MO815 – Análise Forense de Documentos Digitais – Fichamento #3

Exposing Digital Forgeries Through Specular Highlights on the Eye [3]

by Micah K. Johnson and Hany Farid

29 de Outubro de 2010

Revisores: Priscila Sabóia — RA 077555
Tiago Carvalho — RA 087343

1 Visão global

Neste artigo é apresentada uma nova técnica para detecção de *splicing* em imagens. Basicamente, os autores detectam se pessoas em uma mesma imagem foram fotografadas sobre diferentes condições de iluminação, com base nos raios de luz refletidos pelos olhos. Eles apresentam resultados da técnica para imagens de entrada sintéticas e reais, com e sem a presença de *splicing*.

2 Resumo

Os autores apresentam uma nova técnica para identificar *splicing* de pessoas através dos raios de luz refletidos nos olhos. Na primeira parte do artigo é apresentado um método para estimar a direção em que a luz é percebida pela câmera, a normal à superfície do olho e a direção da fonte de luz, todas em relação ao próprio olho. Depois, a posição de uma única fonte de luz é estimada levando em conta as posições de brilhos nos olhos e suas respectivas direções de luz estimadas. Com base nesta única fonte de luz, é calculada a média e o desvio padrão dos erros angulares para realizar um teste de hipóteses, onde é decidido se as pessoas na imagem foram ou não fotografadas sobre diferentes condições de iluminação.

O método para estimar a direção da fonte de luz em relação a cada olho se baseia na reflexão especular da luz, que é a reflexão da luz de uma superfície, neste caso o olho, onde o raio da reflexão incidente se reflete em um ângulo θ_r igual ao ângulo de incidência θ_i. Tais ângulos são medidos em respeito ao vetor normal à superfície no olho. Então, a direção da luz L, pode ser descrita em termos da direção do raio refletido R e da normal à superfície N. Eles assumem que o olho é um reflector perfeito, ou seja, a direção V em que a luz é percebida (direção da câmera) é igual à direção de reflexão R. Logo, é possível obter a direção da luz da seguinte forma

$$L = 2(V^T N)N - V. \tag{1}$$

Para estimar a normal à superfície N e a direção da câmera V em um sistema de coordenadas comum, primeiro é necessário estimar a transformação projetiva H que descreve a transformação de coordenadas

do mundo em coordenadas da imagem. Para estimar esta calibração com somente uma única imagem, eles utilizam o conhecimento de características geométricas dos olhos. No caso o limbus, região que separa a esclera (parte branca do olho) da íris (parte colorida), é modelado como um círculo.

No entanto, a imagem do limbus será uma elipse na imagem capturada pela câmera, exceto quando os olhos estão exatamente de frente para a câmera. Intuitivamente a distorção de uma elipse para um círculo está relacionada com a pose e posição dos olhos em relação à câmera. Desta forma eles procuram a transformação que alinha a imagem do limbus à imagem de um círculo. Além disso assumem que todos os pontos do limbus são coplanares e pertencem ao plano $Z = 0$. Por isso, H se reduz à uma matriz de transformação projetiva planar com dimensão 3x3, onde pontos do mundo X e pontos da imagem x são representados por vetores homogêneos $2D$. Pontos do limbus no sistema de coordenadas do mundo devem satisfazer a equação implícita do círculo, $f(X, \alpha) = (X_1 - C_1)^2 + (X_2 - C_2)^2 - r^2 = 0$, onde o vetor $\alpha = (C_1\ C_2\ r)^T$ representa o centro e o raio do círculo.

Considerando que os pontos X_i, $i = 1, ..., m$, satisfazem a equação do círculo, e um modelo de câmera pinhole ideal, pontos no mundo X_i são mapeados para pontos na imagem x_i da seguinte forma: $x_i = HX_i$. Sendo assim, H e α são estimados utilizado a seguinte função de erro nestes dois parâmetros,

$$E(\alpha, H) = \sum_{i=1}^{m} \min_{X^*} \|x_i - HX_i^*\|^2, \tag{2}$$

onde $\hat{X}$ está no círculo parametrizado por α. Esta função de erro é minimizada por mínimos quadrados não lineares utilizando iteração de Levenberg-Marquadt. Segundo os autores, para que H seja estimada de forma única, a função de erro deve incorporar os dois olhos.

Depois de estimar H e α, eles decompõem H em termos dos parâmetros intrínsecos e extrínsecos, com o objetivo de obter a matriz $\hat{H}$, que transforma coordenadas do mundo em coordenadas da câmera, e a matriz de rotação R.

Os parâmetros intrínsecos da câmera são a distância focal f, o centro da câmera, a distorção geométrica introduzida pelo sistema ótico e a proporção do pixel (relação largura/comprimento). Eles assumem que o centro da câmera é o centro da imagem, não há distorção radial, e a proporção do pixel é 1, restando conhecer apenas a distância focal f. Os parâmetros extrínsecos são a matriz de rotação R e o vetor de translação t. Uma vez que os pontos do limbus são coplanares, a decomposição de H é dada por:

$$H = \lambda K \begin{pmatrix} r_1 & r_2 & t \end{pmatrix}, \tag{3}$$

onde λ é um fator de escala, os vetores colunas r_1 e r_2 são a primeira e a segunda coluna da matriz de rotação R, t é o vetor de translação e K é a seguinte matriz 3x3:

$$K = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

Com a distância focal f conhecida é possível estimar $\hat{H}$ diretamente:

$$\frac{1}{\lambda} K^1 H = \begin{pmatrix} r_1 & r_2 & t \end{pmatrix}, \tag{5}$$

$$\hat{H} = \begin{pmatrix} r_1 & r_2 & t \end{pmatrix} = \frac{1}{\lambda} K^1 H, \tag{6}$$

onde λ é escolhido de forma que r_1 e r_2 sejam unitários. A partir disto, a matriz de rotação R completa é dada por:

$$R = \begin{pmatrix} r_1 & r_2 & r_1 \times r_2 \end{pmatrix}. \tag{7}$$

No caso em que a distância focal é desconhecida, primeiro deve-se estimar f para depois obter $\hat{H}$. Para estimar f decompõe-se $\hat{H}$ da seguinte forma:

$$H = \lambda \begin{pmatrix} fc_y c_z & fc_y s_z & ft_x \\ f(s_x s_y c_z - c_x s_z) & f(s_x s_y s_z + c_x c_z) & ft_y \\ c_x s_y c_z + s_x s_z & c_x s_y s_z - s_x c_z & t_z \end{pmatrix}, \tag{8}$$

onde $c_x = \cos(\theta_x)$ e $s_x = \sin(\theta_x)$, $c_y = \cos(\theta_y)$ e $s_y = \sin(\theta_y)$, e $c_z = \cos(\theta_z)$ e $s_z = \sin(\theta_z)$.

De acordo com a matriz, existem oito parâmetros desconhecidos a serem estimados: a distância focal; o fator de escala λ; os ângulos de rotação θ_x, θ_y e θ_z, referentes à matriz R; e três coordenadas do vetor de translação t. No entanto, os autortes consideram a sub matriz 2x2 da matriz H, à esquerda e à cima, diminuindo o número de termos desconhecidos para 4: θ_x, θ_y, θ_z, e $\hat{f} = \lambda f$. Estes são estimados minimizando a seguinte função de erro usando mínimos quadrados não lineares:

$$E(\theta_x\, \theta_y\, \theta_z,\, \hat{f}) = (\hat{f} c_y c_z - h_1)^2 + (\hat{f} c_y s_z - h2)^2 + (\hat{f}(s_x s_y c_z - c_x s_z) - h4)^2 + (\hat{f}(s_x s_y s_z + c_x c_z) - h5)^2, \tag{9}$$

onde h_i corresponde à i-ésima entrada da matriz H estimada. Eles aplicam o método de Gauss-Newton para minimizar $E(\cdot)$. Com base nos parâmetros estimados encontra-se duas possíveis estimativas para o valor da distância focal f:

$$f_1 = \frac{\hat{f}(c_x s_y c_z + s_x s_z)}{h_7} \; e \; f_2 = \frac{\hat{f}(c_x s_y s_z - s_x c_z)}{h_8}, \tag{10}$$

Sendo que distância focal f é obtida com a média ponderada de f_1 e f_2:

$$f = \frac{h_7^2 f_1 + h_8^2 f_2}{h_7^2 + h_8^2}. \tag{11}$$

Na etapa de calibração de câmera foi estimado o centro do círculo ao se estimar $\alpha = (C_1\ C_2\ r)$. Sendo assim, o centro do limbus no sistema de coordenadas do mundo é dado por $X_c = (C_1\ C_2\ 1)^T$. Logo no sistema de coordenadas da câmera o centro passa a ser $x_c = \hat{H} X_c$. A direção da câmera, vetor que parte do centro do limbus para a origem do sistema de coordenadas da câmera é

$$v = -\frac{x_c}{\|x_c\|}. \tag{12}$$

O vetor normal à superfície do olho N é estimado a partir de um modelo 3D do olho. Tal modelo consiste de duas esferas com os centros localizados no mesmo eixo Z, só que distantes $d = 4,7\, mm$. A esfera maior representa a esclera e tem raio $r_1 = 11,5\, mm$. Já a esfera menor representa a córnea e tem raio $r_2 = 7,8\, mm$. O limbus é definido como um círculo de raio $p = 5,8mm$, o qual é resultante da intersecção das duas esferas. A distância entre o centro da esfera menor e o plano contendo o limbus é igual a $q = 5,25\, mm$.

Considerando o brilho especular na posição $S = (S_x\ S_y)$, medido em relação ao centro do limbus em tal modelo, a normal à superfície é determinada pela intersecção do raio que deixa S ao longo da direção V com a borda da esfera. Esta interseção pode ser computada resolvendo um sistema quadrático para k, que representa a distância entre S e o ponto de intersecção:

$$k^2 + 2(S_x V_x S_y V_y + q k V_z)k + (S_x^2 + S_y^2 + q^2 - r_2^2) = 0. \tag{13}$$

onde q e r_2 são especificados pelo modelo 3D dos olhos e a direção do observador $V = (V_x\ V_y\ V_z)^T$ no sistema de coordenadas do mundo é dado por $V = R^{-1} v$, onde v é a direção do observador no sistema de coordenadas da câmera e R é a matriz de rotação calculada no passo de calibração de câmera.

Para completar é necessário conhecer S no sistema de coordenadas do mundo a partir do sistema de coordenadas da câmera. Sendo x_s o brilho especular nas coordenadas da imagem, a transformação inversa H^{-1} mapeia as coordenadas do ponto de brilho especular nas coordenadas do mundo. Desta forma, o ponto de brilho no sistema de coordenadas do mundo é $X_s = H^{-1}x_s$. Sendo o centro C e o raio r do limbus no sistema de coordenadas do mundo, as coordenadas do brilho especular S dentro do modelo do olho 3D é $S = \frac{q}{r}(X_s - C)$. Então, conhecendo k, V e S é possível obter o vetor normal à superfície, dado por:

$$N = \begin{pmatrix} S_x + kV_x \\ S_y + kV_y \\ q + kV_z \end{pmatrix}. \tag{14}$$

Com V e N estimados, agora basta aplicar a Equação 1 que os relaciona com o vetor L para encontrar a direção da luz estimada. De forma a comparar as estimativas através da imagem, a direção da fonte de luz é convertida para as coordenadas da câmera, $l = RL$.

Depois de estimar a direção da fonte de luz para cada pessoa da foto, a técnica verifica se todos elas partiram de uma mesma fonte de luz. Para cada um dos pontos de brilho especular p_i, o ângulo entre a direção do vetor que parte de p_i para a ponto de uma fonte de luz na posição x e a direção estimada l_i é igual a

$$\theta_i(x) = \arccos\left(l_i^T \frac{x - p_i}{||x - p_i||}\right), \tag{15}$$

onde p_i é a posição do i-ésimo brilho especular. Dadas M direções de luz estimadas l_i e as respectivas posições de brilho especular p_i, a posição da fonte de luz pode ser estimada maximizando, através do método não-linear do gradiente conjugado, a função de erro dada por

$$E(x^*) = \sum_{i=1}^{M} \left(l_i^T \frac{x^* - p_i}{||x^* - p_i||}\right). \tag{16}$$

Assim, x^* denota a posição da fonte de luz estimada. Com este valor é possível calcular o erro angular $\theta_i(x^*)$ entre a direção estimada da luz l_i e a direção do vetor que parte do brilho especular e alcança a posição da luz estimada x^*. Depois calcula-se a média μ e o desvio padrão σ dos erros angulares.

Para decidir se uma imagem possui ou não *splicing*, os autores utilizam um método estatístico denominado "Teste de Hipóteses", com a seguinte equação:

$$z = \frac{\mu - \mu^0}{\sigma^0/\sqrt{m}}, \tag{17}$$

onde μ^0 e σ^0 representam a média e o desvio padrão conhecidos para imagens normais, sem *splicing*. Se a significância do Teste de Hipóteses for menor que um determinado nível escolhido (e.g. 1%), então o erro angular médio dos raios de luz é maior que o esperado e considerado inconsistente. Caso contrário a estimativa não pode ser considerada inconsistente. Por fim a significância do Teste de Hipóteses é dada em termos da função de erro padrão

$$p(z) = \frac{1}{2}\left(1 - erf\left(\frac{z}{\sqrt{2}}\right)\right). \tag{18}$$

A técnica foi testada para imgens sintéticas e reais. As imagens sintéticas foram geradas utilizando o software "pbrt". Cada imagem com 1200x1600 pixels, onde a córnea ocupou menos de 0.01% da imagem. Foram testadas 12 posições diferentes para os olhos, que foram iluminados com duas fontes de luz: uma fonte fixa alinhada com a câmera e uma fonte colocada em uma de quatro posições diferentes.

Neste caso, a posição do limbus e dos raios de luz especulares foram extraídos automaticamente. Nos testes onde a distância focal era conhecida, o erro angular médio foi de 2.8° com desvio padrão de 1.3° e erro máximo de 6.8°. Já com distância focal desconhecida o erro angular médio foi de 2.8° com desvio padrão de 1.3° e erro máximo de 6.3°.

Em um segundo cenário de teste, um homem é fotografado em um ambiente de iluminação controlada. Sua posição e as posições das fontes de luz foram ajustadas manualmente e a posição do limbus e dos brilhos especulares foram manualmente selecionados. Nos testes onde a distância focal era conhecida, o erro angular médio foi de 8.6° e com a distância desconhecida a média foi de 10.5°.

Para 20 imagens retiradas do Flickr, elipses foram ajustadas manualmente ao redor do limbus. Os raios especulares foram localizados selecionando um limite retangular ao redor dos raios de luz e computando o centróide da seleção. No total foram estimadas 88 direções de luz (44 pessoas). O erro angular médio foi de 6.4°, com desvio padrão de 2.8° e erro máximo de 12.8°. De forma geral os erros médios das falsificações sempre foram maiores que os erros das imagens autênticas.

3 Contribuições

A principal contribuição do autor é propor uma técnica para identificar *splicing* compostas por pessoas fotografadas em cenários diferentes. Além disso, dado que o método estima a direção da fonte de luz em um sistema de coordenadas 3D do mundo, ele é mais robusto em relação aos métodos que funcionam estimando a direção da fonte de luz em 2D que apresentam um grau de ambiguidade.

4 Defeitos/Desvantagens

As principal desvantagem do artigo está na validação fraca dos resultados, uma vez que ele só apresenta testes realizados com 4 falsificações como resultado. Uma limitação do método na área de detecção de *splicing* é que este só é aplicado quando a composição leva em conta pessoas, especificamente fotografadas em ambientes fechados (a troca do fundo da cena por exemplo não seria detectada). Além disso, o método necessita que o raio de luz esteja claramente distinguível na imagem, ou seja, em imagens de baixa resolução a aplicação do método pode não ser recomendada.

5 Trabalhos correlatos

5.1 Exposing Digital Forgeries by Detecting Inconsistencies in Lighting. [1]

Relação com o artigo avaliado: Método para identificação de *splicing* baseado na estimação da direção da fonte de iluminação através de uma única imagem.

Descrição: quando uma imagem é fabricada através da composição de duas ou mais imagens, é relativamente complicado realizar o casamento da iluminação das imagens utilizadas para realizar a composição. É possível se detectar falsificações em uma imagem determinando a direção da iluminação de uma região marcada na imagem. Através de modelos compostos para diferentes tipos de iluminação, os autores estimam a iluminação das regiões indicadas. Quando o ângulo entre as posições de iluminação estimadas para partes diferentes da imagem excede um limiar determinado, essa imagem é dita como sendo uma falsificação.

Os pontos fracos do método provêm de uma grande dependência do conhecimento do usuário, bem como da determinação de um *threshold* correto entre imagens que possuem ou não composição.

5.2 Exposing Digital Forgeries in Complex Lighting Environments. [2]

Relação com o artigo avaliado: Assim como o artigo principal, identifica splicing em imagens utilizando características de iluminação, no entanto, é projetado para ambientes de iluminação complexa.

Descrição: Os autores utilizam a modelagem da irradiância da imagem através de harmônicos esféricos. Uma vez marcadas diferentes regiões da imagem, o método calcula a representação da iluminação de cada uma das regiões e compara essas representações entre si. Como principal desvantagem, é mostrada a grande iteração com o usuário, deixando o método dependente de um conhecimento prévio de sua parte.

6 Extensões

Uma possível extensão para o artigo seria a substituição do método de teste de hipótese por um método de aprendizado de máquina. Isso poderia ajudar a determinar melhor o acerto do método, além de uma validação mais exaustiva do método, o que daria uma maior credibilidade ao mesmo.

7 Notas

Relevância: 8 / Originalidade: 9.0 / Qualidade científica: 9.0 / Apresentação: 7.0 / Nota final: 8.25

Referências

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

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

[3] M.K. Johnson and H. Farid. Exposing digital forgeries through specular highlights on the eye. In *9th International Workshop on Information Hiding*, Saint Malo, France, 2007.