Méthode de détection d'attaques par usurpation d'identité basée sur l'étude du bruit des images

Hoai Phuong NGUYEN[1], Agnès DELAHAIES[1], Florent RETRAINT[2], Frédéric MORAIN-NICOLIER[1]

[1]Laboratoire CREsTIC, Université de Reims Champagne-Ardenne

[2]Laboratoire LM2S, ICD, Université de Technologie de Troyes

hoai-phuong.nguyen@etudiant.univ-reims.fr, agnes.delahaies@univ-reims.fr
florent.retraint@utt.fr, frederic.nicolier@univ-reims.fr

Résumé – Les systèmes d'authentification utilisant la reconnaissance faciale sont vulnérables aux attaques par usurpation d'identité. Cet article propose une nouvelle méthode exploitant le comportement statistique du bruit dans l'image pour détecter ce type d'attaques. Ce comportement est étudié à travers un modèle statistique du bruit reposant sur le calcul de la variance locale. La construction d'un test d'hypothèse statistique permet de détecter la présence d'une éventuelle attaque.

Abstract – This paper aims to study the problem of spoofing attack detection for facial recognition systems. Real faces and falsified faces present in front of a security system have differences of micro-textures on their surface, which are exploited to discriminate face spoofing images. Our method exploits the statistic behavior of the distribution of noise's local variances (NLV), which performs differently between images of real faces and fake ones. The detection problem was casted in the framework of hypothesis testing. A statistical test was proposed.

1 Introduction

L'authentification par reconnaissance faciale devient actuellement une solution complémentaire permettant de renforcer la sécurité des systèmes d'information. Cependant, ces méthodes de reconnaissance faciale sont vulnérables aux attaques d'usurpation d'identité. En effet, un attaquant peut contourner le processus d'authentification en présentant tout simplement, devant la caméra, une photo du visage d'un utilisateur légitime obtenue facilement sur les réseaux sociaux. Pour la sécurité de ces systèmes, il est donc vital de pouvoir identifier et éliminer toutes ces attaques.

Plusieurs approches ont été proposées dans la littérature pour la détection des attaques d'authentification par reconnaissance faciale. Elles peuvent être divisées entre deux grandes catégories : méthodes intrusives et non intrusives. Les approches intrusives demandent la coopération de l'utilisateur lors de l'authentification. Par exemple, il est demandé à l'utilisateur de changer la position de sa tête, de cligner les yeux ou de prononcer quelques mots spécifiques en respectant un certain protocole prédéfini [1],[2]. Ces approches peuvent être difficiles à mettre en œuvre par les utilisateurs. Les approches non intrusives, quant à elles, sont plus ou moins transparentes pour les utilisateurs. On privilégiera donc ce type d'approche pour développer notre méthode.

Matta et al. [3] ont proposé une approche fondée sur l'analyse des textures utilisant des motifs binaires locaux (*Local Binary Patterns* ou LBP). Kim et al. [4] ont combiné l'exploitation du modèle LBP avec une étude fréquentielle pour détecter les attaques utilisant des masques papier 2D.

Des structures graphiques locales LGS (*Local Graph Structure*) sont également utilisées [5],[6],[7]. Bao et al. [8] ont proposé une méthode qui exploite la variation des champs des flux optiques générés lors des mouvements de l'objet. En effet, un objet 3D et un plan 2D vont produire des champs de flux optiques qui varient très différemment. Une autre méthode est proposée dans [9] qui analyse la variation de l'intensité des pixels entre deux images capturées avec deux mises au point différentes.

Dans cet article, nous proposons une nouvelle méthode non-intrusive qui exploite le comportement statistique des petites textures présentes à la surface de l'objet imagé. Ces textures viennent par exemple de l'imperfection du processus d'impression d'une photographie ou de l'effet moiré présent lors de l'utilisation d'un écran LCD. Ces petites textures peuvent être interprétées comme un bruit venant s'additionner au processus naturel de formation d'une image numérique. Ainsi, il est proposé d'étudier le comportement statistique des pixels de l'image afin de distinguer un comportement légitime d'une attaque.

Le reste de cet article est organisé comme suit. Premièrement, dans la deuxième section, nous proposons l'utilisation d'un modèle de la variance locale afin de caractériser le comportement statistique des images. Dans la section suivante, il est proposé de montrer l'impact d'une attaque sur le modèle de la variance locale. Ensuite, la construction d'un test d'hypothèse statistique permet de réaliser la détection d'une éventuelle attaque. Les résultats obtenus sur une base de données

composée de près de 2000 images légitimes et non légitimes sont présentés dans la section 5.

2 Modèle de la variance locale

La variance locale est calculée à d'un partir d'un bloc 8x8 de l'image de bruit, c'est à dire l'image dans lequel le contenu débruité a été retiré. On note σ^2 la valeur de la variance locale d'un bloc donné :

$$\sigma^2 = \frac{1}{63}\sum_{i=0}^{63}(N_i - \overline{N})^2 \tag{1}$$

où N_i est la valeur du i-ème pixel du bloc et $\overline{N}$ représente la moyenne des valeurs du bloc étudié avec $\overline{N} = \frac{1}{64}\sum_{j=0}^{63} N_j$. Par conséquent, on a :

$$N_i - \overline{N} = \frac{1}{64}\sum_{j=0}^{63}(N_i - N_j). \tag{2}$$

Les N_i sont des variables indépendantes et identiquement distribuées. Selon le théorème central limite [10], la distribution des $N_i - \overline{N}$ approche une distribution gaussienne d'espérance nulle. Nous savons que le carré d'une variable gaussienne suit une distribution du khi-deux avec un degré de liberté. En comparant les fonctions génératrices des moments, une variable aléatoire de loi du khi-deux avec un degré de liberté échelonnée par une constante suit une distribution Gamma de paramètre de forme égale à $1/2$ [11] :

$$Y_i = \frac{1}{63}(N_i - \overline{N})^2 = \frac{\sigma_N^2}{63}(\frac{N_i - \overline{N}}{\sigma_N})^2 \xrightarrow{d} \Gamma(\frac{1}{2}, \beta_i). \tag{3}$$

d'où $\xrightarrow{d}$ signifie la convergence en distribution et β_i est un paramètre dépendant de la distribution des N_i.

La variance σ^2 est donc considéré comme la somme de 64 variables Y_i de loi Gamma et corrélées entre eux. Il résulte de [12] que la fonction génératrice de σ^2 peut être exprimée comme suit :

$$M_{\sigma^2}(t) = [det(\mathbf{I}_{64} - t\mathbf{DC})]^{-\frac{1}{2}} \tag{4}$$

avec $det(.)$ le déterminant, $\mathbf{I}_{64}$ est la matrice d'unité d'ordre 64, $\mathbf{D}$ une matrice diagonale de même ordre composée par les coefficients β_i et $\mathbf{C}$ la matrice de covariance définie par :

$$\mathbf{C} = \begin{pmatrix} 1 & \sqrt{\rho_{1,2}} & \dots & \sqrt{\rho_{1,64}} \\ \sqrt{\rho_{2,1}} & 1 & \dots & \sqrt{\rho_{2,64}} \\ . & . & \dots & . \\ \sqrt{\rho_{64,1}} & \sqrt{\rho_{64,2}} & \dots & 1 \end{pmatrix} \tag{5}$$

avec $\rho_{i,j}$ le coefficient de corrélation entre Y_i et Y_j.

En notant $\{\lambda_i\}_{i=0}^{63}$ les valeurs propres de la matrice $\mathbf{DC}$, l'équation 4 s'écrit comme suit :

$$M_{\sigma^2}(t) = \prod_{i=0}^{63}(1 - t\lambda_i)^{-\frac{1}{2}}. \tag{6}$$

On peut donc approximer la distribution de σ^2 par une distribution Gamma $\Gamma(\alpha, \beta)$ [1] ,dont les paramètres sont déterminées par les relations suivantes :

$$\alpha = \frac{(\sum_{i=0}^{63}\lambda_i)^2}{2\sum_{i=0}^{63}\lambda_i^2}, \quad \beta = \frac{\sum_{i=0}^{63}\lambda_i^2}{\sum_{i=0}^{63}\lambda_i}. \tag{7}$$

La figure 1 montre un exemple de la distribution de la variance locale obtenue à partir d'une image et la distribution Gamma théorique correspondante.

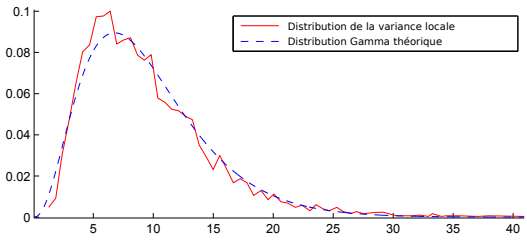

FIGURE 1 – Histogramme de la variance locale du bruit comparée avec son raccordement par une distribution gamma.

3 Caractéristiques d'une attaque

Lors d'une attaque, il est proposé de détecter la présence de petites textures dans l'image. Ces dernières peuvent être interprétées comme un bruit venant s'ajouter au processus naturel de formation de l'image numérique. Il est difficile d'étudier séparément le bruit de texture et le bruit issu du processus d'acquisition. Cependant, comme le montre la figure 2, la présence du bruit de texture provoque une modification de la distribution de la variance locale.

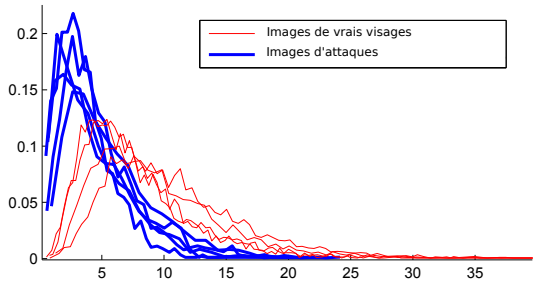

FIGURE 2 – Distribution de la variance locale du bruit (NLV) pour différentes images.

Ainsi, le comportement de la distribution de la variance locale peut être exploité pour détecter une éventuelle attaque et cette dernière sera caractérisée par une variation des paramètres α et β.

1. La densité de probabilité d'une distribution Gamma $\Gamma(\alpha, \beta)$ est donnée comme suit : $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x) \quad \forall x \in \mathbb{R}_+$.

4 Test d'hypothèse

Le bruit d'une image naturelle n'est pas constant dans l'image et dépend du contenu de l'image. En effet, une partie importante du bruit vient du processus de conversion des photons en électrons. De ce fait, le bruit est hétéroscédastique et la variance dépend de l'espérance.

En conséquence, il est proposé de regrouper les blocs 8x8 en fonction de leur espérance. Ainsi, l'intervalle $[0, .., 255]$ est divisié en N plages de largeur identique. Les blocs dont l'espérance se situe dans la même plage sont regroupés. Soit $\mathcal{N}_i$ l'ensemble des blocs du i-ème groupe, d'après la section 2, les $\mathcal{N}_i$ suivent une loi Gamma.

Pour une image donnée, il est possible que le nombre d'éléments d'un groupe n'est pas assez suffisant statistiquement. De ce fait, il est proposé d'étudier les M groupes dont la taille est suffisante ($\mathsf{M} < \mathsf{N}$). Sans perte de généralité, on suppose que ces M populations sont les $\mathcal{N}_i$ avec $(i = 1, .., \mathsf{M})$.

Par conséquent, le problème de détection d'attaques évoqué dans ce document revient à décider entre les deux hypothèses suivantes :

$$\begin{cases} \mathcal{H}_0 : \mathcal{N}_i \sim \Gamma(\alpha_{0i}, \beta_{0i}) \\ \mathcal{H}_1 : \exists i, \mathcal{N}_i \nsim \Gamma(\alpha_{0i}, \beta_{0i}) \end{cases} \tag{8}$$

avec α_{0i} et β_{0i} des paramètres connus.

Soit (α_i, β_i) les paramètres de la distribution Gamma qui caractérisent la population $\mathcal{N}_i$, le test d'hypothèse 8 peut se réécrire comme suit :

$$\begin{cases} \mathcal{H}_0 : (\alpha_i, \beta_i) = (\alpha_{0i}, \beta_{0i}) \\ \mathcal{H}_1 : \exists i, (\alpha_i, \beta_i) \neq (\alpha_{0i}, \beta_{0i}) \end{cases} \tag{9}$$

Sachant que $(\alpha_i, \beta_i) = (\alpha_{0i}, \beta_{0i})$ ssi $(\alpha_i\beta_i, \beta_i) = (\alpha_{0i}\beta_{0i}, \beta_{0i})$, on peut réecrit le test ci-dessus sous la forme matricielle suivant :

$$\begin{cases} \mathcal{H}_0 : C\theta = \Phi_0 \\ \mathcal{H}_1 : C\theta \neq \Phi_0 \end{cases} \tag{10}$$

avec $C = I_{2\mathsf{M}}$ la matrice d'unité d'ordre 2M et

$$\theta = [\alpha_1\beta_1, \beta_1, \alpha_2\beta_2, \beta_2, ..., \alpha_{\mathsf{M}}\beta_{\mathsf{M}}, \beta_{\mathsf{M}}]'$$
$$\Phi_0 = [\alpha_{01}\beta_{01}, \beta_{01}, \alpha_{02}\beta_{02}, \beta_{02}, ..., \alpha_{0\mathsf{M}}\beta_{0\mathsf{M}}, \beta_{0\mathsf{M}}]'.$$

En notant κ_{ij} le j-ème cumulant de la i-ème population, nous obtenons :

$$\kappa_{i1} = \alpha_i\beta_i, \kappa_{ij} = (j-1)!\ \alpha_i\beta_i^j, \text{ et } \kappa_{i,j+1}/\kappa_{ij} = j\beta_i \tag{11}$$

Soient

$$\eta_{i0} = \kappa_{i0}, \eta_{ij} = \kappa_{i,j+1}/\kappa_{ij}, \qquad pour\ j \geq 1 \tag{12}$$
$$\eta_i' = (\eta_{i0}, \eta_{i1}, \eta_{i2}, \eta_{i3}) \quad pour\ i = 1, 2, .., \mathsf{M} \tag{13}$$
$$\eta' = [\eta_1', \eta_2', .., \eta_{\mathsf{M}}'], \tag{14}$$

nous obtenons

$$\eta_i = w^*. \begin{bmatrix} \alpha_i\beta_i \\ \beta_i \end{bmatrix} \tag{15}$$

avec

$$w^{*'} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 \end{bmatrix} \quad \text{pour } i = 1, 2.., \mathsf{M} \tag{16}$$

En posant $w = diag(w^*, w^*, .., w^*)$, nous obtenons la relation linéaire suivante :

$$\eta = w\theta \tag{17}$$

Selon des travaux de Tripathi et al. dans [13], il est possible de proposer un test statistique pour chosir entre les 2 hypothèses ci-dessus. Le test statistique est donné par l'équation suivante :

$$Q = (C\hat{\theta} - \Phi_0)'(C(w'\hat{\Sigma}^{-1}w)^{-1}C')(C\hat{\theta} - \Phi_0) \tag{18}$$

avec $\hat{\theta} = (w'\hat{\Sigma}^{-1}w)^{-1}w^{-1}\hat{\Sigma}^{-1}\hat{\eta}$ et $\hat{\eta}$ une estimation de η.

$\hat{\Sigma}$ représente une estimation de la matrice de covariance Σ de η définie par l'équation suivante :

$$\Sigma = J_2J_1VJ_1'J_2' \tag{19}$$

avec V la matrice de covariance dont les éléments sont en fonction des μ_{ij}, le j-ème moment du i-ème population.

$$V = \tilde{V} + \tilde{V}^T - diag(\tilde{V})$$
$$\tilde{V} = diag(V_1, V_2, ..., V_N), \text{avec}$$
$$V_i = \frac{1}{n_i}\begin{bmatrix} \mu_{i2} - \mu_{i1}^2 & 0 & 0 & 0 \\ \mu_{i3} - \mu_{i2}\mu_{i1} & \mu_{i4} - \mu_{i2}^2 & 0 & 0 \\ \mu_{i4} - \mu_{i3}\mu_{i1} & \mu_{i5} - \mu_{i3}\mu_{i2} & \mu_{i6} - \mu_{i3}^2 & 0 \\ \mu_{i5} - \mu_{i4}\mu_{i1} & \mu_{i6} - \mu_{i4}\mu_{i2} & \mu_{i7} - \mu_{i4}\mu_{i3} & \mu_{i8} - \mu_{i4}^2 \end{bmatrix}$$

J_1 est J_2 sont des matrices Jacobiennes données par les expressions suivantes :

$$J_1 = diag(J_{11}, J_{12}, ..., J_{1\mathsf{M}})$$
$$J_2 = diag(J_{21}, J_{22}, ..., J_{2\mathsf{M}})$$

dont les J_{1i} et les J_{2i} permettent de réaliser les transformations suivantes :

$$J_{1i} : (\mu_{i1}, \mu_{i3}, \mu_{i2}, \mu_{i4}) \longrightarrow (\kappa_{i1}, \kappa_{i2}, \kappa_{i3}, \kappa_{i4})$$
$$J_{2i} : (\kappa_{i1}, \kappa_{i2}, \kappa_{i3}, \kappa_{i4}) \longrightarrow (\eta_{i1}, \eta_{i2}, \eta_{i3}, \eta_{i4})$$

La distribution asymptotique de la statistique Q sous l'hypothèse $\mathcal{H}_0$ suit une loi du khi-deux à 2M degrés de liberté. Ainsi le test de niveau α rejete l'hypothèse nulle $\mathcal{H}_0$ lorsque la statistique Q est plus grande que le quantile d'ordre $1 - \alpha$ de la loi du khi-deux à 2M degrés de liberté :

$$Q > \chi^2_{2\mathsf{M},\alpha} \tag{20}$$

5 Résultats expérimentaux

Comme indiqué dans la formulation du problème, les images doivent être obtenues avec les mêmes conditions d'acquisition. Par conséquent, nous ne pouvons pas tester notre méthode sur les bases de données publiques existantes. Ainsi, nous avons construit une base de données obtenue avec un téléphone portable Samsung Galaxy S6. La base de données contient 1172

FIGURE 3 – Des exemplaires de la base de données : à gauche photo d'un vrai visage, à droite la version falsifiée.

photos d'une personne, dont 522 sont des photos de son vrai visage, 650 sont des photos de son visage imprimé sur papier. Pour calculer la statistique Q, on a pris $N = 20$ et $M = 10$ pour que le nombre minimal d'échantillons par groupe choisie soit supérieur à 100.

Le figure 4 présente la fonction d'efficacité du détecteur (courbe ROC) obtenue à partir de la statistique Q. La puissance de détection est supérieure à 95% quelque soit le taux de fausse alarme fixé.

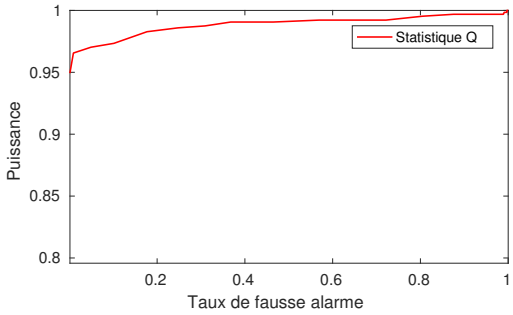

FIGURE 4 – Fonction d'efficacité du détecteur Q

6 Conclusion

Les systèmes d'authentification utilisant la reconnaissance faciale sont connues pour être vulnérables aux attaques par usurpation d'identité. Dans cet article, nous proposons une nouvelle méthode de détection de ces attaques en reposant sur l'étude du bruit des images. La construction d'un test d'hypothèse statistique a permis de détecter ces attaques. En effet, dans l'hypothèse où les conditions d'acquisition des images restent stables, les performances du test sont très bonnes.

Références

[1] L. Sun, G. Pan, Z. Wu, and S. Lao, "Blinking-based live face detection using conditional random fields," in *Proceedings of the 2007 International Conference on Advances in Biometrics*. 2007, ICB'07, pp. 252–260, Springer-Verlag.

[2] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in "liveness" assessment," *Trans. Info. For. Sec.*, vol. 2, no. 3, pp. 548–558, sep. 2007.

[3] J. Maatta, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *2011 International Joint Conference on Biometrics (IJCB)*, oct. 2011, pp. 1–7.

[4] G. Kim, S. Eum, J. Suhr, D. Kim, K. Park, and J. Kim, "Face liveness detection based on texture and frequency analyses," in *2012 5th IAPR International Conference on Biometrics (ICB)*, mar. 2012, pp. 67–72.

[5] Housam Khalifa Bashier, Siong Hoe Lau, Pang Ying Han, Liew Yee Ping, and Chiang Mee Li, "Face spoofing detection using local graph structure," in *Proceedings of the 2014 International Conference on Computer, Communications and Information Technology*. 2014, Atlantis Press.

[6] H.K. Bashier, S.H. Lau, Y.H. Pang, Y.P. Liew, and M.L. Chiang, "Face spoofing detection based on improved local graph structure," in *2014 International Conference on Information Science and Applications (ICISA)*, 2014, pp. 1–4.

[7] M.F.A. Abdullah, M.S. Sayeed, K. Sonai Muthu, H.K Bashier, A. Azman, and S.Z. Ibrahim, "Face recognition with symmetric local graph structure (SLGS)," *Expert Systems with Applications*, vol. 41, no. 14, pp. 6131–6137, oct. 2014.

[8] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *International Conference on Image Analysis and Signal Processing*, avr. 2009, pp. 233–236.

[9] S. Kim, S. Yu, K. Kim, Y. Ban, and S. Lee, "Face liveness detection using variable focusing," in *2013 International Conference on Biometrics (ICB)*, jun. 2013, pp. 1–6.

[10] M. Blum, "On the central limit theorem for correlated random variables," *Proceedings of the IEEE*, vol. 52, no. 3, pp. 308–309, mar. 1964.

[11] T.H. Thai, R. Cogranne, and F. Retraint, "Statistical model of quantized DCT coefficients : Application in the steganalysis of jsteg algorithm," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 1980–1993, 2014.

[12] M.S. Alouini, A. Abdi, and M. Kaveh, "Sum of gamma variates and performance of wireless communication systems over nakagami-fading channels," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 6, pp. 1471–1480, nov. 2001.

[13] Ram C. Tripathi, Ramesh C. Gupta, and Robert K. Pair, "Statistical tests involving several independent gamma distributions," *Annals of the Institute of Statistical Mathematics*, vol. 45, no. 4, pp. 773–786.