

Caçadores de fraudes

Pesquisa na área da computação forense sugere conjunto de passos para identificar documentos digitais falsificados

MANUEL ALVES FILHO
manuel@reitoria.unicamp.br

Pesquisa desenvolvida no Instituto de Computação (IC) da Unicamp propõe nova abordagem para a identificação de fraudes em documentos digitais. Em vez de utilizar um método único para cada tipo de caso, como se faz usualmente, o estudo sugere a combinação de diferentes técnicas [metodologia denominada de multianálise] como forma de refinar a investigação. O trabalho consistiu na tese de doutoramento do cientista da computação Anselmo Castelo Branco Ferreira, que foi orientado pelo professor Anderson de Rezende Rocha e coorientado pelo professor Jefersson Alex dos Santos, do Departamento de Ciência da Computação (DCC) da Universidade Federal de Minas Gerais (UFMG).

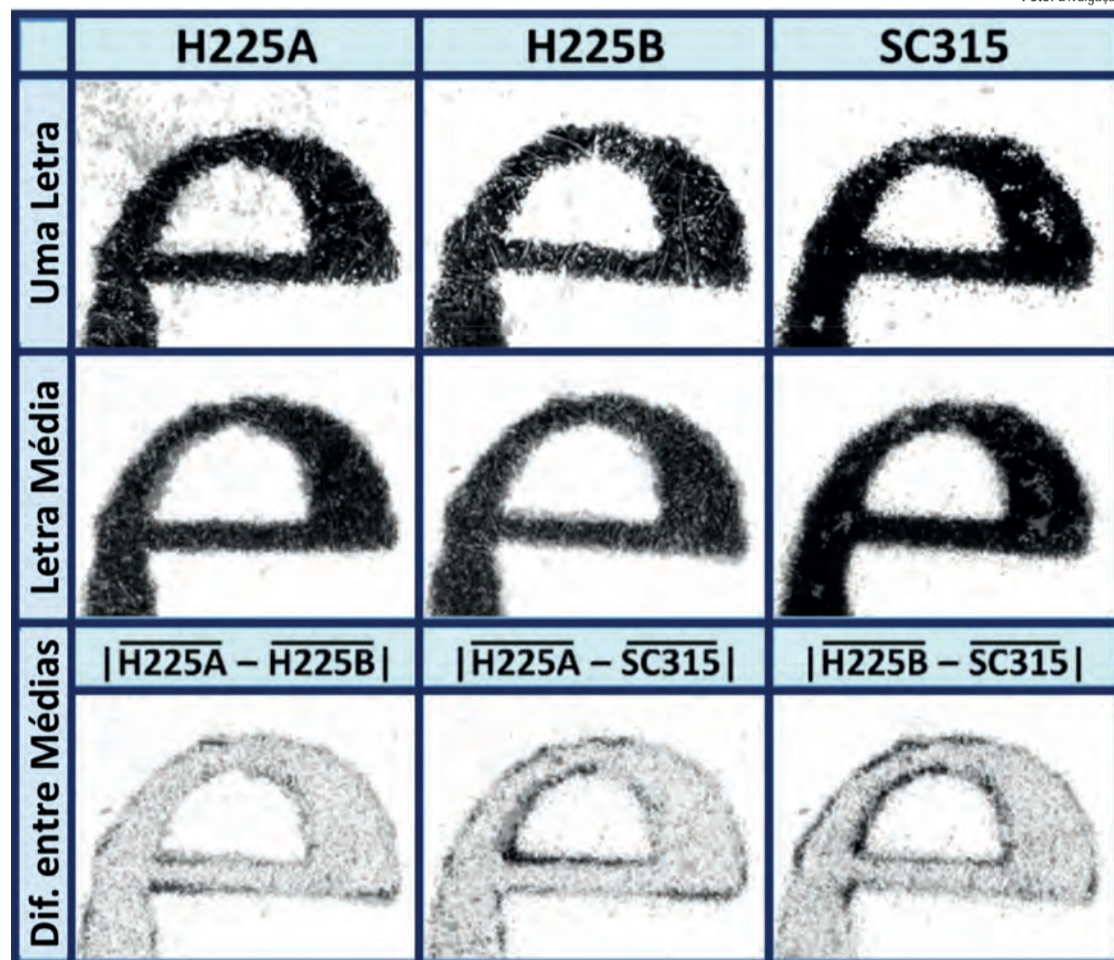
Os crimes envolvendo a adulteração de documentos digitais, observa Ferreira, têm aumentado significativamente em tempos recentes, acompanhando o crescimento das redes sociais, blogs e sites. “Está se tornando cada vez mais difícil constatar a autenticidade de fotografias e vídeos compartilhados nesses espaços, dado que as ferramentas de edição de imagens também estão cada vez mais acessíveis”, afirma. Comprovar se um documento é autêntico não constitui, porém, o único desafio para quem atua na área da computação forense.

De acordo com o autor da tese de doutorado, outra dificuldade é descobrir a origem dos documentos. “Identificar qual impressora gerou um contrato fraudado, por exemplo, também é tarefa das mais complexas”, diz. Normalmente, esse tipo de fraude tem sido investigado por meio de abordagens específicas. “Ainda nesse contexto, quando pensamos na identificação de falsificações em fotografias, normalmente as técnicas exploram um ou poucos aspectos ou artefatos de cada vez. Ou seja, muitas vezes uma técnica verifica um único ponto, como a iluminação da foto. Ocorre que esse tipo de exame não é tão efetivo, porque desconsidera vários outros elementos. O que a tese propõe é realizar uma multianálise do documento, promovendo a combinação de diferentes técnicas para checar distintos atributos do objeto investigado”, acrescenta Ferreira.

Dito de modo simplificado, o trabalho do pesquisador estabelece um conjunto de passos, vários deles originais, para promover a checagem tanto da autenticidade quanto da origem de documentos digitais. Assim, utilizando novamente o exemplo da fotografia, ao invés de investigar somente a iluminação, o modelo também consideraria outras propriedades da imagem, como cor, tom de pele das pessoas e eventuais inconsistências nas bordas, que são analisadas por algoritmos dotados da capacidade de aprender com esses dados de entrada.

Em seguida, outro algoritmo faz a fusão dessas análises e determina se o objeto em questão foi ou não alvo de alguma manipulação. “O cumprimento dessa rota é importante porque a fraude pode eventualmente driblar um algoritmo, mas dificilmente conseguirá driblar vários deles”, pontua o professor Rocha. Em relação à origem do documento, o procedimento é semelhante. No caso de um texto impresso, por exemplo, os pesquisadores avaliam o tipo e o corpo das letras, a área de mancha da impressão, o tom do toner etc, e fazem a comparação com textos gerados por impressoras suspeitas, de modo a encontrar diferenças ou semelhanças.

O orientador da tese observa que, para realizar esse tipo de análise, os documentos digitais passam por pré-processamentos. “Nós ampliamos ou reduzimos várias vezes uma imagem, para enfatizar determinadas propriedades da mesma. Ou a decomparamos em diferentes canais de informação. Ou, ainda, aplicamos determinadas perturbações. Como muitos fraudadores aplicam filtros nas imagens, na tentativa de dificultar a comprovação da falsificação, nós também filtramos a imagem diversas vezes e analisamos o comportamento desta após cada etapa de filtragem, que pode ser vista como um processo de perturbação. Se ela não foi filtrada anteriormente, o comportamento será um. Se foi, será outro”, pormenoriza.



Exemplo de detecção de fraude usado na pesquisa: checando a origem e a autenticidade de documentos

Uma analogia interessante em relação a esse processo seria com a compressão de textos. “Se temos um texto original nunca foi comprimido e o comprimimos pela primeira vez, ele diminui de tamanho; se ele já foi comprimido anteriormente, dificilmente diminuirá e, possivelmente, até aumentará de tamanho. O mesmo acontece na filtragem; se a imagem já foi filtrada e a filtramos novamente, certos elementos da mesma serão excluídos, mas, certamente, são elementos diferentes em relação àquela imagem que nunca foi filtrada e está sofrendo a primeira filtragem”, reforça o docente do IC.

Outro aspecto importante, conforme o autor da tese, é que o método proposto pode ser expandido e agregar outras técnicas de análise. Assim, também pode ser aplicado, em tese, a outros tipos de documentos, como vídeos e áudio. “Existe essa possibilidade. Em relação a um arquivo de áudio, o principal objetivo de uma investigação é identificar quem é o locutor. Nesse caso, o princípio seria o mesmo, ou seja, utilizar dois ou três algoritmos para analisar diferentes propriedades da voz, para depois confrontá-las com os atributos da voz do suspeito. Em relação a um vídeo, a probabilidade é ainda maior, visto que este nada mais é que um conjunto de fotos em movimento. Nesse caso, seria possível analisar cada quadro separadamente e depois combinar os dados obtidos”, infere

Ferreira, que cumpriu parte do doutorado na Universidade Politécnica de Milão, Itália, e contou com bolsa sanduíche concedida pelo Conselho de Desenvolvimento Científico e Tecnológico (CNPq).

O professor Rocha reconhece que, a despeito dos avanços registrados, os desafios da computação forense são grandes, principalmente por causa da habilidade demonstrada pelos falsificadores, que estão sempre desenvolvendo novas maneiras de fraudar documentos digitais. “Eu costumo dizer que participamos de uma disputa de gato e rato com eles. Se nós criamos um bom algoritmo para identificar fraudes, eles desenvolvem algum método para driblar essa identificação. Daí a grande vantagem do modelo proposto pela tese, que refina bastante a análise. Imagine um sistema biométrico que identifica a digital do dedo polegar. Agora, imagine um sistema que identifica, além do polegar, também o indicador e o anelar. Mais: imagine se adicionarmos ao sistema uma técnica para identificar, adicionalmente, a face. Isso dificultaria muito a fraude, e é nessa direção que estamos caminhando”.

Questionado se o método poderia, por exemplo, ajudar a identificar falsificação de obras de arte, o docente do IC afirma que há essa possibilidade, com as devidas restrições a serem respeitadas para esse novo problema. “Recentemente, outro orientando meu

teve um trabalho aceito para apresentação em um evento científico nos Estados Unidos. Ele desenvolveu um algoritmo para aprender os padrões típicos presentes nas pinturas de Van Gogh, cuja técnica é personalíssima. Nos testes que fizemos, com o uso de fotografias das obras em alta resolução, conseguimos apontar com mais de 90% de precisão se determinado quadro foi pintado por ele. Agora, outro aluno dará continuidade à pesquisa, estendendo a análise a outros artistas e refinando os resultados obtidos com essa análise preliminar”, adianta Rocha.

Na opinião dos pesquisadores, o Brasil tem adotado boas práticas para estimular a pesquisa no âmbito da investigação forense. Em 2014, por exemplo, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), agência de fomento vinculada ao Ministério da Educação, lançou um edital por meio do qual foram selecionados 15 projetos forenses nas áreas de computação, biologia, química etc. Os objetivos são desenvolver estudos originais e formar recursos humanos qualificados para auxiliar a Polícia Federal e outras entidades nas suas investigações. Esses projetos temáticos estão divididos em diversas instituições do país e o IC/Unicamp conta com dois deles, um em Multimídia Forense, coordenado pelo professor Rocha, e outro em Informática Forense, coordenado pelo professor Ricardo Dahab.

Medida semelhante tem sido tomada pelos Estados Unidos. A Defense Advanced Research Projects Agency (Darpa) selecionou recentemente dez projetos de pesquisas para financiamento, dentro de um esforço mundial para o desenvolvimento de soluções que contribuam para o processo de combate às fraudes de documentos digitais. A equipe coordenada pelo professor Rocha participa, de forma colaborativa, de um desses projetos juntamente com as universidades de Notre Dame, Purdue, NYU e UCSC, nos EUA, e Politécnica de Milão e Universidade de Siena, na Itália.

Retornando ao Brasil, o docente do IC considera que a promulgação do marco civil da internet, em 2014, poderá facilitar o uso das metodologias científicas para subsidiar eventuais decisões judiciais. “O grupo da Unicamp é o mais ativo no país na área da computação forense. Temos desenvolvido estudos importantes e mantido relações de cooperação com destacados grupos estrangeiros. Um dos nossos objetivos é implantar na Universidade um Laboratório Multidisciplinar de Análise Forense, cujo projeto tem sido discutido desde 2012”, relata Rocha.

Publicação

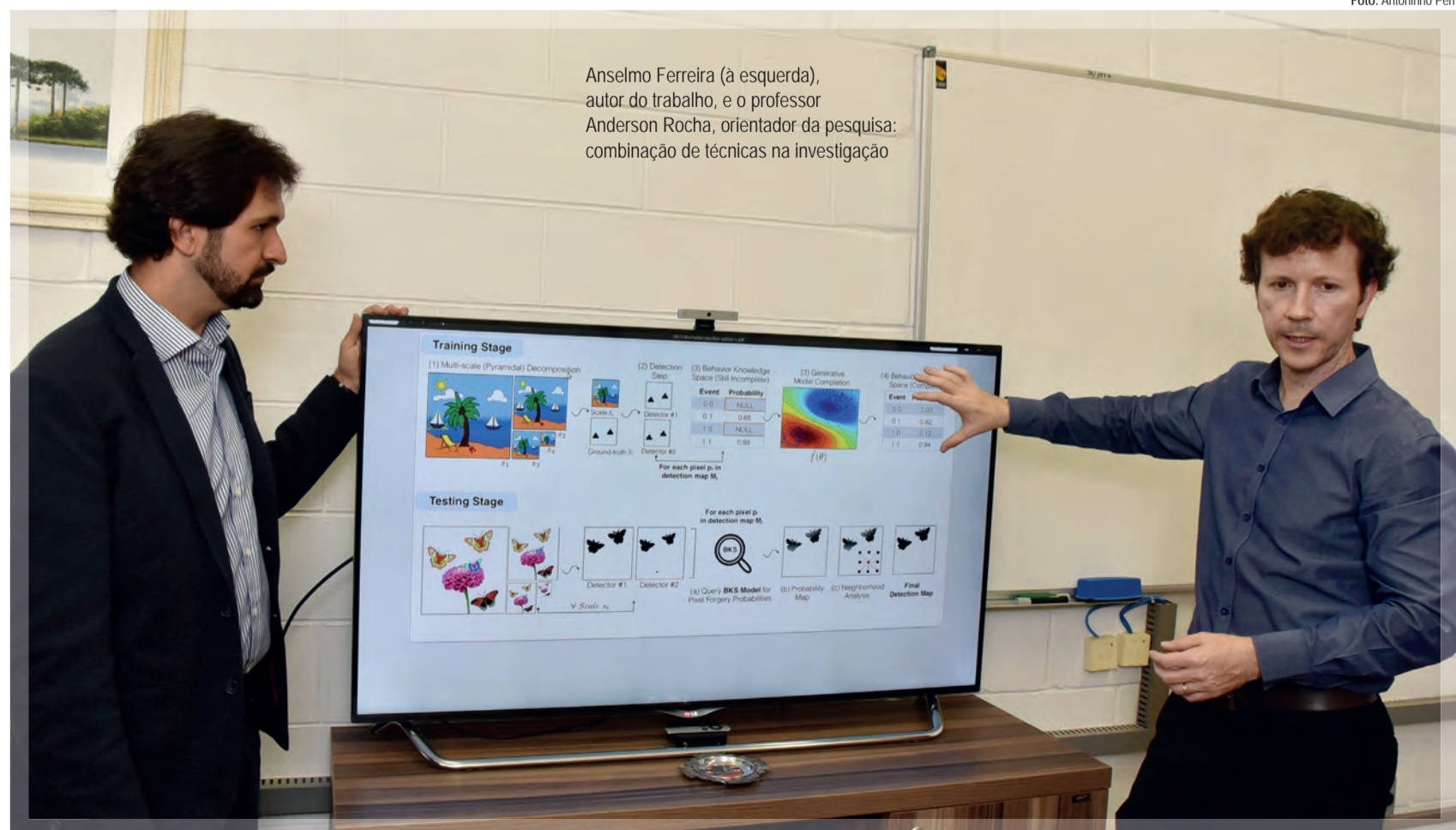
Tese: “Técnicas de multi-análise para investigação forense de documentos digitais”

Autor: Anselmo Castelo Branco Ferreira

Orientador: Anderson de Rezende Rocha

Coorientador: Jefersson Alex dos Santos

Unidade: Instituto de Computação (IC)



Anselmo Ferreira (à esquerda), autor do trabalho, e o professor Anderson Rocha, orientador da pesquisa: combinação de técnicas na investigação

Foto: Antoninho Perri