Revista Eletrônica Paulista de Matemática

ISSN 2316-9664
Volume 10, dez. 2017
Edição Ermac

# Detecção de *spoofing* facial: uma abordagem baseada nas máquinas de Boltzmann restritas

Face spoofing detection: an approach based on the restricted Boltzmann machines

**Gustavo Botelho de Souza**
UFSCar - Universidade Federal de São Carlos
gustavo.botelho@gmail.com

**Aparecido Nilceu Marana**
UNESP - Universidade Estadual Paulista "Júlio de Mesquita Filho"
nilceu@fc.unesp.br

**João Paulo Papa**
UNESP - Universidade Estadual Paulista "Júlio de Mesquita Filho"
papa@fc.unesp.br

**Resumo**

A identificação de pessoas tem um papel essencial em nossa sociedade. Nos últimos anos, a Biometria vem se configurando como opção robusta e conveniente para este fim. Apesar da maior praticidade do reconhecimento facial, sistemas de reconhecimento pela face são os primeiros a sofrer com ataques de apresentação de características sintéticas (fotografias) por meliantes (ataques de *spoofing*). Neste sentido, métodos capazes de detectar automaticamente se a face capturada pela câmera do sistema biométrico é real ou artificial se tornam indispensáveis. Neste trabalho propõe-se uma nova abordagem baseada no descritor LBP (*local binary patterns*) e nas máquinas de Boltzmann restritas para a extração das características de textura mais relevantes das faces apresentadas a fim de detectar ataques de *spoofing* facial com maior acurácia. Resultados obtidos sobre a base de imagens NUAA indicam que o método proposto apresenta boas taxas de acerto, mesmo em casos de pouca variabilidade interclasse, como na base avaliada.

**Palavras-chave:** Reconhecimento facial. Detecção de *spoofing*. Máquinas de Boltzmann restritas. Campos aletórios de Markov. Métodos estocásticos e estatísticos.

**Abstract**

In the last years, Biometrics emerged as a robust and convenient solution for people identification. Despite the practicality of face recognition, identification systems based on face are the first to suffer with presentation of synthetic traits (facial photographs) by criminals (spoofing attacks). In this sense, methods able to automatically detect whether the face captured by the camera of the biometric system is real or fake become essential. In this work we propose a new approach based on the LBP (local binary patterns) descriptor and on the restricted Boltzmann machines for the extraction of the most relevant texture features from facial images in order to detect spoofing attacks with greater accuracy. Results obtained on the NUAA dataset indicate that the proposed method presents good accuracy rate, even in case of low interclass variability, as in the evaluated database.

**Keywords:** Face recognition. Spoofing detection. Restricted Boltzmann machines. Markov random fields. Stochastic and statistical methods.

# 1 Introdução

A identificação de pessoas desempenha um papel importante em nossa sociedade. Nos últimos anos, a Biometria, isto é, o reconhecimento automatizado de indivíduos por meio de suas características físicas, fisiológicas ou comportamentais (tais como face, íris, impressão digital, termograma facial, forma de andar, de digitar, dentre outras) vem se configurando como opção de segurança robusta e conveniente para tal fim (JAIN; ROSS; NANDAKUMAR, 2011).

Quando os primeiros sistemas biométricos foram propostos, acreditava-se que, por reconhecerem as pessoas por "algo que elas são", as fraudes seriam bastante dificultadas: para burlar tais sistemas seria necessário simular características biológicas de outrem (JAIN et al., 2004). Entretanto, apesar da certa dificuldade em fraudar tais mecanismos de segurança em comparação com os sistemas baseados em senhas e cartões, hoje, dada sua disseminação pela sociedade, criminosos já desenvolveram mecanismos de ataque capazes de driblar os sensores de captura das características biométricas, simulando traços de usuários válidos, técnica conhecida como *spoofing* (MENOTTI et al., 2015; SILVA; MARANA; PAULINO, 2015). Frente a esta realidade, torna-se necessário o desenvolvimento de métodos preventivos, isto é, de contramedida.

Dentre as características biométricas, a face se configura como uma opção bastante conveniente e natural aos usuários dos sistemas de identificação, dada sua extração não intrusiva, rápida e sem a necessidade de muita colaboração da pessoa sendo identificada. Além disto, atualmente, imagens de boa qualidade são facilmente registradas, tendo em vista as câmeras disponíveis, inclusive em dispositivos móveis. Apesar de toda esta praticidade, os sistemas de reconhecimento facial são os primeiros a sofrer com tentativas de fraude visto que meliantes, nos dias atuais, podem facialmente obter imagens faciais de outros indivíduos na rede mundial, em resolução considerável, e apresentá-las aos sistemas de reconhecimento. Desta maneira, para que os sistemas de reconhecimento facial continuem se configurando como boa alternativa, enquanto soluções de segurança, técnicas anti-*spoofing* precisam ser agregadas.

A maioria dos métodos de detecção de *spoofing* facial trabalha com descritores tradicionais, extraindo características referenciadas como *handcrafted*, isto é, pré-definidas na formulação da técnica (TAN et al., 2010). Entretanto, resultados recentes têm mostrado que os métodos baseados em redes neurais, que trabalham com características autoaprendidas a partir dos dados de treinamento, vêm superando as técnicas estado-da-arte em muitas tarefas complexas, em especial por permitirem a identificação e extração das características mais relevantes a cada problema (HINTON, 2002).

Neste trabalho, um novo método para deteção de *spoofing* em sistemas de reconhecimento facial é proposto, valendo-se do descritor *local binary patterns* (LBP) (OJALA; PIETIKÄINEN; HARWOOD, 1996) e das máquinas de Boltzmann restritas (HINTON, 2002), redes neurais estocásticas baseadas em energia, para o aprendizado e extração das características de textura mais relevantes das imagens faciais a fim de classificá-las em reais ou sintéticas com maior acurácia. Resultados na base NUAA indicam que a técnica proposta apresenta desempenho superior à do método proposto pelos próprios autores da base.

# 2 *Local binary patterns* (LBP)

Por ser simples e ao mesmo tempo muito eficaz, o descritor de textura *local binary patterns* (LBP) (OJALA; PIETIKÄINEN; HARWOOD, 1996) tem atraído a atenção de inúmeros pesquisadores que lidam com a análise de imagens. Em geral, neste descritor, um valor inteiro é

associado a cada pixel da imagem com base em comparações efetuadas entre seu tom de cinza original e os tons de seus pixels vizinhos. Na análise de um pixel $p$ da imagem, compara-se seu tom de cinza com o tom de cada vizinho $q$. Caso o tom de $q$ seja maior ou igual ao do $p$, associa-se o *label* "1" a $q$, caso contrário "0".

Definido um raio de vizinhança $R$ e um número de vizinhos $P$, após efetuar as comparações e percorrer os pixels vizinhos de $p$ em sentido horário a partir de seu vizinho superior esquerdo, pode-se construir um número binário com base nos *labels* associados a tais pixels. Este número é então convertido para decimal e associa-se o valor ao pixel central $p$ sob análise. Em termos matemáticos o valor LBP de um pixel $p$ é dado por:

$$LBP_{P,R} = \sum_{q=0}^{P-1} l(c_q - c_p) \cdot 2^q, \tag{1}$$

onde $c_p$ indica o tom de cinza de $p$, $c_q$ o tom de cinza do vizinho $q$ e $l(x)$ corresponde à função de limiarização definida por:

$$l(x) = \begin{cases} 1, & se\ x \geq 0, \\ 0, & se\ x < 0. \end{cases} \tag{2}$$

Em geral, após efetuar tal procedimento para todos os pixels de uma dada imagem, um histograma é construído com a frequência da ocorrência de cada decimal possível a fim de caracterizá-la. A Figura 1 ilustra o processo de cálculo do decimal associado a um pixel $p$ considerando-se $R = 1$ e $P = 8$ (vizinhança $3 \times 3$). Após associar um *label* a cada vizinho dada a comparação de seu tom de cinza com o tom de cinza do pixel central, encontra-se o valor binário para o pixel central ao percorrer seus vizinhos em sentido horário. Tal valor é então convertido para decimal.

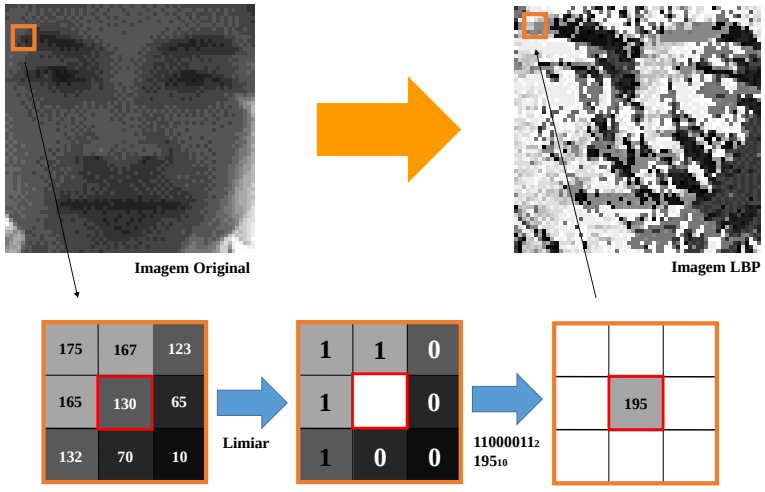


Figura 1: Cálculo do valor LBP para um pixel central sob análise em uma dada imagem facial.

# 3 Máquinas de Boltzmann restritas (RBM)

As máquinas de Boltzmann (BM, do inglês, *Boltzmann machines*) (HINTON; SEJNOWSKI, 1983) correspondem a redes de unidades de processamento estocásticas, similares aos modelos de redes neurais conhecidos. Elas podem ser usadas no aprendizado de importantes aspectos de distribuições de probabilidade, a partir de amostras destas distribuições (ACKLEY; HINTON; SEJNOWSKI, 1985).

Por permitirem a existência de ligações entre todos os nós da rede, as BMs apresentam algoritmo de aprendizado complexo e que consome muito tempo. Entretanto, o aprendizado pode ser bastante facilitado ao se impor restrições em sua topologia, dando-se origem assim às chamadas máquinas de Boltzmann restritas (*restricted Boltzmann machines* - RBM) (RUMELHART; MCCLELLAND, 1986; HINTON, 2002).

Uma RBM (HINTON, 2002) corresponde a um campo aleatório de Markov (MRF, do inglês, *Markov random field*) associado a um grafo não-direcionado bipartido. As RBMs podem ser vistas também como redes neurais estocásticas. Os vértices (neurônios) se dispõem em duas camadas, uma visível e outra escondida, com arestas (sinapses) apenas entre vértices de camadas diferentes. Como as BMs, estas máquinas também podem ser utilizadas no aprendizado de aspectos de distribuições de probabilidade dadas amostras destas distribuições (TANG; SALAKHUTDINOV; HINTON, 2012).

Os neurônios da camada visível são responsáveis pela observação, isto é, captura de informações do ambiente: por exemplo, pode-se ter um neurônio associado a cada pixel da imagem sob análise, isto é, extraindo características do pixel. Já na camada escondida, os neurônios modelam dependências e relações entre os vértices da primeira camada (por exemplo, dependências entre pixels da imagem sob análise). A Figura 2 ilustra a arquitetura de uma RBM. Pode-se observar as duas camadas (visível e escondida) e as ligações intercamadas apenas.

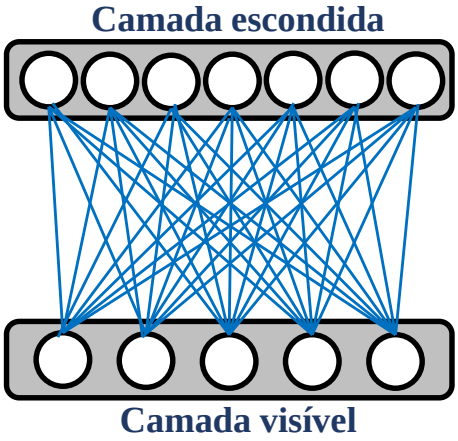


Figura 2: Exemplo da arquitetura de uma RBM.

Assumindo-se os neurônios da camada visível $\mathbf{v}$ e escondida $\mathbf{h}$ como unidades binárias, isto é, $\mathbf{v} \in \{0,1\}^m$ e $\mathbf{h} \in \{0,1\}^n$, onde $m$ e $n$ indicam a quantidade de nós em cada camada, tem-se a tradicional BB-RBM (*Bernoulli-Bernoulli restricted Boltzmann machine*), uma vez que os neurônios seguem a distribuição de Bernoulli. A função de energia de uma BB-RBM é dada por:

$$E(\mathbf{v},\mathbf{h}) = -\sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m}\sum_{j=1}^{n} v_i h_j w_{ij}, \tag{3}$$

onde $\mathbf{a}$ e $\mathbf{b}$ correspondem aos *biases* da camada visível e escondida, respectivamente, e $w_{ij}$ corresponde ao peso da conexão entre os neurônios $i$ da camada visível e $j$ da camada escondida.

A probabilidade da rede se encontrar em uma configuração $(\mathbf{v},\mathbf{h})$ é dada por:

$$P(\mathbf{v},\mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v},\mathbf{h})}, \tag{4}$$

onde $Z$ corresponde à função de partição, isto é, um fator de normalização calculado com base em todas as configurações possíveis das unidades da camada visível e escondida. De maneira similar, a probabilidade marginal de uma configuração da camada visível é dada por:

$$P(\mathbf{v}) = \frac{1}{Z}\sum_{\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})}. \tag{5}$$

Uma vez que uma BB-RBM é um grafo bipartido, as ativações dos neurônios da camada visível e dos neurônios da camada escondida são mutuamente independentes. Obtém-se assim as seguintes probabilidades condicionais:

$$P(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^{m} P(v_i|\mathbf{h}) \tag{6}$$

e

$$P(\mathbf{h}|\mathbf{v}) = \prod_{j=1}^{n} P(h_j|\mathbf{v}), \tag{7}$$

onde:

$$P(v_i = 1|\mathbf{h}) = \phi\left(\sum_{j=1}^{n} w_{ij} h_j + a_i\right) \tag{8}$$

e

$$P(h_j = 1|\mathbf{v}) = \phi\left(\sum_{i=1}^{m} w_{ij} v_i + b_j\right), \tag{9}$$

onde $\phi(\cdot)$ corresponde à função sigmoidal.

Seja $\theta = (\mathbf{W}, \mathbf{a}, \mathbf{b})$ o conjunto de parâmetros de uma BB-RBM, os quais são aprendidos através de um algoritmo de treinamento que visa maximizar o produtório das probabilidades de ocorrência de todos os dados (vetores) de treinamento $\mathscr{V}$, conforme segue:

$$\arg\max_{\theta} \prod_{\mathbf{v}\in\mathscr{V}} P(\mathbf{v}). \tag{10}$$

Uma das abordagens mais empregadas para resolver este problema se dá por meio do método denominado divergência contrastiva (*contrastive divergence* - CD) (HINTON, 2002), o qual, em suma, simula o processo de amostragem de Gibbs para a convergência da rede, inicializando as unidades visíveis com os vetores de treinamento.

Vale ressaltar que, na presença de dados reais, como quando se trabalha com imagens em tons de cinza, deve-se empregar outro tipo de RBM, a chamada Gaussian-Bernoulli RBM (GB-RBM) (NAIR; HINTON, 2014), a qual modela o vetor de entrada por meio de unidades que seguem a distribuição gaussiana. Deste modo, a Equação 3 deve ser reescrita como:

$$E(\mathbf{v},\mathbf{h}) = \frac{1}{2}\sum_{i=1}^{m}\frac{(v_i - a_i)^2}{\sigma_i^2} - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{v_i}{\sigma_i} h_j w_{ij}. \tag{11}$$

Dada a modificação nas unidade visíveis, é necessário reformular suas probabilidades condicionais. Assim, a Equação 8 deve ser reescrita como:

$$P(v_i|\mathbf{h}) = \mathcal{N}\left(v_i \middle| \sum_{j=1}^{n} w_{ij} h_j + a_i, \sigma_i^2\right), \tag{12}$$

onde, em ambas as equações, $\sigma_i^2$ corresponde à variância da distribuição gaussiana $\mathcal{N}$ do vértice visível $i$. Em geral, normaliza-se os vetores de treinamento para apresentarem valores seguindo distribuição normal com média zero e variância unitária a fim facilitar o aprendizado, eliminando a necessidade de se estimar $\sigma_i^2$, com $i = 1, 2, ..., m$.

# 4 Abordagem proposta

Neste trabalho, propõe-se o uso de uma GB-RBM discriminativa, seguindo Montavon, Orr e Müller (2012) e Souza, Marana e Papa (2017), a fim aprender as características mais relevantes e classificar imagens faciais em reais ou sintéticas com maior acurácia com base em informações de textura, muitos usadas neste tipo de problema, extraídas pelo descritor LBP (*local binary patterns*) (OJALA; PIETIKÄINEN; HARWOOD, 1996). A arquitetura da GB-RBM discriminativa é idêntica à de uma GB-RBM tradicional, exceto pelo fato de que são inseridos, na camada visível da rede, dois neurônios adicionais que servem para representar a classe do vetor (imagem) de entrada: se real ou falsa.

No treinamento, conforme mostra a Figura 3, após converter cada imagem facial conhecida para sua versão LBP e normalizá-la como dito (média zero e variância unitária), os valores de seus pixels servem de entrada para os nós visíveis tradicionais da GB-RBM e os dois valores referentes à classe da imagem ("1 e 0" se face real, ou "0 e 1" se falsa) são considerados como ativação para os dois neurônios especiais. Dadas todas as imagens de treinamento e suas respectivas classes, procede-se então com a divergência contrastiva a fim de ajustar os parâmetros, isto é, o modelo interno da rede, à distribuição de tais amostras conhecidas a fim de poder classificar amostras de teste posteriormente.

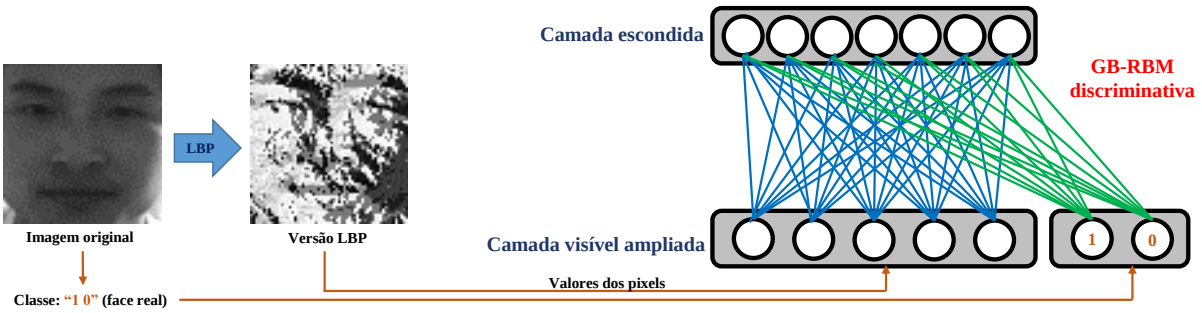


Figura 3: Treinamento da arquitetura proposta: os tons de cinza das imagens de treinamento (pré-processadas pelo LBP e normalizadas) e os dois valores referentes às suas classes alimentam a GB-RBM discriminativa.

Para determinar a classe de uma dada imagem facial de teste, aplica-se o LBP sobre a mesma, servindo seus pixels (após normalização) de entrada para os nós visíveis tradicionais da GB-RBM, e verifica-se qual configuração dos dois neurônios especiais (se "1 e 0", ou "0 e 1") apresenta maior probabilidade de ocorrência segundo a distribuição de probabilidades aprendida pela rede. Para isto, calcula-se a energia livre (HINTON, 2002) fornecida pela GB-RBM discriminativa ao se apresentar a imagem de teste e os valores "1 e 0" e ao se apresentar a mesma imagem e os valores "0 e 1" à sua camada visível. A configuração dos dois nós adicionais responsável pela energia livre mínima indica a classe da imagem de teste.

# 5 Experimento e resultados

A arquitetura proposta foi avaliada sobre a base de imagens NUAA (TAN et al., 2010), a qual apresenta 3.491 imagens obtidas de faces reais e sintéticas (fotografias) para treinamento e 9.123 imagens para teste de métodos anti-*spoofing*. As imagens foram obtidas por meio de *webcams* convencionais a partir das faces de diferentes indivíduos em termos de idade e gênero bem como em diferentes sessões. A Figura 4 ilustra algumas das imagens que compõem tal base. Conforme pode-se observar, mesmo visualmente, é difícil identificar quais faces são reais e quais são falsas.

A GB-RBM discriminativa avaliada possuía 4.098 neurônios visíveis (4.096 para os pixels das imagens $64 \times 64$ e 2 neurônios adicionais para identificar suas classes) e 2.000 neurônios escondidos. A rede foi treinada sobre 6.982 imagens faciais (3.491 imagens de treinamento originais da base acrescidas de suas versões equalizadas) por 10 épocas (iterações), utilizando *learning rate* de 0,001, *momentum* de 0,5 (e 0,9 nas 5 últimas épocas), bem como *weight decay* de 0,0002.

O método proposto obteve acurácia de **93,6%** na classificação das faces sobre as 9.123 imagens de teste, enquanto a técnica proposta pelos próprios autores da base NUAA, que se vale de descritores *handcrafted*, obteve acurácia de **92,0%**. Em complemento, as taxas de falsa aceitação (FAR - *false acceptance rate*) e falsa rejeição (FRR - *false rejection rate*) da abordagem proposta foram de apenas 7,24% e 4,97%, respectivamente.

Figura 4: Exemplo de imagens da base de detecção de *spoofing* facial NUAA (TAN et al., 2010). As imagens faciais da base são mostradas em tons de cinza e suas capturas logo acima.

# 6 Conclusão

Com base nos resultados obtidos neste trabalho, pode-se perceber que as redes neurais podem aprender e extrair importantes características a partir de amostras conhecidas dos problemas com que lidam, neste caso, detecção de *spoofing* facial, propiciando um desempenho robusto mesmo em tarefas complexas como o teste sobre a base NUAA, onde há grande similaridade interclasses e variabilidade intraclasse. A rede apresentada, baseada nas máquinas de Boltzmann restritas, ao aprender boas características de textura a partir de informaçoes extraídas pelo descritor *local binary patterns* (LBP) das faces reais e sintéticas, obteve acurácia superior na detecção de ataques à do método proposto pelos próprios autores da base de imagens analisada, que se vale apenas de descritores *handcrafted*. As taxas de erro de falsa aceitação e falsa rejeição obtidas também foram bastante baixas, evidenciando que a abordagem proposta se configura como boa alternativa para a detecção de *spoofing* em sistemas de reconhecimento facial.

# 7 Referências bibliográficas

ACKLEY, D. H.; HINTON, G. E.; SEJNOWSKI, T. J. A learning algorithm for Boltzmann Machines. **Cognitive Science**, v. 9, n. 1, p. 147-169, 1985.

HINTON, G. E. Training products of experts by minimizing contrastive divergence. **Neural Computation**, v. 14, n. 8, p. 1711-1800, 2002.

HINTON, G. E.; SEJNOWSKI, T. J. Optimal perceptual inference. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 1983. **Proceedings...** Washington, DC: IEEE, 1983. p. 448-453.

JAIN, A. K. et al. Biometrics: a grand challenge. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 17., 2004, Cambridge. **Proceedings...** Cambridge: IEEE, 2004. p. 935-942.

JAIN, A. K.; ROSS, A.; NANDAKUMAR, K. **Introduction to biometrics**. New York: Springer, 2011.

MENOTTI, D. et al. Deep representations for iris, face, and fingerprint spoofing detection. **IEEE Transactions on Information Forensics and Security**, v. 10, n. 4, p. 864-879, 2015.

MONTAVON, G.; ORR, G. B.; MÜLLER, K. R. **Neural networks**: tricks of the trade. 2. ed. Heidelberg: Springer, 2012.

NAIR, V.; HINTON, G. E. Implicit mixtures of restricted Boltzmann machines. In: CONFERENCE NEURAL INFORMATION PROCESSING SYSTEMS, 2008, [S.l.]. **Advances in Neural Information Processing Systems 21.** v. [S.l.: s.n.]: 2008. p. 1145-1152. Disponível em: <https://papers.nips.cc/paper/3536-implicit-mixtures-of-restricted-boltzmann-machines>. Acesso em: 10 out. 2017.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. **Pattern Recognition**, v. 29, n. 1, p. 51-59, 1996.

RUMELHART, D. E.; MCCLELLAND, J. L. **Parallel distributed processing**: explorations in the microstructure of cognition. Cambridge: MIT Press, 1986.

SILVA, M. V. da; MARANA, A. N.; PAULINO, A. A. On the importance of using high resolution images, third level features and sequence of images for fingerprint spoof detection. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2015, Brisbane. **Proceedings...** Piscataway: IEEE, 2015. p. 1807-1811.

SOUZA, G. B.; MARANA, A. N.; PAPA, J. P. Detecção de ataques a sistemas de reconhecimento facial: uma abordagem baseada nas Máquinas de Boltzmann Restritas. In: ENCONTRO REGIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 2017, Bauru. **Caderno de trabalhos completos e resumos.** Bauru: Unesp, Faculdade de Ciências, 2017. p. 465-467. Disponível em: <http://www.fc.unesp.br/Home/Departamentos/Matematica/ermac/caderno-ermac_2017.pdf>. Acesso em: 20 nov. 2017.

TAN, X. et al. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 2010, Heraklion. **Lecture notes in computer science.** Berlin: Springer, 2010. p. 504-517.

TANG, Y.; SALAKHUTDINOV, R.; HINTON, G. E. Robust Boltzmann machines for recognition and denoising. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2012, Providence. **Proceedings...** Washington, DC: IEEE Computer Society, 2012.

Artigo recebido em jun. 2017 e aceito em out. 2017.