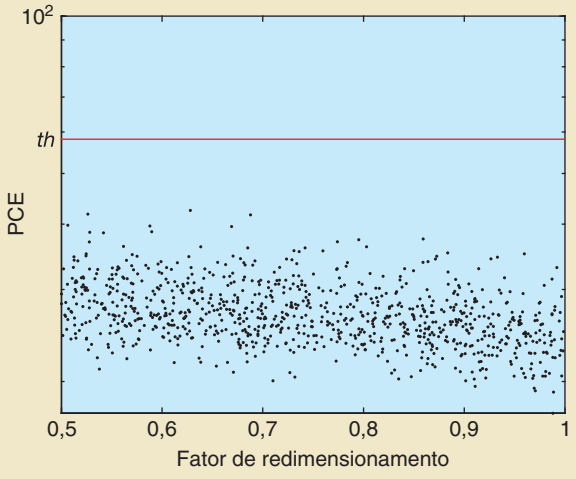
to e deslocamentos espaciais, reduzindo, portanto, a PCE e aumentando substancialmente a taxa de falsos positivos.

As impressões digitais estimadas dos três canais de cor foram combinadas numa única impressão digital usando a a regra de conversão linear usada para converter imagens coloridas para tons de cinza.

$$\hat{\mathbf{K}} = 0.3\hat{\mathbf{K}}_R + 0.6\hat{\mathbf{K}}_G + 0.1\hat{\mathbf{K}}_B.$$

Todas as outras imagens investigadas neste teste também foram convertidas para tons de cinza antes do emprego dos detectores descritos na subseção sobre identificação do dispositivo em "Identificação da Câmera por meio da Impressão Digital do Sensor".

A câmera foi usada posteriormente para capturar 720 imagens compostas por fotos e várias cenas em interiores e exteriores sob condições de iluminação de amplo espectro e configurações de zoom por um período de 4 anos. Todas as imagens foram tiradas com resolução total CCD com configuração de JPEG de alta qualidade. Cada imagem foi primeiramente recortada aleatoriamente até 50% em cada dimensão. O canto superior esquerdo da região recortada também foi escolhido aleatoriamente, com distribuição uniforme dentro do quarto superior esquerdo da imagem. A parte recortada foi reamostrada (*downsampled*) numa razão de redimensionamento $r \in [0{,}5,\ 1]$ escolhida aleatoriamente. Finalmente, as imagens foram convertidas a tons de cinza e processadas usando-se compressão JPEG com 85% de qualidade.



[FIG7] PCE_{peak} para 915 imagens não correlacionadas à câmera. O limiar de detecção T é novamente sublinhado com uma linha horizontal e corresponde a $P_{FA} = 10^{-5}$.

O limite de detecção T foi escolhido para obtermos a probabilidade de ocorrência de falsos positivos $P_{FA} = 10^{-5}$. O algoritmo de identificação da câmera foi aplicado com $r_{min} = 0.5$ em todas as imagens. Em somente dois casos não houve detecção (Figura 6).

Na figura, a PCE é mostrada como uma função da razão de redimensionamento escolhida aleatoriamente. Não houve

APÊNDICE: FILTRO DE REMOÇÃO DE RUÍDO

O filtro de remoção de ruído (*denoising filter*) usado nas seções experimentais deste tutorial é construído no domínio *wavelet*. Ele foi originalmente descrito em [22].

Vamos supor que a imagem seja uma imagem de 512 × 512 em tons de cinza. Imagens maiores podem ser processadas em blocos; imagens coloridas têm os ruídos removidos para cada canal de cor separadamente. Os coeficientes wavelet de alta frequência da imagem com ruído são modelados como uma mistura aditiva de sinais i.i.d. (independentes, identicamente distribuídos) com média zero (imagem sem ruído) e um ruído branco gaussiano estacionário $N(0, \sigma_0^2)$) (o componente do ruído). O filtro de remoção de ruído é construído em dois estágios. No primeiro estágio, estimamos a variância da imagem local e, no segundo estágio, o filtro Wiener local é usado para obtermos uma estimativa da imagem sem ruído no domínio *wavelet*. Descreveremos agora os passos a serem seguidos:

1) Calcule a decomposição wavelet em quatro níveis da imagem com ruído com os filtros ortogonais de quadratura espelhados de Daubechies. Descrevemos o procedimento para um nível fixo (executado para bandas de alta frequência em todos os quatro níveis). Denote a sub-banda vertical, horizontal e diagonal como $h[i, j]$, $v[i, j]$, $d[i, j]$, onde (i, j) passa por um conjunto de índices $\mathcal{J}$ que depende do nível de composição.

2) Em cada sub-banda, estime a variância local da imagem original sem ruído para cada coeficiente wavelet usando a estimação MAP para quatro tamanhos de uma vizinhança $W \times W$ quadrada $\mathcal{N}$, para $W \in \{3, 5, 7, 9\}$.

$$\hat{\boldsymbol{\sigma}}_W^2[i, j] = \max\left(0, \frac{1}{W^2}\sum_{(i,j)\in\mathcal{N}} h^2[i, j] - \sigma_0^2\right),\ (i, j) \in \mathcal{J}.$$

Tome a menor das quatro variâncias como estimativa final.

$$\hat{\boldsymbol{\sigma}}^2(i, j) = \min(\boldsymbol{\sigma}_3^2[i, j], \boldsymbol{\sigma}_5^2[i, j], \boldsymbol{\sigma}_7^2[i, j], \boldsymbol{\sigma}_9^2[i, j]),\ (i, j) \in \mathcal{J}.$$

3) Os coeficientes *wavelet* sem ruído são obtidos usando-se o filtro de Wiener

$$h_{den}[i, j] = h[i, j]\frac{\hat{\boldsymbol{\sigma}}^2[i, j]}{\hat{\boldsymbol{\sigma}}^2[i, j] + \sigma_0^2}$$

e da mesma maneira, para $v[i, j]$, e $d[i, j]$, $(i, j) \in \mathcal{J}$.

4) Repita os passos 1-3 para cada nível e cada canal de cor. A imagem sem ruído é obtida aplicando-se a transformada *wavelet* inversa aos coeficientes wavelet sem ruído.

Em todos os experimentos, usamos $\sigma_0 = 2$ [para uma faixa dinâmica de imagens de (0-255)] para sermos conservadores e para garantir que o filtro extraia uma parte considerável do ruído PRNU mesmo para câmeras com muito ruído.

detecção em duas imagens altamente texturizadas. Em todas as detecções bem sucedidas, os parâmetros de recorte e redimensionamento foram detectados com precisão acima de dois pixels em cada dimensão.

Para testar a taxa de identificações errôneas, usamos 915 imagens de mais de 100 câmeras diferentes, baixadas da Internet com resolução nativa. As imagens foram recortadas para quatro megapixels (o tamanho das imagens da Canon G2) e sujeitas ao mesmo recorte aleatório, redimensionamento e compressão JPEG, tal como as 720 imagens discutidas acima. O limiar do algoritmo de identificação da câmera foi estabelecido no mesmo valor que no experimento anterior. Todas as imagens foram corretamente classificadas como não originadas pela câmera testada (Figura 7). Para verificar experimentalmente a taxa teórica de falsos positivos, milhões de imagens teriam de ser examinadas, algo, infelizmente, impossível.

A IMPRESSÃO DIGITAL PODE SER ESTIMADA A PARTIR DE IMAGENS CAPTURADAS PELA CÂMERA, CALCULANDO-SE A MÉDIA DOS COMPONENTES DE SEU RUÍDO. O SINAL DOMINANTE NA IMPRESSÃO DIGITAL É A NÃO-UNIFORMIDADE DA RESPOSTA FOTÔNICA (PRNU) CAUSADA PELA SENSIBILIDADE VARIÁVEL DOS PIXELS À LUZ.

DETECÇÃO DE ADULTERAÇÃO

A Figura 8(a) mostra a imagem original sem processamento, tirada por uma câmera digital Olympus C765 equipada com um sensor CCD de quatro megapixels. Usando o Photoshop, a garota do meio foi coberta por pedaços da casa que está ao fundo [Figura 8(b)]. Em seguida, a imagem adulterada foi armazenada nos formatos TIFF e JPEG (com compressão JPEG com 75% de qualidade). O resultado correspondente do algoritmo de detecção de fraudes, mostrado nas Figuras 8(c) e (d) é o mapa binário **Z**, realçado com uma grade quadrada. As duas últimas figuras mostram o mapa Z após a adulteração ter sido submetida à remoção de ruído usando um filtro Wiener 3×3 [Figura 8(e)], seguido de compressão JPEG com 90% de qualidade; quando a imagem forjada foi processada usando correção gama com $\gamma = 0{,}5$ e novamente salva usando compressão JPEG com 90% de qualidade [Figura 8(f)]. Em todos os casos, a região forjada foi corretamente detectada.

Mais exemplos de casos de detecção de adulteração usando esse algoritmo, inclusive resultados de testes feitos com número maior de adulterações automaticamente criadas, assim como imagens não adulteradas, podem ser encontrados na publicação original [13].

CONCLUSÕES

Este tutorial descreve vários métodos forenses digitais baseados no fato de que cada sensor de imagem deixa uma impressão digital na forma de ruído em todas as fotos que tira. O principal componente da impressão digital é a PRNU, que é decorrente da capacidade variável dos pixels de converter a luz em elétrons. Com as diferenças entre os pixels devido às imperfeições do processo de fabricação e a falta de homogeneidade do silício, a impressão digital é, basicamente, um sinal estocástico de espectro espalhado, sendo, portanto, robusto frente à distorção.

[FIG8] Uma foto original (a) tirada com uma câmera Olympus C765; uma imagem forjada (b) e sua detecção a partir de uma foto adulterada armazenada no formato TIFF (c); JPEG com 75% de qualidade, com ruído removido usando filtro Wiener 3×3 e salva como (d), JPEG com 90% de qualidade (e) e com correção gama com $\gamma = 0{,}5$ e armazenada como JPEG com 90% de qualidade (f).

Como a dimensionalidade da impressão digital é igual ao número de pixels, a impressão digital é única para cada câmera e a probabilidade de haver duas câmeras com a mesma impressão digital é extremamente baixa. A impressão digital também é estável ao longo do tempo. Todas essas propriedades fazem dela uma excelente ferramenta de detecção, adequada para muitas tarefas forenses, tais como identificação de dispositivos, correlação com um dispositivo e detecção de adulteração.

Este tutorial descreve os métodos usados para estimar a impressão digital a partir de imagens capturadas pela câmera e os métodos para detecção de impressões digitais. O estimador é derivado usando-se o princípio de máxima verossimilhança a partir de um modelo simplificado de saída de dados do sensor. Em seguida, o modelo é usado para formular a detecção da impressão digital como um problema de teste de hipóteses de dois canais, para o qual o detector de verossimilhança generalizada é derivado. Devido à complexidade, o teste da razão de verossimilhança generalizada (GLRT, na sigla em inglês) é substituído por um detector simplificado, porém mais rápido, calculado por meio da transformada rápida de Fourier.

O desempenho dos testes forenses apresentados é demonstrado brevemente por meio de imagens reais. Por todo o texto, referências a artigos previamente publicados ajudam o leitor interessado a encontrar informações técnicas mais detalhadas.

Para complementar, lembramos que existem abordagens que combinam a detecção de ruídos com a classificação por meio do aprendizado de máquina [14] - [16]. As referências [14], [17] e [18] estendem aos scanners os métodos forenses baseados em sensores. Uma versão mais antiga desse método forense foi testada para câmeras de telefones celulares em [16] e [19], quando os autores mostram que combinar os métodos forenses baseados em sensores com os métodos que identificam a marca da câmera pode reduzir a ocorrência de falsos positivos. No entanto, é improvável que a melhoria relatada em [19] seja válida para a nova versão do método forense baseado no ruído do sensor, apresentado neste tutorial, pois os resultados parecem ser fortemente influenciados pelos efeitos não corrigidos, discutidos em "Estimação da Impressão Digital de Sensores" na seção de mesmo nome. O problema de se comparar um grande número de imagens foi discutido em [20] usando uma abordagem ad-hoc. A anisotropia do ruído de imagens para a classificação de imagens geradas por scanners, câmeras e computadores foi discutida em [21].

AGRADECIMENTOS

O trabalho realizado neste artigo foi financiado pela bolsa de pesquisas FA9550-06-1-0046 do *Air Force Office of Scientific Research*. As opiniões e conclusões contidas neste artigo são do autor e não representam políticas oficiais, explícitas ou implícitas, do *Air Force Office of Scientific Research* ou do governo americano.

AUTOR

Jessica Fridrich (fridrich@binghamton.edu) é professora de engenharia elétrica e da computação na Binghampton University (SUNY). Ela concluiu o seu doutorado em ciência de sistemas na Universidade de Binghampton em 1995 e o mestrado em matemática aplicada na Universidade Técnica da República Tcheca, em Praga, em 1987. Seus principais interesses são esteganografia, estego-análise, marcação digital e análise forense de imagens digitais. Desde1995, ela já recebeu 18 bolsas de pesquisas, a maioria delas para projetos sobre *data embedding* e estego-análise, gerando mais do que 80 artigos publicados e sete patentes nos EUA. Ela é membro do IEEE e da ACM.

REFERÊNCIAS

[1] J. R. Janesick, "Scientific Charge-Coupled Devices," in *Monograph,* vol. PM83. Bellingham, WA: SPIE, 2001.

[2] G. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 3, pp. 267–276, Mar. 1994.

[3] S. M. Kay, "Fundamentals of Statistical Signal Processing," in *Estimation Theory,* vol. 1. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[4] S. M. Kay, "Fundamentals of Statistical Signal Processing," in *Detection Theory*, vol. 2. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[5] A. El Gamal, B. Fowler, H. Min, and X. Liu, "Modeling and estimation of FPN components in CMOS image sensors," in *Proc. SPIE, Solid State Sensor Arrays: Development and Applications II*, San Jose, CA, Jan. 1998, vol. 3301-20, pp. 168–177.

[6] T. Filler, J. Fridrich, and M. Goljan, "Using sensor pattern noise for camera model identification," in *Proc. IEEE ICIP 08*, San Diego, CA, Sept. 2008.

[7] C. R. Holt, "Two-channel detectors for arbitrary linear channel distortion," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 3, pp. 267–273, Mar. 1987.

[8] B. V. K. Vijaya Kuma and L. Hassebrook, "Performance measures for correlation filters," *Appl. Opt.*, vol. 29, no. 20, pp. 2997–3006, 1990.

[9] M. Goljan and J. Fridrich, "Camera identification from cropped and scaled images," in *Proc. SPIE Electronic Imaging, Forensics, Security, Steganography, and Watermarking of Multimedia Contents X*, San Jose, CA, Jan. 28–30, 2008, vol. 6819, pp. 0E-1–0E-13.

[10] M. Goljan and J. Fridrich, "Camera identification from printed images," in *Proc. SPIE Electronic Imaging, Forensics, Security, Steganography, and Watermarking of Multimedia Contents X*, San Jose, CA, Jan. 28–30, 2008, vol. 6819, pp. OI-1–OI-12.

[11] M. Goljan, Mo Chen, and J. Fridrich, "Identifying common source digital camera from image pairs," in *Proc. IEEE ICIP 07*, San Antonio, TX, 2007.

[12] M. Chen, J. Fridrich, and M. Goljan, "Source digital camcorder identification using ccd photo response non-uniformity," in *Proc. SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, San Jose, CA, Jan. 28–Feb. 1, 2007, vol. 6505, pp. 1G–1H.

[13] M. Chen, J. Fridrich, and M. Goljan, and J. Luká , "Determining image origin and integrity using sensor noise," *IEEE Trans. Inform. Sec. Forensics*, vol. 3, no. 1, pp. 74–90, Mar. 2008.

[14] H. Gou, A. Swaminathan, and M. Wu, "Robust scanner identification based on noise features," in *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, San Jose, CA, Jan. 29–Feb. 1, 2007, vol. 6505, pp. 0S–0T.

[15] N. Khanna, A. K. Mikkilineni, G. T. C. Chiu, J. P. Allebach, and E. J. *Delp*, III, "Forensic classification of imaging sensor types," in *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, San Jose, CA, Jan. 29–Feb. 1, 2007, vol. 6505, pp. 0U–0V.

[16] B. Sankur, O. Celiktutan, and I. Avcibas, "Blind identification of cell phone cameras," in *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, San Jose, CA, Jan. 29–Feb. 1, 2007, vol. 6505, pp. 1H–1I.

[17] T. Gloe, E. Franz, and A. Winkler, "Forensics for flatbed scanners," in *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, San Jose, CA, Jan. 29–Feb. 1, 2007, vol. 6505, pp. 1I–1J.

[18] N. Khanna, A. K. Mikkilineni, G. T. C. Chiu, J. P. Allebach, and E. J. Delp, III, "*Scanner identification using sensor pattern noise," in Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, San Jose, CA, Jan. 29–Feb. 1, 2007, vol. 6505, pp. 1K–1L.

[19] Y. Sutcu, S. Bayram, H. T. Sencar, and N. Memon, "Improvements on sensor noise based source camera identification," in *Proc. IEEE Int. Conf. Multimedia and Expo*, July 2007, pp. 24–27.

[20] G. J. Bloy, "Blind camera fingerprinting and image clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 30, no. 3, pp. 532–534, Mar. 2008.

[21] N. Khanna, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Forensic techniques for classifying scanner, computer generated and digital camera images," in *Proc. IEEE ICASSP*, Mar. 31–Apr. 4, 2008, pp. 1653–1656.

[22] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Phoenix, AZ, Mar. 1999, vol. 6, pp. 3253–3256. SP

Robert C. Maher

Análise Forense de Áudio

Autenticidade, realce e interpretação

O campo da análise forense de áudio envolve muitos tópicos familiares à comunidade de processamento digital de sinais (PDS), tais como reconhecimento de voz, identificação do interlocutor e realce da qualidade do sinal. Muito se tem a ganhar com a aplicação da teoria moderna de PDS aos problemas de interesse da comunidade forense e este artigo foi escrito para dar às pessoas interessadas nessa tecnologia algumas idéias sobre os problemas e desafios enfrentados pelos peritos nos laboratórios de análise forense de áudio. Por outro lado, este artigo também discute as frustrações e armadilhas enfrentadas pelos peritos em processamento de sinais no seu trabalho forense rotineiro devido às normas e práticas do sistema judicial.

Digital Object Identifier 10.1109/MSP.2008.931080

INTRODUÇÃO

A análise forense de áudio consiste na aquisição, análise e avaliação de gravações de áudio que possam vir a ser apresentadas como prova admissível num tribunal ou órgão oficial. Provas forenses na forma de áudio normalmente são obtidas como parte de uma investigação policial ou inquérito oficial devido a um acidente ou outro tipo de ocorrência civil.

As principais preocupações em relação à análise forense de material de áudio são (i) determinar a autenticidade da prova em áudio [1], [2]; (ii) realçar a qualidade das gravações para melhorar a inteligibilidade e a audibilidade de sons de má qualidade [3], [4] e (iii) interpretar e documentar provas sonoras, identificando os interlocutores, transcrevendo o diálogo e reconstruindo cenas do crime ou acidente e a sua ordem cronológica. As exigências para que as provas sejam consideradas admissíveis por um tribunal significa que as téc-

 1053-5888/09/$25.00©2009IEEE

nicas empregadas devem comprovadamente, perante o tribunal, ser imparciais, possuir estatísticas de confiabilidade, ser reconhecidas, não-destrutíveis e amplamente aceitas pelos peritos da área.

Uma técnica moderna de realce da voz, por exemplo, pode parecer a escolha óbvia para melhorar a qualidade de uma gravação de videovigilância antes de transcrevê-la, mas talvez seja necessário convencer o tribunal de que o "realce" não alterou o significado ou a interpretação do diálogo gravado. Se o ruído for removido pelo procedimento de realce da voz, os advogados de defesa e de acusação e o próprio tribunal podem acreditar que um fonema subjacente também foi inadvertidamente (ou deliberadamente!) alterado. O tribunal poderia julgar, por exemplo, que a técnica de processamento de sinais alteraria a interpretação de uma expressão veemente, como "Eu não fiz isso!" para a frase "Eu também fiz isso!", com várias implicações óbvias para a transcrição. Da mesma forma, determinar a cadeia de custódia e a autenticidade ao manusear arquivos de áudio digital é uma tarefa que vai bem além dos procedimentos operacionais padrão dos laboratórios de pesquisa acadêmica da área de PDS. Sendo assim, devido a essas considerações legais específicas, este artigo precisa incluir ou pouco da história e das práticas da análise forense do áudio, mesmo que os tópicos e técnicas pareçam obsoletos para aqueles de nós que atuamos com pesquisas na área processamento de sinais.

AS GRAVAÇÕES FORENSES DE ÁUDIO GERALMENTE POSSUEM RUÍDOS, DISTORÇÕES, INTERFERÊNCIAS SONORAS E OUTROS PROBLEMAS RELACIONADOS À ÁREA DE PROCESSAMENTO DE SINAIS QUE IMPEDEM UMA ANÁLISE ADEQUADA.

HISTÓRIA

Nos últimos 40 anos, a análise forense de material de áudio se tornou uma profissão reconhecida. Antes do surgimento das ferramentas de PDS, a maioria dos analistas forenses de áudio trabalhava exclusivamente com fitas magnéticas analógicas e equipamentos eletrônicos típicos de um estúdio de gravação; tais como filtros analógicos, equipamentos de *playback* com velocidade variável, compressores de ganho e equipamentos de teste (osciloscópios, microscópios e espectrogramas de voz).

Desde a década de 60, o Federal Bureau of Investigation (FBI) tem analisado gravações de áudio para realce da inteligibilidade e autenticação da voz. Departamentos de polícia do mundo inteiro também desenvolveram procedimentos e normas para o manuseio de material de áudio para fins de perícia forense, de acordo com as leis e costumes locais.

A autenticidade das gravações em fitas analógicas era (e ainda é) analisada por meio da revelação magnética, que usa uma suspensão coloidal de partículas magnéticas finas num fluído que evapora após ser borrifado na fita [1]. Quando secas, as partículas magnéticas aderem aos domínios magnéticos gravados na fita, revelando, sob um microscópio, os padrões magnéticos subjacentes. Determinar a autenticidade dos arquivos digitais modernos pode ser um problema, conforme descrito mais adiante neste artigo.

Antes da era digital, o examinador forense de áudio, encarregado de realçar um sinal, transcrever os diálogos e identificar o interlocutor, tocaria a fita original (ou uma cópia de referência) repetidamente, ajustando as configurações do filtro e de ganho para obter o melhor resultado subjetivo

ANÁLISE FORENSE DE ÁUDIO E A LEI

Nos Estados Unidos, a base legal para a admissibilidade de provas acústicas na forma de conversas gravadas é, geralmente, a decisão judicial de 1958 do caso *United States versus McKeever (69 F.Supp. 426, 430, S.D.N.Y. 1958)*. O juiz do caso McKeever teve que decidir sobre a admissibilidade de uma conversa gravada envolvendo o réu. No caso McKeever, uma transcrição foi apresentada ao júri, ao invés da reprodução da fita em si. Mesmo assim, a decisão do juiz estabeleceu uma série de exigências que são utilizadas até hoje, com alguma variação, na maior parte dos tribunais federais dos Estados Unidos para julgar a autenticidade de gravações de áudio. Os sete princípios que regem a determinação da autenticidade de materiais de áudio, elaborados a partir do caso McKeever, estão relacionados na Tabela 1.

Os princípios 1 e 2 deixaram de ser tão relevantes quanto eram no final da década de 50, quando os gravadores de fita não eram os equipamentos ubíquos e comuns que os juízes e júris conhecem hoje. Os princípios 3 e 4 tratam do envolvimento de peritos de áudio que podem analisar a fita física e as características magnéticas sutis do equipamento de gravação para determinar se há montagens ou outras questões que indiquem danos involuntários ou adulteração deliberada. O princípio 5 exige uma cadeia de custódia adequada da gravação e o princípio 6 exige que os participantes da gravação sejam identificados pela voz ou por testemunhas que corroborem com a gravação. Finalmente, o princípio 7 exige que a conversa gravada seja espontânea e não coagida.

Em termos legais, tanto a autenticidade da prova física, composta pela fita e pelo sistema de gravação, quanto as implicações legais da transcrição – geralmente sujeitas à interpretação e ações contenciosas – são assuntos a serem resolvidos pelo tribunal. Como a conversa gravada geralmente é obtida fora do tribunal, os participantes não estão sob juramento e as testemunhas talvez não possam ser consultadas, o tribunal precisa determinar se a gravação é admissível como prova.

TABELA 1] PRINCÍPIOS BASEADOS NO CASO MCKEEVER PARA DETERMINAR A AUTENCIDADE DE MATERIAL DE ÁUDIO

1) QUE O EQUIPAMENTO DE GRAVAÇÃO TENHA SIDO CAPAZ DE GRAVAR A CONVERSA QUE ESTÁ SENDO OFERECIDA COMO PROVA.
2) QUE O OPERADOR DO EQUIPAMENTO TIVESSE COMPETÊNCIA PARA OPERÁ-LO.
3) QUE A GRAVAÇÃO SEJA AUTÊNTICA E CORRETA.
4) QUE A GRAVAÇÃO NÃO TENHA SOFRIDO ALTERAÇÕES, ADIÇÕES OU EXCLUSÕES.
5) QUE A GRAVAÇÃO TENHA SIDO PRESERVADA DA FORMA DEMONSTRADA NO TRIBUNAL.
6) QUE OS INTERLOCUTORES SEJAM IDENTIFICADOS.
7) QUE A CONVERSA EXTRAÍDA TENHA OCORRIDO DE FORMA VOLUNTÁRIA E DE BOA FÉ, SEM QUALQUER TIPO DE COERÇÃO.

TESTEMUNHAS ESPECIALISTAS

Nos Estados Unidos, todos os estados e circunscrições federais utilizam uma série de normas para aceitar o testemunho de especialistas em assuntos específicos. Essas normas geralmente são baseadas no caso Frye de 1923 *(Frye versus United States, 54 App. D.C. 46, 293F, 1013, DC CT APP 1923)*, no caso Daubert *(Daubert versus Merrel Down Pharmaceuticals, 509, U.S. 579 1993)* ou em interpretações similares. A norma Frye exige que os métodos e técnicas do especialista sejam amplamente aceitos pela comunidade científica. Daubert usa as Normas Federais sobre Provas *(Federal Rules of Evidence)* como base para exigir que a relevância e a confiabilidade científica do testemunho do especialista sejam comprovadas. Casos posteriores, como o de *Kumho Tire Co versus Carmichael (526 U.S),* expandiram as normas de Daubert para além dos testemunhos científicos, abrangendo também o conhecimento técnico e outros tipos de conhecimento especializado (como engenharia de áudio).

O INTERVALO DE 18½ MINUTOS

O divisor de águas na área de análise forense de áudio ocorreu em 1974, com a investigação de uma conversa entre o presidente americano Richard Nixon e o Chefe do Estado Maior H.R. Haldeman, gravada no Gabinete Executivo da Casa Branca em 1972. Os investigadores descobriram que a gravação continha um trecho inexplicável de 18½ minutos, durante os quais eram ouvidos zumbidos, mas nenhuma voz identificável. Devido à natureza altamente específica da prova técnica, o juiz John J. Sirica, do Foro da Comarca de Columbia, EUA, designou um Painel Consultivo sobre Fitas da Casa Branca, para fornecer orientação técnica ao tribunal. O painel consultivo consistia de seis especialistas técnicos indicados em conjunto pelo conselho do presidente e pelo promotor, sob orientação do tribunal "... para estudar aspectos relevantes da fita e dos sons nela gravados" [6]. Os membros do painel eram pessoas conhecidas das comunidades de engenharia e acústica: Richard H. Bolt *(Bolt, Beranek, and Newman)*, Franklin S. Cooper *(Haskins Laboratories)*, James L. Flanagan *(AT&T Bell Laboratories)*, John G. McKnight *(Magnetic Reference Laboratory)*, Thomas G. Stockham Jr. *(University of Utah)*, e Mark R. Weiss *(Federal Scientific Corp.)*.

O painel consultivo realizou uma série de análises objetivas da fita propriamente dita, dos seus sinais magnéticos, dos sinais elétricos e acústicos gerados na reprodução da fita e das propriedades do equipamento de gravação usado para produzir os sinais magnéticos na fita. As análises incluíram a observação dos sinais de áudio e o desenvolvimento magnético dos padrões de domínio e assinaturas do cabeçote da fita. Por fim, o painel determinou que o intervalo de 18½ minutos se devia à exclusão de diversos trechos, realizada com um modelo específico de gravador, diferente daquele que produziu a gravação original. A conclusão do painel foi baseada, principalmente, nas assinaturas características de início e parada presentes na fita investigada.

A análise feita pelo painel se tornou, rapidamente, a abordagem padrão de avaliação da autenticidade de gravações de áudio por peritos forenses:

1) Observe fisicamente todo o comprimento da fita.
2) Anote o comprimento e a integridade mecânica da fita, bobinas e cartucho.
3) Verifique se a gravação é contínua ou se possui sequências nas quais a gravação foi iniciada/interrompida ou apagada.
4) Ouça a fita inteira com atenção.
5) Use o processamento não destrutivo de sinais quando necessário para melhorar a inteligibilidade.

Outros casos de análise forense de áudio incluem a tentativa de reconstrução de um crime utilizando, como prova acústica, uma gravação feita num ditafone do Departamento de Polícia de Dallas, envolvendo, supostamente, o assassinato do presidente Kennedy [7]; a interpretação de sons de fundo de um gravador de voz (caixa preta) de uma cabine de controle de uma aeronave; e o uso de técnicas de identificação de voz para identificar gravações de Osama Bin Laden e outros terroristas [9].

AUTENTICIDADE

Como outros tipos de evidência forense, o áudio pode estar sujeito à adulteração acidental ou deliberada. O tribunal deve ser convencido da autenticidade e integridade da prova em áudio. O que é autenticidade? Como ela pode ser demonstrada? Como o PDS pode ajudar?

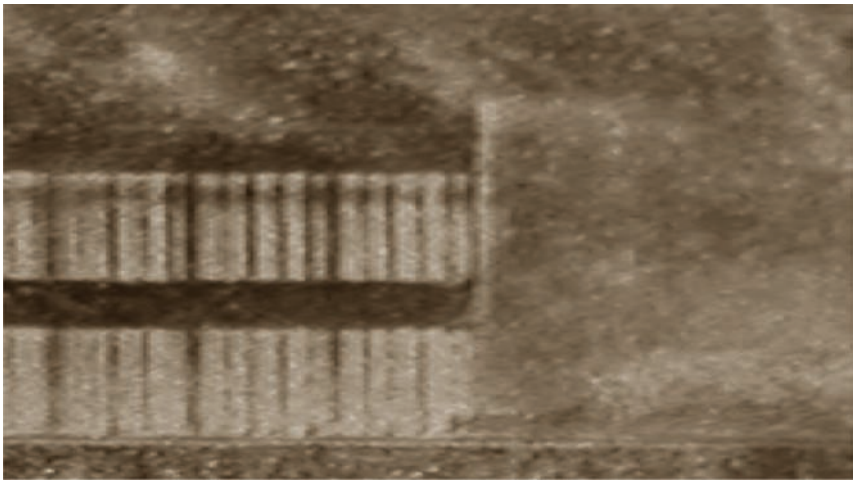
MUITO SE TEM A GANHAR COM A APLICAÇÃO DA TEORIA MODERNA PDS AOS PROBLEMAS DE INTERESSE DA COMUNIDADE FORENSE, MAS OS PERITOS DEVEM ESTAR FAMILIARIZADO COM AS NORMAS E PRÁTICAS DO SISTEMA JUDICIAL.

A gravação em áudio feita para fins forenses deve ser produzida tendo em mente a verificação de sua autenticidade. Por exemplo, a gravação deve incluir uma introdução, apresentando informações relevantes, tais como o local, data, hora, participantes, modelo e número de série do gravador e da mídia usada para a gravação, etc. A gravação deve ser feita em sessão contínua, sem pausas ou interrupções. A verificação da autenticidade pode ser facilitada permitindo-se, de forma deliberada, a inclusão de sons de fundo identificáveis, como sons de relógio ou transmissões de rádio, na gravação.

GRAVAÇÕES DE ÁUDIO PARA ANÁLISE FORENSE

A autenticidade das provas forenses de áudio sempre se baseou em gravações feitas em fitas magnéticas analógicas. Um perito forense faz uma série de observações e testes para avaliar a integridade da gravação [2], [5], [10]. Apesar do surgimento dos gravadores digitais com memória flash de estado sólido, muitos órgãos responsáveis pela aplicação da lei nos Estados Unidos ainda confiam quase que exclusivamente nos gravadores de fita cassete e microcassete analógicos, pois esses já fazem parte do seu inventário e já são velhos conhecidos, assim como os problemas relacionados à autenticidade de dados digitais, como será mencionado posteriormente nessa seção. O procedimento seguido para confirmação da autenticidade de peças de áudio para

análise forense em mídias como fitas analógicas geralmente segue a estratégia empregada pelo painel consultivo nas fitas da Casa Branca. A metodologia exige que o perito verifique a integridade física do meio de gravação, a qualidade do áudio gravado e a consistência das assinaturas magnéticas presentes na fita. Os detalhes do processo são descritos a seguir.



[FIG1] As duas barras listradas no lado esquerdo da figura são traços magnéticos de uma gravação de dois canais. A linha vertical próxima ao centro e a imagem borrada à esquerda são resultados da energização do cabeçote de apagamento e a destruição dos padrões magnéticos subjacentes gravados anteriormente. Como um segmento contínuo e inalterado de fita magnética não apresentaria essa assinatura de apagamento, o perito forense suspeitaria que a fita foi alterada deliberadamente após a gravação original (de [11]).

ANÁLISE E INSPEÇÃO FÍSICA

O perito documenta a condição e as propriedades da gravação usada como prova, inclusive o comprimento e a condição da fita, das bobinas e do cartucho; os números de série; o número de lote e as configurações magnéticas da fita (número de faixas; sistema mono ou estéreo, etc.). É preciso verificar a fita em si à procura de emendas ou outro tipo de alteração e verificar e testar o gravador usado para gravar a fita.

AUDIÇÃO CRÍTICA

O perito ouve cuidadosamente a gravação inteira e anota eventuais alterações ou irregularidades aparentes. Qualquer prova audível de edição, montagem ou descontinuidade nos sons de fundo, ruídos, tons e afins, são anotados.

OBSERVAÇÃO DA ASSINATURA MAGNÉTICA E DA ONDA

A condição dos sinais magnéticos presentes na fita é examinada por meio de técnicas de revelação magnética e comparada às assinaturas de referência de gravações obtidas do mesmo gravador. Um exemplo de revelação magnética é descrito na Figura 1. É analisada a consistência das assinaturas magnéticas relacionadas ao cabeçote de gravação e de apagamento, assim como as transições entre parar e gravar, gravar e pausar, regravações e assim por diante. O perito também observa e mede as ondas elétricas obtidas durante a reprodução da fita.

PREPARAÇÃO DO RELATÓRIO.

Enfim, o perito analisa as observações e elabora um relatório explicando porque a fita é considerada autêntica, se é uma cópia ou se foi alterada de alguma forma após a gravação original.

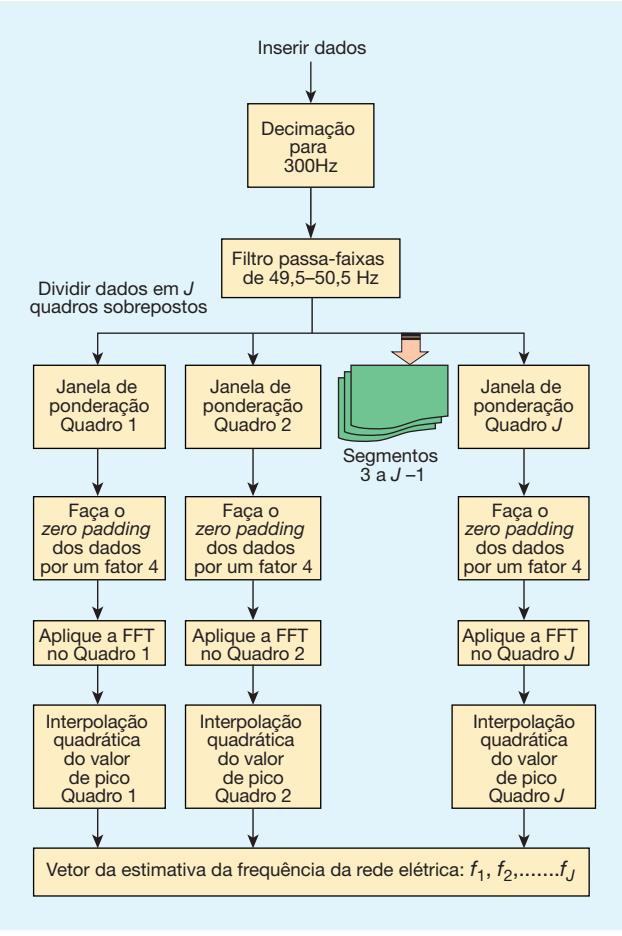
MÍDIA DIGITAL

A questão da autenticidade fica mais complicada com gravações digitais por que as provas de adulteração ou alteração são mais difíceis de verificar do que a montagem mecânica ou assinaturas de *overdubbing* em fitas analógicas físicas. Uma gravação digital pode ser codificada com uma soma de verificação *(checksum)*, processada com uma marca d'água digital incorporada, ou codificada de outra forma, mas é difícil excluir a possibilidade de que o próprio conteúdo de áudio tenha sido editado ou manipulado antes de ter sido recodificado de forma ilícita. Em geral, é essencial haver informações auxiliares e uma cadeia de custódia meticulosa.

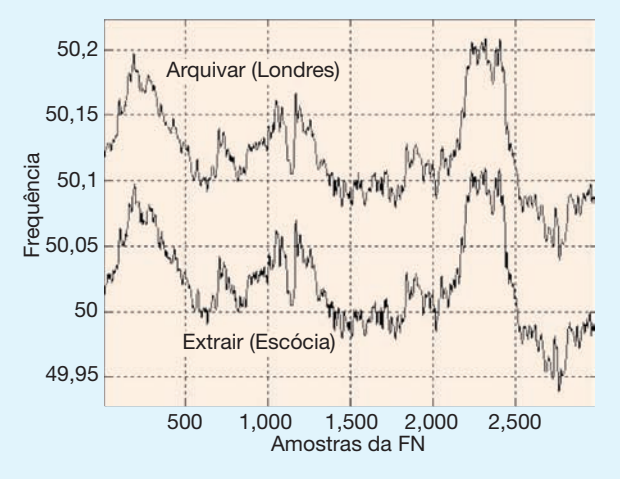
Uma atividade recente de autenticação de áudio que se aplica a gravações digitais é a análise dos sinais residuais decorrentes do acoplamento da freqüência da corrente de energia elétrica no sistema de gravação de áudio [12-16]. A frequência nominal (FN) é de 60Hz nos Estados Unidos e 50Hz em vários outros lugares do mundo. Ela não é precisamente constante, mas varia em 1/20,5 Hz ocasionalmente, de forma imprevisível, devido a pequenas diferenças entre a carga do sistema elétrico e a geração do sistema. O campo magnético alternado, que emana das linhas de corrente alternada, pode causar zumbidos na gravação, o que é indesejável. No entanto, quando há zumbido, as flutuações características de freqüência podem ser rastreadas, revelando uma data e hora específica, desde que haja dados suficientes disponíveis, uma vez que a FN é a mesma em toda a região geográfica atendida por uma rede de corrente alternada síncrona. Portanto, a FN extraída de uma gravação forense poderia ser comparada à um banco de dados de FNs conhecidas da rede de energia elétrica, a fim de obter-se a data e a hora da gravação [15].

Um sistema proposto de extração da FN de gravações de áudio é exibido na Figura 2. Uma comparação da amostra da FN obtida de uma gravação de áudio e a FN referencial de um sistema de energia elétrica é exibida na Figura 3.

O procedimento para obtenção da FN nem sempre é aplicável na prática, pois o acoplamento do campo magnético do sistema de energia elétrica pode ser mínimo em equipamentos de áudio bem projetados que usam microfones condensadores ou piezelétricos ou quando equipamentos movidos a bateria são usados em uma área distante da rede de energia elétrica.



[FIG2] Procedimento proposto para processamento da FN (de [15])



[FIG3] Correspondência automatizada encontrada por meio das técnicas descritas. A forma de onda extraída sofreu *offset* de 0.1 Hz para auxiliar na comparação visual (de [15]).

Apesar de não existir um banco de dados de FN oficial em nenhum lugar do mundo, vários países estão buscando uma forma de criar e armazenar seus registros para uso futuro [15]

REALCE

As gravações de áudio para análise forense geralmente contêm ruídos, distorções, interferências sonoras e outros desafios no âmbito do processamento de sinais, que podem impedir uma análise adequada.

Talvez o maior problema do realce, do ponto de vista dos peritos de áudio forense, seja a gravação clandestina de uma conversa por meio de um microfone escondido. A natureza ilícita do sistema de gravação frequentemente leva ao posicionamento errado do microfone em relação aos participantes, provocando interferências do vento e outros sons ambientes, além da fricção e o abafamento devido ao contato do microfone com a roupa.

Quando um sinal de áudio gravado contém ruídos aditivos indesejados, é necessário realçar a razão sinal-ruído antes da reprodução (vide [4], [17]-[25]). O processo de realce geralmente é realizado *offline*, de forma iterativa, usando-se uma cópia digital da gravação probatória, para que a prova original seja mantida intacta.

O procedimento de realce forense é quase sempre realizado em dados monofônicos como um processo cego *single-ended*, uma vez que os únicos dados disponíveis consistem do próprio sinal degradado pelo ruído. O processo de realce deve, portanto, ser flexível e adaptável, para que o perito possa lidar com a interferência variável ao longo do tempo e a corrupção acústica. A maioria dos peritos usa o processamento tanto no domínio do tempo, quanto no domínio da frequência, a fim de fornecer ao ouvinte (estenógrafo, juiz ou júri) um sinal melhor ou mais inteligível do que o sinal ruidoso original.

A meta do realce forense de áudio pode ser melhorar a inteligibilidade e reduzir a fadiga do ouvinte decorrente da transcrição da voz ou ajudar a revelar sons de fundo sutis ou idiossincráticos, que podem ser provas investigativas importantes. Primeiramente, o perito ouve toda a gravação de forma crítica, fazendo anotações em relação ao tempo e à qualidade dos sons identificáveis e ruídos extrínsecos, assim como à qualidade sonora geral do material. Caso o perito determine que o realce é necessário, várias ferramentas de PDS de áudio estão disponíveis.

MÉTODOS COMUNS DE PDS

Os principais procedimentos de realce de áudio para análise forense são os detectores de nível no domínio do tempo e filtros no domínio da frequência.

DETECÇÃO DE NÍVEL NO DOMÍNIO DO TEMPO

O realce no domínio do tempo trata o envelope de amplitude do sinal de áudio gravado. Um exemplo é o ganho de compressão, no qual o nível total (intensidade sonora) do sinal é ajustado para permanecer relativamente constante: trechos mais silenciosos são amplificados e trechos altos são atenuados ou ignorados.

O método tradicional de domínio do tempo para redução de ruído, tanto analógico como digital, usa um nível ou limiar (threshold) de sinal específico, que indica a provável presença do sinal desejado. O limiar é configurado (na maioria das vezes manualmente) alto o suficiente, para que, quando o sinal desejado estiver ausente (quando houver uma pausa entre as sequências ou trechos, por exemplo), não haja nenhum sibilo ou outro tipo

de ruído no fundo. O limiar, no entanto, não deve ser tão alto a ponto de influenciar o sinal desejado, quando este estiver presente. Se os sinais recebidos estiverem abaixo do limiar, presume-se que exista apenas ruído, sendo que o nível do sinal de saída é reduzido ou *gated*, conforme o caso. Neste contexto, o termo *"gated"* significa que não se permite a passagem do sinal. Por meio do monitoramento contínuo do sinal de saída em relação ao limiar, o método de detecção no domínio do tempo permite a passagem ou não do sinal de saída, conforme a variação do sinal de entrada. Em contextos diferentes, esse tipo de sistema de detecção de nível no domínio do tempo é chamado de *squelch*, expansor de faixa dinâmica ou *gate* de ruído. Esse processo pode amenizar o ruído do sinal recebido, fazendo com que o sibilo desapareça durante a pausa entre as palavras ou frases, mas pode não ser eficaz, pois não reduz o ruído de fundo quando o *gate* está aberto e o sinal desejado está supostamente presente.

A META DO REALCE PODE SER MELHORAR A INTELIGIBILIDADE E REDUZIR A FADIGA DO OUVINTE DECORRENTE DA TRANSCRIÇÃO DE UM DIÁLOGO OU AJUDAR A REVELAR SONS DE FUNDO SUTIS OU IDIOSSINCRÁTICOS QUE POSSAM SER PROVAS IMPORTANTES NUMA INVESTIGAÇÃO.

Um diagrama de bloco de um compressor/expansor de sinais de tempo discreto é exibido na Figura 4. O bloco identificado como "detector de nível" é um detector do envelope de amplitude ou do seguidor de pico. Uma abordagem é o uso de uma comparação de onda completa não linear, como a mostrada abaixo:

$$\begin{aligned} &\text{if } (|x[n]| > c[n-1]) \quad c[n] = \alpha\, c[n-1] \\ &\text{else } \; c[n] = \beta\, c[n-1], \end{aligned} \tag{1}$$

onde α é o coeficiente de ataque escolhido para fornecer o rastreamento desejado quando o nível de entrada estiver aumentando e β o coeficiente de decaimento escolhido para seguir o envelope do sinal em declínio. Sendo assim, é comum escolher $\alpha > 1$ e $0 < \beta < 1$.

O limiar de ganho *(gain threshold)*, c_0, determina o nível no qual a função de controle de ganho se torna eficaz. Se, por exemplo, a expansão do ganho (por *squelch* ou redução do ganho em níveis baixos) for necessária, um cálculo do ganho, como

$$f(c) = \begin{cases} 1, & \text{if } c \geq c_0 \\ (c/c_0)^{\rho-1}, & \text{if } c < c_0 \end{cases} \tag{2}$$

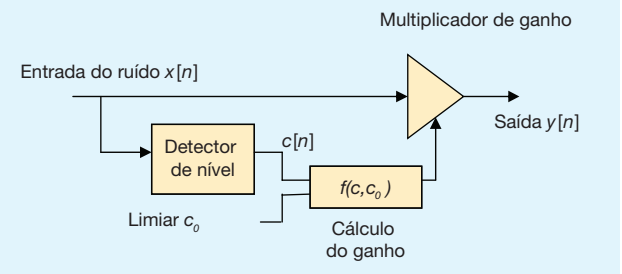
pode ser utilizado [26]. O parâmetro ρ define o fator de expansão: $\rho > 1$ faz com que $f(c)$ seja reduzido quando o nível detectado, $c[n]$, for menor que o limiar c_0. Quanto maior o valor de ρ, , mais abrupta é a alteração do ganho em níveis baixos. Uma abordagem complementar é usada para obter a compressão ou limitação do ganho (redução do ganho em níveis altos).

A forma usual de ilustrar a compressão/ expansão de ganho é por meio de um gráfico que compara o nível de saída com o nível entrada, conforme mostrado na Figura 5. Observe que as curvas de compressão/ expansão mostram o ajuste do nível de saída em relação ao nível do envelope de entrada e não o valor instantâneo do envelope de saída.

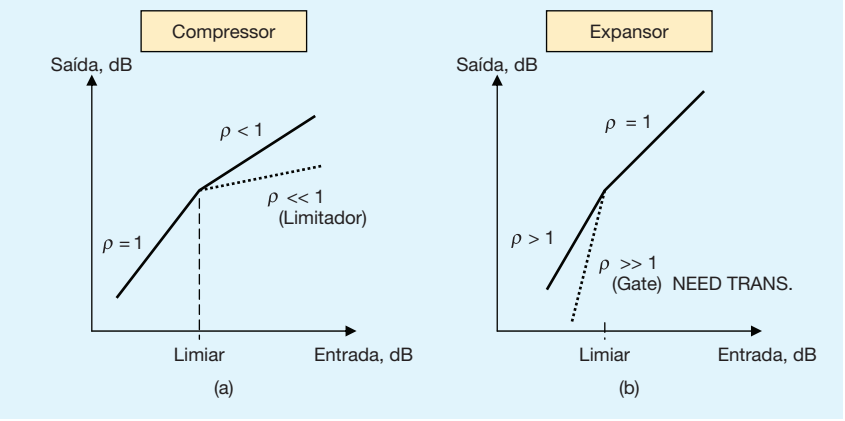
A detecção do nível para a separação dos segmentos que contenham apenas ruídos dos segmentos que contenham sinal e ruído em gravações de áudio para análise forense não é especificamente eficaz, exceto quando longos trechos de ruídos de fundo forem interrompidos de forma intermitente pelo sinal desejado, como no caso de uma conversa gravada num ambiente ruidoso com longas pausas e intervalos.

A simples aplicação do limiar de *gate* não remove o ruído quando o sinal desejado está presente: o *gate* simplesmente "se abre" quando o limiar é excedido. Além disso, o *gate* pode se abrir com um barulho repentino, um clique ou outro tipo de som alto que faça com que o nível do sinal exceda o limiar. Nesse caso, a qualidade do sinal da saída é boa somente quando o sinal for forte o suficiente para mascarar a presença do ruído.

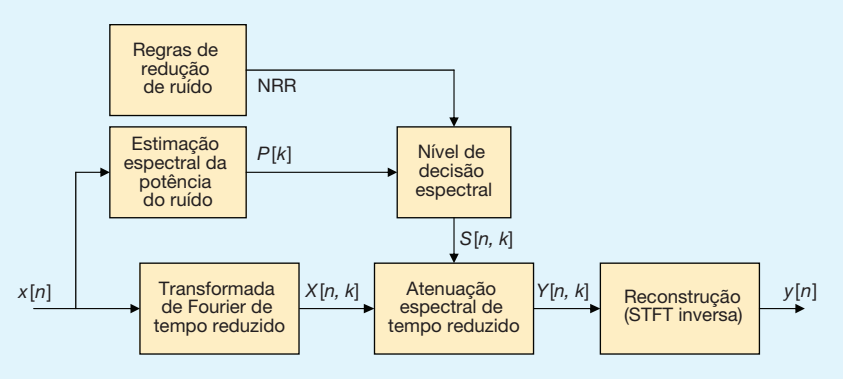
Outra consideração prática é que a alteração do ganho entre o modo de "*gate* aberto" e "*gate* fechado" deve ser feita de forma cuidadosa, para evitar efeitos audíveis de modulação do ruído. O termo *gain pumping* é usado por engenheiros de som e significa o som audível do ruído que emana quando o *gate* se abre e que desaparece quando o *gate* se fecha. No entanto, as funções do compressor e expansor de ganho no domínio do tempo podem ser úteis em situações em que o realce forense for realizado.



[FIG4] Diagrama de blocos de um compressor/ expansor de ganho básico



[FIG5] Ilustração do comportamento do (a) compressor e do (b) expansor de ganho.



[FIG6] Estrutura do sistema de redução do ruído espectral

A eficiência dos métodos do domínio do tempo pode ser aumentada por meio do processamento digital dos sinais. O controle cuidadoso dos tempos de ataque e liberação do gate, ou seja, o quão rapidamente o processador responde às alterações do sinal de entrada, por exemplo, pode minimizar os artefatos que, de outra forma, confundiriam o perito. Outras melhorias trazidas pelo PDS são: o uso de um software de controle com "visão do futuro", para causar a variação automática do limiar quando houver alteração do nível de ruído, e a divisão da decisão de uso do gate em duas ou mais sub-bandas de frequência. Usar várias bandas de frequência com gates individuais permite que o limiar seja estabelecido de forma mais otimizada quando o ruído variar de uma banda de frequência para outra. Por exemplo, caso o ruído seja um ruído ''surdo" ou zumbido de baixa frequência, o limiar pode ser alto o suficiente para remover o ruído da banda de baixa frequência, ao mesmo tempo em que mantém um limiar mais baixo nas faixas de alta frequência. Apesar dessas melhorias, os métodos de processamento no domínio do tempo para realce forense do áudio ainda são limitados, pois o processador não consegue distinguir ruídos do sinal desejado a não ser com base no nível do envelope do sinal absoluto.

FILTRAGEM NO DOMÍNIO DA FREQUÊNCIA

Os métodos de realce forense do áudio no domínio da frequência geralmente utilizam alguma forma de subtração espectral. Como o nome diz, a subtração espectral envolve a formulação da estimativa do espectro do ruído (potência do ruído como uma função da frequência) e, posteriormente, subtração dessa estimativa do sinal principal. A saída com ruído reduzido é criada por meio da reconstrução do sinal a partir do espectro subtraído. Idealmente, toda a energia espectral abaixo do limiar da estimativa de ruído é removida, para que os componentes do sinal desejado excedam o nível do ruído. Se a estimativa do ruído for suficientemente precisa, a técnica pode ser útil e eficaz [20], [21].

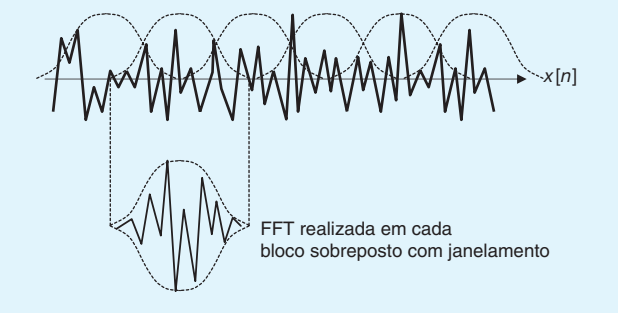
Infelizmente, se o nível de ruído real for diferente do espectro de ruído estimado, a redução do ruído fica incompleta e sujeita à artefatos de áudio indesejáveis. A energia espectral residual próxima ao limiar do ruído tem o som de um assobio ou sino, às vezes chamado de canto de passarinho ou ruído musical [22-25]. Sistemas de subtração espectral práticos possibilitam a atualização frequente da estimativa do nível de ruído e incorporam as técnicas de redução do ruído musical. Sendo assim, a qualidade e a eficácia da técnica de subtração espectral dependem da realização de atividades forenses específicas e das respectivas exigências de processamento.

A estrutura básica da redução do ruído espectral é ilustrada na Figura 6. A transformada de Fourier de tempo curto (STFT, na sigla em inglês) produz uma sequência temporal de quadros espectrais ou *snapshots*, $X[n, k]$, geralmente com um salto entre os quadros sobrepostos, conforme mostra a Figura 7 [27], [28]. Em cada quadro, o algoritmo de redução do ruído espectral toma decisões sobre quais componentes são o sinal desejado e quais são atribuíveis ao ruído.

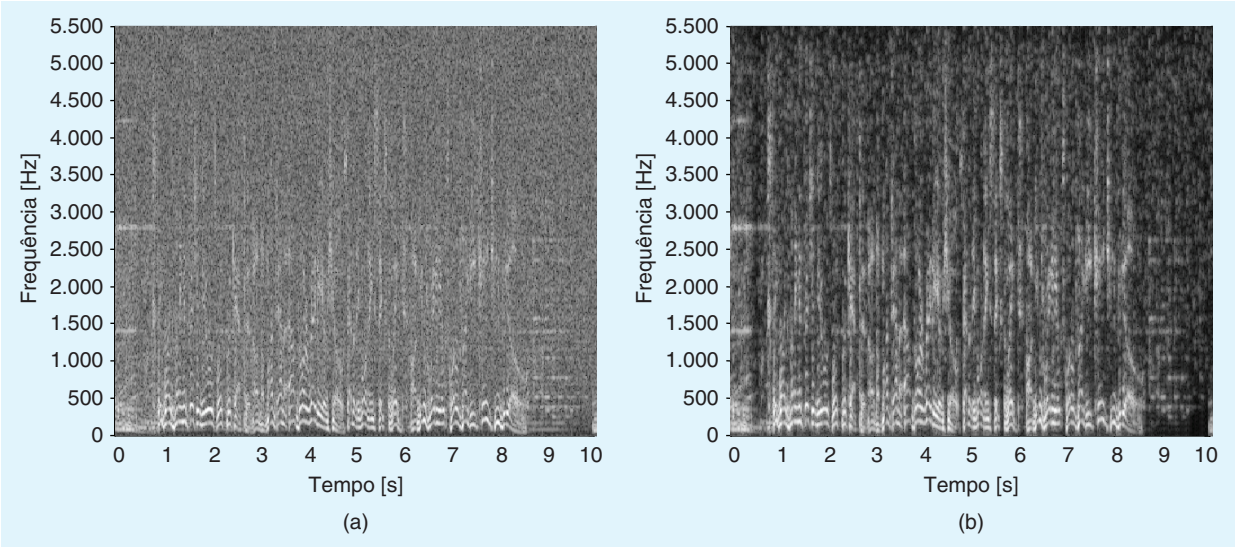
Um método híbrido *single-ended* de redução de ruído expande a detecção do nível no domínio do tempo e os conceitos de subtração espectral no domínio da frequência, fornecendo um meio de distinguir o comportamento coerente dos componentes do sinal desejado e o comportamento não coerente (não correlacionado) do ruído aditivo [4], [24], [25], [29]. Esse procedimento identifica características que se comportam de forma consistente em uma janela de tempo reduzido e atenua ou remove características que apresentam flutuações randômicas ou inconsistentes. Como um método *single-ended*, a distinção entre ruído e sinal pode não ser perfeita, mas para muitos sinais forenses importantes (voz com ruído, por exemplo), o processo pode ser suficientemente confiável para melhorar o sinal de saída para análise posterior.

O problema de simplesmente utilizar a filtragem do ruído espectral é que sinais comuns, como a voz humana, contêm componentes fricativos e oclusivos, importantes para a sua inteligibilidade. A detecção do padrão iterativo também é normalmente empregada, para que os componentes fricativos sejam permitidos, principalmente nas fronteiras entre os intervalos sem sinais sonoros e intervalos com componentes sonoros. Isso porque a presença e a audibilidade de fonemas consoantes de prefixos e sufixos são uma característica essencial para o reconhecimento da fala [25].

Um exemplo de redução do ruído espectral com variação temporal (no domínio da frequência) é exibido na Figura 8.



[FIG7] Transformada Fourier de tempo reduzido usando o processamento baseado em blocos sobrepostos.



[FIG8] Exemplo de redução do ruído espectral. (a) diagrama de magnitude da STFT da voz com ruído. O contraste baixo e o sombreamento cinza revelam o ruído de fundo. (b) Magnitude da STFT da versão realçada da voz de (a). O maior contraste entre os componentes do sinal desejado (brancos) e o nível de fundo (preto) mostra, de forma qualitativa, a relação sinal-ruído melhorada (de [25]).

INTERPREÇÃO

Na conclusão de uma análise forense de áudio, o solicitante geralmente exige que o perito elabore um relatório descrevendo e interpretando a prova, os métodos utilizados e a base estatística das opiniões fornecidas.

Devemos observar que, apesar do melhor desempenho dos sistemas modernos de identificação de voz e de interlocutores, não há registro de nenhum julgamento nos Estados Unidos em que o juiz tenha aceitado como prova uma transcrição ou identificação feita por computador. O desafio para as pesquisas acerca do realce de áudio e da inteligibilidade de vozes é demonstrar um desempenho adequado aos padrões exigidos por um tribunal para determinar se um réu é culpado ou inocente, tendo os tribunais americanos sempre preferido a interpretação de um ser humano especializado, conforme descrito abaixo. Mostramos, a seguir, alguns exemplos da fase de interpretação em projetos de análise forense de áudio.

A ANÁLISE FORENSE DE ÁUDIO CONSISTE NA AQUISIÇÃO, ANÁLISE E AVALIAÇÃO DE GRAVAÇÕES DE ÁUDIO QUE POSSAM SER APRESENTADAS COMO PROVA ADMISSÍVEL EM UM TRIBUNAL OU OUTRO ÓRGÃO OFICIAL.

IDENTIFICAÇÃO ESPECTROGRÁFICA E AUDITIVA DA VOZ

A análise forense de um diálogo gravado pode dar origem a um litígio judicial sobre a identidade de um ou mais participantes da conversa. O suspeito ou a outra parte de um processo civil pode negar ser o indivíduo que proferiu as palavras registradas, principalmente se o registro tiver sido feito por telefone e não tenha havido nenhuma testemunha ocular para identificar o interlocutor visualmente. Nesse caso, o perito de áudio forense pode ser encarregado de identificar o suspeito ou afirmar que ele não proferiu as palavras na gravação em questão.

O método auditivo-espectrográfico de identificação da voz para a análise forense é baseado no julgamento de um perito treinado que compara um exemplar desconhecido de voz com um ou mais exemplares conhecidos [30]-[33].

Como o próprio nome do método indica, o objetivo do perito é fornecer uma opinião com base na comparação auditiva (por meio da audição meticulosa do material) e na comparação visual dos espectrogramas da voz.

Num caso comum, o perito começa ouvindo a gravação do interlocutor desconhecido de maneira crítica e identificando frases específicas que sejam peculiares e relativamente livres de ruídos.

O perito então marca uma sessão de gravação com o suspeito para criar exemplares cujo ritmo, ênfase e enunciado combinem com as frases escolhidas da voz desconhecida. O suspeito repete cada frase várias vezes para produzir gravações com o mesmo padrão de tempo e de voz dos exemplares desconhecidos.

Posteriormente, o perito gera arquivos de áudio individuais contendo os exemplos produzidos para cada frase distinta. Os arquivos são usados para comparações auditivas do tipo “A-B” e também para criar espectrogramas para comparação visual do formato, características espectrais discretas e outros padrões.

Apesar de o perito em análise auditivo-espectrográfica poder usar o processamento de sinais para realce e exibição espectral, são as suas observações auditivas e visuais que servem de base para sua opinião sobre a possibilidade de os exemplares serem similares ou não à gravação desconhecida. O perito expressa, especificamente, um dos seguintes pareceres:

1) identificação positiva (a voz desconhecida corresponde à do exemplo, proferida pelo suspeito)
2) identificação provável
3) sem decisão
4) provável eliminação
5) eliminação positiva (o exemplo da voz do suspeito é diferente da voz desconhecida).

A confiança do perito auditivo-espectrográfico em sua experiência anterior no método de correspondência subjetiva de padrões tem provocado controvérsias ao longo dos anos acerca da veracidade e da base científica da análise e, portanto, a sua admissibilidade pelo tribunal [31]. Como não existem estudos científicos aceitos que quantifiquem totalmente a taxa de erros aceitável da análise auditivo-espectrográfica, os tribunais devem julgar a admissibilidade do testemunho do perito conforme o caso [32], [33].

O PROBLEMA DE UTILIZAR APENAS A FILTRAGEM DO RUÍDO ESPECTRAL É QUE SINAIS COMUNS, COMO A FALA HUMANA, CONTÉM COMPONENTES FRICATIVOS E OCLUSIVOS, IMPORTANTES PARA A SUA INTEGIBILIDADE.

INVESTIGAÇÕES DE QUEDA DE AERONAVES

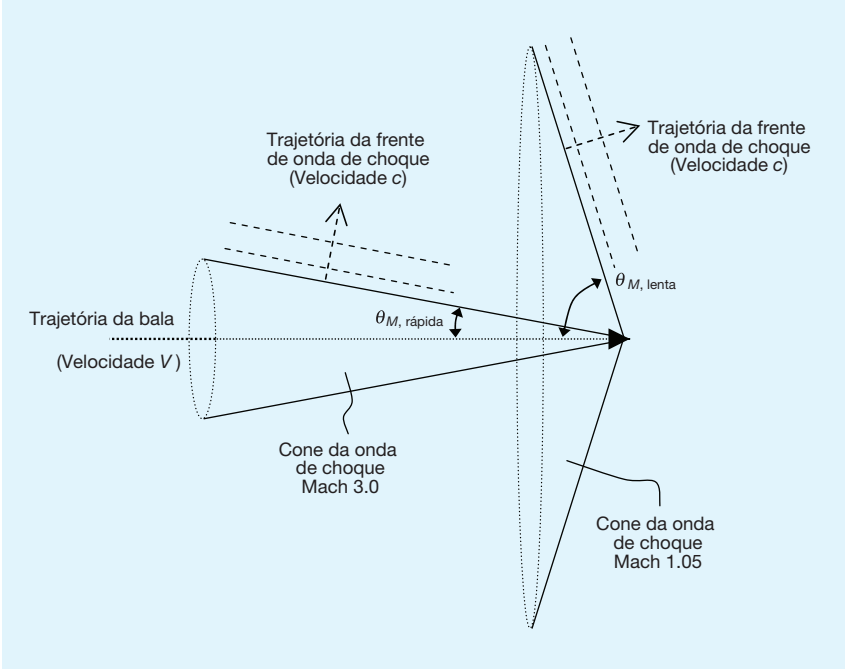
As aeronaves comerciais modernas e algumas aeronaves militares, empresariais e particulares são equipadas com um gravador de dados de voo (FDR, na sigla em inglês) e um gravador de voz na cabine de controle (CVR, na sigla em inglês). O FDR mantém um registro dos parâmetros do voo, tais como hora, altitude, orientação da aeronave, velocidade no ar e outros. O CVR possui canais de áudio, para gravar comunicações via rádio, e um microfone localizado no painel superior acima do assento do piloto, na cabine de controle, para captar conversas e sons de fundo. O FDR mantém um registro de pelo menos 25 horas, enquanto o CVR grava uma sequência de 30 a 120 minutos, para que sejam documentados, no mínimo, os últimos 30 minutos dos sons da cabine de controle em caso de queda ou outro incidente de segurança.

A transcrição das conversas da tripulação é importante para a investigação de acidentes com aeronaves.

Além do diálogo gravado, o CVR reúne outras informações não verbais, como avisos audíveis e sinais de alerta, ruídos mecânicos provenientes da estrutura da aeronave e o som dos motores. Os dados do CVR também podem ajudar os investigadores do acidente a identificar os índices de respiração e outros sinais sutis relacionados ao estresse e esforço da tripulação e outras informações sobre a situação.

Num importante caso envolvendo uma investigação forense de áudio com dados de um CVR, peritos do Conselho Nacional de Segurança nos Transportes dos EUA puderam, por meio da análise meticulosa do conteúdo do CVR do voo 427 da USAir (um Boeing 737), que caiu nos arredores de Pittsburgh em 1994, entender o comportamento dos motores da aeronave, o momento, as reações e as atitudes do piloto e do co-piloto durante o acidente. A investigação incluiu, entre outros detalhes, testes para determinar a capacidade do microfone da cabine de controle de captar o som através da vibração da estrutura [4].

Uma investigação de 1997 analisou os dados do CVR de uma aeronave Beechcraft 1900C, envolvida num acidente em 1992, com base nas características do sinal do microfone da cabine e de um canal de CVR para testar a teoria de que a separação do motor da asa durante o voo ocorreu após um *whirl flutter* da turbina, atribuído a um rompimento no suporte do motor [35].



[FIG9] Comportamento da onda de choque de projéteis supersônicos relativamente rápidos (Mach 3.0) e lentos (Mach 1.05) (informação utilizada com autorização de [38])

ANÁLISE ACÚSTICA DE DISPAROS

A análise forense da gravação de disparos de armas de fogo pode ajudar a verificar as declarações de testemunhas oculares (e auditivas) e auxiliar na reconstrução da cena de um crime. A prova em áudio pode incluir a explosão causada durante o disparo *(muzzle blast)*; a assinatura da onda de choque, provocada quando o projétil viaja em velocidade supersônica; a chegada do som refletido e reverberado de obstáculos próximos e até mesmo a característica do som da ação mecânica da arma de fogo, quando a gravação é feita em local próximo à posição de disparo [35]—[38].

A possível prova de áudio forense relacionada aos disparos pode incluir vários componentes, conforme descrito a seguir.

EXPLOSÃO DO DISPARO

Uma arma de fogo convencional usa a combustão rápida da pólvora para empurrar o projétil para fora da arma. A ejeção rápida de

gás quente da boca da arma de fogo causa uma onda de choque acústico e um ruído estrondoso chamado de explosão do disparo *(muzzle blast)*. Essa explosão geralmente dura somente alguns milésimos de segundo. Quando um microfone está posicionado próximo ao cano da arma, o som direto da explosão do disparo é o principal sinal acústico. Este geralmente causa uma sobrecarga no microfone e/ou no dispositivo de gravação devido ao nível altíssimo de pressão do som. Microfones posicionados a distâncias maiores geralmente exibem o som refletido e as reverberações vindas de várias direções, além de outros efeitos do ambiente sonoro.

AÇÃO MECÂNICA

A ação mecânica da arma de fogo inclui o som do mecanismo de engatilhamento e de disparo, o posicionamento do novo projétil pelo sistema de carregamento automático ou manual e, possivelmente, sons do cartucho vazio sendo ejetado e lançado ao solo. Esses sons mecânicos são bem sutis, se comparados aos altos níveis de pressão do som decorrentes da explosão do disparo; portanto, a detecção da ação mecânica só é possível quando o microfone usado na gravação estiver bem perto do atirador.

DESDE A DÉCADA DE 60, O FEDERAL BUREAU OF INVESTIGATION (FBI) TEM ANALISADO GRAVAÇÕES DE ÁUDIO PARA REALCE DA INTELIGIBILIDADE E AUTENTICAÇÃO DA VOZ.

PROJÉTIL SUPERSÔNICO

Além da explosão do disparo e da ação mecânica, uma terceira fonte de informação acústica de um tiro surge quando o projétil viaja em velocidade supersônica [35], [38]. A velocidade do projétil, V, depende do tamanho da carga, da massa da bala e de outros fatores balísticos. Uma bala supersônica gera um padrão de onda de choque característico conforme cruza o ar. A onda de choque se expande, formando um cone atrás da bala, conforme a frente da onda de choque se propaga para fora, à velocidade do som, denotada por c. O cone da onda de choque possui um ângulo interno θ_M que é relacionado à velocidade da bala pela fórmula $\theta_M = \arcsin(1/M)$, onde $M = V/c$ é o número Mach. A Figura 9 ilustra a geometria da onda de choque.

Um projétil em alta velocidade com V muito maior que c resulta num número Mach muito maior que uma unidade, causando, portanto, uma onda de choque com ângulo de cone estreito, onde a frente da onda de choque se propaga para fora, quase que perpendicularmente ao percurso do projétil. Por exemplo, um projétil que viaja a 1.000 m/s em temperatura ambiente tem $M = 2.67$, resultando em $\theta_M = \sim 22°$. Por outro lado, quando a bala viaja somente um pouco acima da velocidade do som, o M é aproximadamente uma unidade e o ângulo Mach é de mais ou menos 90°, o que significa que a propagação da onda de choque é, basicamente, paralela à trajetória do disparo. Além disso, o projétil desacelera durante o percurso devido à fricção com o ar, fazendo com que o ângulo Mach aumente quando a velocidade é reduzida na queda. Portanto, a estimativa forense da posição e orientação do disparo, com base no momento relativo da chegada da onda de choque e na explosão do disparo (mais o som refletido nos objetos ao redor), deve considerar o número Mach estimado ao longo do percurso do projétil.

QUESTÕES RELACIONADAS AO EXAME DO DISPARO

Os desafios do áudio forense para os peritos em disparos advêm da natureza impulsiva das assinaturas sônicas tanto da explosão do disparo quanto da onda de choque do projétil, quando presente. Apesar de durarem centenas de milésimos de segundos nas trilhas sonoras dos filmes de Hollywood e videogames, os sons da explosão do disparo de uma arma de fogo duram, na vida real, somente de 1 a 3 milésimos de segundo. A assinatura da onda de choque, com ou sem alta pressão, dura algumas centenas de microssegundos. De fato, do ponto de vista forense, as gravações de disparos de armas de fogo, mantidas nas bibliotecas de efeitos sonoros, revelam mais sobre a resposta ao impulso acústico do ambiente ao redor do que sobre a arma de fogo em si. Isso porque essas gravações contêm, deliberadamente, um alto nível de ecos e reverberações artificiais criados com o fim de aumentar o impacto emocional da cena. Na verdade, as testemunhas que ouvem um disparo real geralmente dizem que o som se parecia mais com "estalos" ou "fogos de artifício" do que com tiros, pelo menos, comparados ao que foram influenciadas a acreditar. Mesmo quando a reverberação não é deliberadamente acrescentada, as gravações de disparos obtidas em áreas acusticamente refletoras, tais como no interior ou exterior de edifícios em áreas urbanas, podem conter uma mistura de disparos e ecos sobrepostos que complicam o processo de análise.

Os níveis altíssimos de pressão do som próximo à arma de fogo podem ultrapassar 150 dB re 20 μPa. Os altos níveis de pressão associados aos sons do disparo podem causar uma perturbação no microfone e na fase de entrada do sinal. Além disso, os momentos de aumento extremamente rápidos são distorcidos pelo sistema de gravação, dificultando a observação. Isso é particularmente verdade para gravações obtidas por telefone.

CONCLUSÕES

Este artigo apresentou uma visão geral sobre as práticas atualmente usadas no campo de áudio forense. Os peritos em processamento de sinais podem e devem considerar o uso de aplicações de áudio forense em seu trabalho de pesquisa e desenvolvimento, devendo entender, no entanto, que as exigências e requisitos do sistema penal podem gerar alguma frustração até que novas técnicas sejam avaliadas e consideradas admissíveis.

Há uma clara necessidade de se instruir os tribunais e as pessoas que compõem o grupo de jurados sobre os pontos fortes e fracos do material de áudio forense. Ouvimos, com frequência, na comunidade de áudio forense, relatos envolvendo o "efeito CSI", que se refere à série da TV americana *"Crime Scene Investigation"*. Os juízes e jurados que conhecem a série, podem chegar ao tribunal com altas expectativas sobre a capacidade dos sistemas de realce de sinais e de identificação de voz, que são completamente infundadas na vida real. Só nos resta aguardar para ver se essas falsas crenças levantarão questões no momento da deliberação pelo júri.

Encorajamos todos os membros da Signal Processing Society do IEEE a se informarem mais sobre os aspectos legais

e técnicos da análise forense de áudio, para dar aos profissionais do sistema penal, judicial e de investigação de acidentes as informações e ferramentas necessárias para a execução de seu importante trabalho na sociedade contemporânea.

AGRADECIMENTOS

O autor agradece aos comentários de todos os revisores anônimos e à orientação de Durand Begault, Christoph Musialik, Richard Sanders e Steven R. Shaw.

AUTOR

Robert C. Maher (rob.maher@montana.edu) obteve o título de bacharel pela Universidade de Washington em St. Louis (1984); o título de mestre pela Universidade de Wisconsin, em Madison (1985) e o título de doutor pela Universidade de Illinois, em Urbana-Champaign (1989), todos na área de Engenharia Elétrica. De 1989 a 1996, ele foi membro do corpo docente do Departamento de Engenharia Elétrica da Universidade de Nebraska, em Lincoln. Em 1996, ele se mudou para Boulder, Colorado, para assumir o cargo de vice-presidente de engenharia da empresa EuPhonics Inc. (1996-1998); gerente de engenharia de produtos de áudio, na 3Com/U.S. Robotics (1998-2001) e professor adjunto da Universidade de Colorado, em Boulder (2002-2002). Em 2002, ele ingressou no Departamento de Engenharia Elétrica e da Computação da Universidade Estadual de Montana, em Bozeman, onde é atualmente titular do departamento e professor. Seu trabalho de pesquisa, ensino e consultoria envolve a aplicação de técnicas de PDS a problemas de engenharia de áudio, acústica, som ambiente e áudio forense.

REFERÊNCIAS

[1] B. E. Koenig, "Authentication of forensic audio recordings," *J. Audio Eng. Soc.*, vol. 38, no. 1/2, pp. 3–33, Jan./Feb. 1990.

[2] *AES Standard for Forensic Purposes—Criteria for the Authentication of Analog Audio Tape Recordings*, AES Standard 43-2000.

[3] B. E. Koenig, D. S. Lacey, and S. A. Killion, "Forensic enhancement of digital audio recordings," *J. Audio Eng. Soc.*, vol. 55, no. 5, pp. 252–371, May 2007.

[4] C. Musialik and U. Hatje, "Frequency-domain processors for efficient removal of noise and unwanted audio events," in *Proc. Audio Engineering Society 26th Conf., Audio Forensics in the Digital Age*, Denver, CO, July 2005, pp. 65–77.

[5] *AES Recommended Practice for Forensic Purposes—Managing Recorded Audio Materials Intended for Examination*, AES Standard 27-1996.

[6] Advisory Panel on White House Tapes, *"The Executive Office Building Tape of June 20, 1972: Report on a technical investigation*," United States District Court for the District of Columbia, May 31, 1974. [Online]. Available: http://www.aes.org/aeshc/docs/forensic.audio/watergate.tapes.report.pdf

[7] National Academy of Sciences, *"Report of the Committee on Ballistic Acoustics*," Washington, D.C.:National Academy Press, 1982. [Online]. Available: http://www.nap.edu/catalog.php?record_id=10264

[8] G. Byrne, *Flight 427: Anatomy of an Air Disaster*. New York: Springer-Verlag, 2002.

[9] J. S. Sachs. (2003, Mar.). Graphing the voice of terror. Popular Sci., pp. 38–43 [Online]. Available: http://www.popsci.com/scitech/article/2003-02/graphing-voice-terror

[10] Scientific Working Group on Digital Evidence. (2008, Jan. 31). *SWGDE Best Practices for Forensic Audio*, Version 1.0 [Online]. Available: http://www.swgde.org/documents/swgde2008/SWGDEBestPracticesforForensicAudioV1.0.pdf

[11] D.R. Begault, B.M. Brustad, and A.M. Stanley, "Tape analysis and authentication using multi-track recorders," in *Proc. Audio Eng. Soc. 26th Conf., Audio Forensics in the Digital Age*, Denver, CO, July 2005, pp. 115–121.

[12] C. Grigoras, "Digital audio recording analysis: The electric network frequency (ENF) criterion," *Int. J. Speech Language Law*, vol. 12, no. 1, pp. 63–76, 2005.

[13] E. B. Brixen, "Techniques for the authentication of digital audio recordings," in *Proc. Audio Engineering Society 122nd Conv.*, Vienna, Austria, 2007, Convention Paper 7014.

[14] C. Grigoras, "Application of ENF analysis method in authentication of digital audio and video recordings," in *Proc. Audio Engineering Society 123rd Conv.*, New York, 2007, Convention Paper 1273.

[15] A. J. Cooper, "The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings—An automated approach," in *Proc. Audio Engineering Society 33rd Conf., Audio Forensics—Theory and Practice*, Denver, CO, June 2008, pp. 1–10.

[16] E. B. Brixen, "ENF—Quantification of the magnetic field," in *Proc. Audio Engineering Society 33rd Conf., Audio Forensics—Theory and Practice,* Denver, CO, June 2008, pp. 1–6.

[17] M. R. Weiss, E. Aschkenasy, and T. W. Parsons, "*Study and development of the INTEL technique for improving speech intelligibility*," Nicolet Scientific Corp., Final Rep. NSC-FR/4023, 1974.

[18] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, 1979.

[19] M. R. Weiss and E. Aschkenasy, "*Wideband speech enhancement (addition)*," Final Tech. Rep. RADC-TR-81-53, DTIC ADA100462, 1981.

[20] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, no. 2, pp. 113–120, 1979.

[21] R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, no. 2, pp. 137–145, 1980.

[22] O. Cappé, "Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor," *IEEE Trans. Speech Audio Processing*, vol. 2, no. 2, pp. 345–349, 1994.

[23] D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis, "Speech enhancement based on audible noise suppression," *IEEE Trans. Speech Audio Processing*, vol. 5, no. 6, pp. 479–514, 1997.

[24] S. Godsill, P. Rayner, and O. Cappé, "Digital audio restoration," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Norwell, MA: Kluwer, 1998, pp. 133–194.

[25] R.C. Maher, "Audio enhancement using nonlinear time-frequency filtering," in *Proc. Audio Engineering Society 26th Conf., Audio Forensics in the Digital Age*, Denver, CO, July 2005, pp. 104–112.

[26] S. J. Orfanidis, *Introduction to Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[27] J.B. Allen and L.R. Rabiner, "A unified approach to short time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558–1564, 1977.

[28] T.F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[29] J. Moorer and M. Berger, "Linear-phase bandsplitting: Theory and applications," *J. Audio Eng. Soc.*, vol. 34, no. 3, pp. 143–152, 1986.

[30] R. H. Bolt, F. S. Cooper, E. E. David, P. B. Denes, J. M. Pickett, and K. N. Stevens, "Identification of a speaker by speech spectrograms," *Science*, vol. 166, no. 3903, pp. 338–342, 1969.

[31] R. H. Bolt, F. S. Cooper, E. E. David, P. B. Denes, J. M. Pickett, and K. N. Stevens, "Speaker identification by speech spectrograms: A scientist's view of its reliability for legal purposes," *J. Acoust. Soc. Amer.*, vol. 47, no. 2, pp. 597–612, 1970.

[32] National Academy of Sciences, Committee on Evaluation of Sound Spectrograms, "On the theory and practice of voice identification," Washington, D.C.: National Academy Press, Rep. 0-309-02973-16, 1979.

[33] F. Poza and D. R. Begault, "Voice identification and elimination using aural-spectrographic protocols," in *Proc. Audio Engineering Society 26th Conf., Audio Forensics in the Digital Age*, Denver, CO, July 2005, pp. 21–28.

[34] R. O. Stearman, G. H. Schulze, and S. M. Rohre, "Aircraft damage detection from acoustic and noise impressed signals found by a cockpit voice recorder," in *Proc. Nat. Conf. Noise Control Engineering*, vol. 1, 1997, pp. 513–518.

[35] B. M. Brustad and J. C. Freytag, "A survey of audio forensic gunshot investigations," in *Proc. Audio Engineering Society 26th Conf., Audio Forensics in the Digital Age*, Denver, CO, July 2005, pp. 131–134.

[36] R. C. Maher, "Modeling and signal processing of acoustic gunshot recordings," in *Proc. IEEE Signal Processing Society 12th DSP Workshop*, Jackson Lake, WY, Sept. 2006, pp. 257–261.

[37] R. C. Maher, "Acoustical characterization of gunshots," in *Proc. IEEE SAFE 2007: Workshop on Signal Processing Applications for Public Security and Forensics*, Washington, D.C., Apr. 2007, pp. 109–113.

[38] R. C. Maher and S. R. Shaw, "Deciphering gunshot recordings," in *Proc. Audio Engineering Society 33rd Conf., Audio Forensics—Theory and Practice*, Denver, CO, June 2008, pp. 1–8.

[SP]


IEEE WAS HERE
Members share fascinating first-person stories of technological innovations. Come read and contribute your story.
IEEE Global History Network
www.ieeeghn.org
IEEE