修士論文要旨 (2013 年度)

虹彩認証システムへのウルフ攻撃による安全性評価
The Security Evaluation of Iris Recognition against Wolf-Attacks

電気電子情報通信工学専攻 丹 寛之
Hiroyuki TAN

1 序論

生体認証とは, 指紋や静脈などの生体情報を, 登録されたその人のテンプレートとの差をスコアとして算出して照合を行う自動認証技術である. 生体認証に関する研究は, 精度の向上や処理速度の向上などに着眼点を置かれる. 精度の向上は, 他人同士の生体情報が同一であると誤判定されるゼロ·エフォート攻撃に対する耐性として評価され, 他人受入率 (False Accept Rate:FAR) で表される. この値が小さいほど, 認証精度が高くゼロ·エフォート攻撃への耐性が高いといえる.

しかし, 複数の生体情報と一致と誤判定される入力情報『ウルフ』を巧みに利用する攻撃 [1] を想定した場合, このなりすましに対する耐性を FAR によって評価することはできない. ウルフへの耐性を評価するためには, 攻撃者に最も有利な入力情報をシステムに与えた時の成功確率を求める必要があり, その最大成功確率をウルフ攻撃確率 (Wolf Attack Probability:WAP) と呼ぶ. WAP は, 最も攻撃成功確率の高い確率を用いるため, FAR より安全性の評価に適している.

本研究では, 生体認証システムの中でも認証精度が非常に高い Daugman の虹彩認証 [2] に対してウルフ攻撃による安全性評価を行い, WAP を考慮した安全性評価が必要であることを示す. 2 章では, 本研究が着眼点を置いた虹彩認証とウルフ攻撃確率について紹介を行う. 3 章では, 本研究で発見したウルフについて解説する. 具体的には, 虹彩領域と非虹彩領域の境界がもつ位相情報の偏りを利用することで, FAR を大きく上回る照合成功確率を実現する符号語のことである. 4 章では, 3 章で発見したウルフがセンサを通して実現できるかを検証し, 虹彩認証が特徴情報の抽出に用いる方式を利用することで, 高い確率で再現可能であることを示す.

なお, 使用する虹彩画像のデータベースは CASIA Iris Ver4.0 に収録されている虹彩画像 Lamp である.

2 虹彩認証とウルフ攻撃確率

FAR による安全性評価はセンサに提示されるものが人間のものであることを前提としており, 悪意あるなりすまし攻撃への耐性は評価できない. このような攻撃に対する安全性評価基準として WAP がある. 2 章では, ウルフ攻撃確率と実際の製品に広く用いられている Daugman の虹彩認証システム [2] について解説する. 具体的には, Telecom Management Sud Paris の公開アルゴリズムであり, Daugman が提案している照合アルゴリズムを実装した OSIRIS ver2.1 である.

2.1 ウルフ攻撃確率 (WAP)

WAP とは, 人工物も含めたセンサに提示可能なあらゆるサンプルを考慮した際の最大のなりすまし成功確率のことであり, 次の式で定義される.

$$\mathrm{WAP} = \max_{s \in S_A} \left(\operatorname*{Ave}_{t \in T_h} \Pr[\mathrm{match}(s,t) = \text{"}accept\text{"}] \right) \quad (1)$$

S_A は人工物も含めたサンプルデータ全体の集合であり, T_h は生体認証システムに保管されているテンプレート全体の集合である. match は生体認証システムで用いられる照合アルゴリズムのことである. 従来の安全性評価尺度である FAR は他人を誤って受け入れてしまい, 誤一致と判定されてしまう確率のことであり, ゼロ·エフォート攻撃によって次の式で定義される.

$$\mathrm{FAR} = \operatorname*{Ave}_{s \in S_h} \left(\operatorname*{Ave}_{t \in T_h} \Pr[\mathrm{match}(s,t) = \text{"}accept\text{"}] \right) \quad (2)$$

ここで S_h は生体情報からなるサンプルデータ全体の集合である. 強力なウルフが存在していたとしても, ウルフの数がサンプルデータのごく一部である場合, FAR によって正しく評価されない. よって, 第三者によるなりすましに対する耐性として WAP を十分に小さく抑えることが重要である.

2.2 虹彩認証

Daugman の虹彩認証での認証の流れは, 1. 虹彩画像の取得, 2. 虹彩領域の抽出, 3. テンプレートのアイリスコードとマスクコードを生成, 4. ハミング距離による照合, 5. 判定結果の出力 となっており, このアルゴリズムは他の生体認証と比較して，FAR が極めて低いことが確かめられている．これらの処理の内, 本研究に深く関係している 3. 4. について解説していく．

2.2.1 テンプレートの生成

虹彩認証では瞳孔と白目との間にある虹彩模様を生体情報としてテンプレートを生成する．このテンプレートは虹彩模様の特徴情報を符号化したアイリスコードと, 抽出する虹彩領域の信頼度を表すマスクコードによって構成されている．

アイリスコードの生成

虹彩領域に対して 2 次元ウェーブレット変換と符号化関数を用いて, 虹彩模様の特徴情報を表す 2bit を生成する．ウェーブレット変換とは画像にある情報を畳み込み積分によって数値化するものである．これに確率密度関数であるガボールフィルタを合わせることで, 虹彩画像の任意の座標にある局所的な情報を正確に得られる．このフィルタにある各パラメータを変更することで, 読み取りたい位相情報を任意のものに変更することができる．ウェーブレット変換によって数値化された位相情報は複素平面上に表すことができ, 実部と虚部それぞれの正負を基準として符号化を行う．この処理を虹彩画像の各ピクセルに対して行い, OSIRIS ではそれぞれのフィルタに合わせて 774bit 生成し, 計 2322bit のアイリスコードを特徴情報とする．

マスクコードの生成

アイリスコードと同じコード長をもつマスクコードを生成する．瞼など非虹彩領域や照明などの環境要因の影響を受けた虹彩領域の bit 値を 0 とする．この bit 値が 0 である場合, その領域におけるアイリスコードは照合から除外される．これにより非虹彩領域等が引き起こす認証精度への悪影響を回避する．OSIRIS では先述のアイリスコードと同じように, 各フィルタに沿った 774bit のマスクコードを計 2322bit 生成する．

図 1: 白黒の領域はアイリスコードの bit 値を表し, 赤の領域はマスクコードによって照合から除外される

1 つのフィルタによって生成された 774bit の符号語の模式図を図 1 に示す．図の白と黒の領域はアイリスコードの bit 値を表し, 赤の領域は照合から除外される領域を表している．

2.2.2 ハミング距離を用いた照合及び判定

OSIRIS では 2.2.1 項で生成された符号語 4644bit をその虹彩画像のテンプレートとする．このテンプレートと登録テンプレートとのハミング距離 $\mathrm{HD}_{\mathrm{raw}}$ を次式より求める．

$$\mathrm{HD}_{\mathrm{raw}} = \frac{\left|\left|(\mathrm{codeA} \oplus \mathrm{codeB}) \cap \mathrm{maskA} \cap \mathrm{maskB}\right|\right|}{\left|\left|\mathrm{maskA} \cap \mathrm{maskB}\right|\right|} \quad (3)$$

$||\cdot||$ は 1 となるビットの個数を表すハミング重みを示す．(3) 式はテンプレート間の単純ハミング距離を求める式であり, そのスコアは照合から除外されない虹彩領域の共通部分のコード長 $||\mathrm{maskA} \cap \mathrm{maskB}||$ に大きく依存する．そこで, コード長に左右されない正規化ハミング距離 $\mathrm{HD}_{\mathrm{norm}}$ を $\mathrm{n} = ||\mathrm{maskA} \cap \mathrm{maskB}||$ として, 次式より求める．n は 2 つの虹彩画像において, 照合から除外されない虹彩領域の共通部分のコード長である．

$$\mathrm{HD}_{\mathrm{norm}} = 0.5 - (0.5 - \mathrm{HD}_{\mathrm{raw}})\sqrt{\frac{\mathrm{n}}{911}} \quad (4)$$

(4) 式の 0.5 は, 他人同士の虹彩画像で, $\mathrm{HD}_{\mathrm{raw}}$ を全ての組み合わせに対して行った平均値である． また, 式中の 911 も同様に他人同士の虹彩画像での $\mathrm{n} = ||\mathrm{maskA} \cap \mathrm{maskB}||$ によって求められた値の平均値である．これらの値は使用する虹彩画像データベースによって上下する．0.5 や n, 911 によってハミング距離 $\mathrm{HD}_{\mathrm{raw}}$ を正規化することで, 認証に用いるコード長を表す n が小さい場合でも FAR を一定に保つ．求められた $\mathrm{HD}_{\mathrm{norm}}$ があらかじめ決められた認証閾値 T よりも小さい場合 ”accept” を返し, 認証閾値 T よりも大きい場合 ”reject” を返す．

3　虹彩認証に対するウルフ攻撃

3 章では, 虹彩認証に対するウルフ攻撃について, その研究成果と安全性評価について述べる. 既存研究として小島ら [3] によるものがあり, 攻撃者がマスクコードを操作することで FAR を大幅に上昇させる手法を提案していた. しかし, 具体的なウルフの提示や具体的な WAP が求められていなかった.

本研究はこれまで探索されていないウルフを探索し, 実験によって WAP を初めて求めたものである. 本章では OSIRIS が生成したテンプレートに対して, 高い誤一致率をもつアイリスコードと誤一致率を上げるマスクコードから構成されるウルフの作成法を示す.

3.1　被マスク率と特徴情報の出現確率分布

$(1,1)$,$(0,1)$,$(0,0)$,$(1,0)$ のうち最大出現確率の 2bit を, 各領域がマスクされない確率ごとにソートして図 2 に示す. 横軸は値が大きくなるほどその領域の被マスク率が上がり, 縦軸は各領域におけるいずれかの 2bit の出現確率を表す. 0.25 から伸びている赤線は, 4 種類の 2bit が平等に出現した独立分布を示し, この赤線から離れるほど特徴情報が偏っていることを意味する. 被マスク率が上がることで, 特徴情報の偏りが明確になっていることがわかる.

3.2　アイリスウルフ

実際の照合では特徴情報が偏りにくい領域で照合を行っており, 偏った特徴情報は無視できる程度の影響かマスクコードによって照合から除外される. これらの無視あるいは除外される領域を持ち, 各領域に偏った特徴情報を設定することで, 高い誤一致率をもつウルフを作成することが可能である.

OSIRIS の 1 つのフィルタに対するアイリスウルフを図 3 に示す. ウルフは虹彩画像における上瞼や白目 · 瞳孔との境界付近を照合に用い, 最もマスクされにくい虹彩領域を照合から除外する性質をもっている.

3.3　ウルフ攻撃実験及び実験結果

OSIRIS のフィルタに合わせて作成したウルフを用いて, なりすまし攻撃のシミュレーション実験を行う.

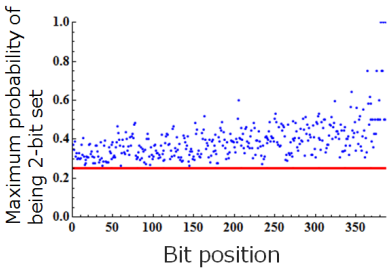

図 2: 各座標の被マスク率とある 2bit の最大出現確率

図 3: アイリスコードの偏りを利用したウルフ

表 1: アイリスウルフを用いたなりすまし成功確率

DB	認証閾値	FAR[%]	WAP[%]
虹彩 DB^I_1	0.393	0.98	42.6
虹彩 DB^I_2	0.375	0.33	14.0
虹彩 DB^I_3	0.378	0.65	33.0

攻撃対象のデータベースには, ウルフが学習に用いたウルフ DB^I 以外の虹彩画像を収録している虹彩 DB^I を用いた. 作成したウルフを 1 回提示し, 各虹彩 DB^I のテンプレート全てと照合を行い, その結果からウルフのなりすまし成功確率 WAP を求める.

高い認証精度を持つ虹彩認証に対して, 作成したウルフによるなりすまし攻撃実験の結果を表 1 に示す. 従来の評価基準である FAR に比べ, WAP はおよそ 40〜50 倍の誤一致率となり, 最高でなりすまし攻撃確率 $WAP = 42,6\%$ を達成した.

また, 虹彩同士の照合における共通部分のコード長の平均値は 1261bit であり, 虹彩とウルフの照合における共通部分のコード長の平均値は 551bit である. この結果より, ウルフを提示した照合では通常の虹彩同士の照合よりも少ないコード長によってなりすましに成功しており, (4) 式によるスコア正規化が, 通常の照合時よりも少ない共通部分のコード長をもつウルフに対して充分に行えていないことが示された.

この結果より, 従来の FAR による安全性評価では高い認証精度をもっている虹彩認証にも無視できない脆弱性があることが示された.

4　安全性評価用の人工虹彩生成法

センサを含めた安全性評価では, 1. 照合アルゴリズムに対するウルフの探索, 2. ウルフを出力する人工物の生成 の 2 つのステップを踏む必要がある. これまでの研究では, 電子データ上でのシミュレーションによる安全性評価であり, 2. 人工物の生成 を考慮していない. 本章では虹彩認証に対して, WAP による安全性評価のための人工虹彩の生成法を提案する.

4.1　人工虹彩画像の生成

アイリスコードを生成する各領域の 1 つ 1 つに注目する. 各領域から生成される 2bit は, その座標における位相情報を符号化したものである. 仮に 2bit を (0,0) とすると, 符号化する前の複素平面上の座標は (-1,-1) と置くことができる. このとき, この情報の振幅情報は 1 であり, 位相情報は $\frac{3\pi}{4}$ である. 各虹彩領域 $\mathrm{I}(\rho,\phi)$ における位相情報を持つ画像は, n 象限目の位相情報を $\frac{(2\mathrm{n}-1)\pi}{4}$ とし, 次のように定義できる.

$$I_{(\rho,\phi)} = \int_{\rho}\int_{\phi} \cos\left(\frac{2\pi(\theta_0-\rho)}{\beta}+\frac{(2n-1)\pi}{4}\right) \times \mathrm{e}^{-\frac{(r_0-\rho)^2}{\alpha^2}} \times \mathrm{e}^{-\frac{(\theta_0-\rho)^2}{\beta^2}} \times \rho d\rho d\phi \quad (5)$$

(5) 式で生成される人工虹彩模様を全ての領域に対して行い, 重ね合わせて図 4 の人工虹彩画像を生成する.

4.2　性能評価実験及び実験結果

各人工虹彩模様から得られる 2bit は元の 2bit と一致するが, 模様が重なり合うことで元のアイリスコードと異なることが予想される. 元のアイリスコードとの差を $\mathrm{HD}_{\mathrm{norm}}$ から求め, 再現性を評価する. 本節では波長 β_{i} をもつ照合アルゴリズムを Π_{i} として実装し, 各波長 β_{i} で生成した人工虹彩画像の性能をアルゴリズム Π_{i} で評価する. なお, 虹彩領域抽出や極座標変換の際に付与されるノイズは無いものとする.

アルゴリズムごとの性能評価実験の結果を表 2 に記す. 表より, 類似率 $\mathrm{HD}_{\mathrm{norm}}$ は 0.04 から 0.27 に収まっており, 全ての人工虹彩画像が照合に成功している. この結果から, 本研究で提案する人工虹彩画像生成法が, 最も類似率が低い場合においても照合に成功する性能を有していることが示された.

図 4: 人工虹彩画像

表 2: アルゴリズム Π_{i} ごとの人工虹彩画像の性能評価

アルゴリズム (β, FAR, 閾値)	平均値	最小値	最大値
$\Pi_1(20, 3.760, 0.464)$	0.111	0.040	0.238
$\Pi_2(24, 1.963, 0.452)$	0.116	0.040	0.246
$\Pi_3(28, 0.670, 0.434)$	0.119	0.046	0.267

5　結論

本研究では, 認証精度が非常に高い Daugman の虹彩認証に対して, 悪意あるなりすまし攻撃であるウルフ攻撃に焦点を当てて, 攻撃成功確率を求めて安全性評価を行った. 1 つ目は, 特徴情報の分布を解析することで, ウルフのような無視できない脆弱性があることを示した. 2 つ目は, 虹彩模様を解析するウェーブレット変換を利用することで, 安全性評価のための人工虹彩生成法を示した. これらの成果より, ウルフ対策と WAP による安全性評価の導入が強く望まれる.

研究業績

- 丹 寛之, 井沼 学, 大塚 玲, 北川 隆, 今井 秀樹, “虹彩パターンの周波数情報の抽出領域による情報量の偏りを用いたウルフ攻撃,”SCIS2012, 1F2-1, 2012.
- 丹 寛之, 井沼 学, 大塚 玲, 米澤 祥子, 今井 秀樹, “虹彩認証アルゴリズムに対するウルフ攻撃研究,” 第 2 回バイオメトリクスと認識・認証シンポジウム, A3-3, 2012.
- 丹 寛之, 井沼 学, 大塚 玲, 米澤 祥子, 今井 秀樹, “Wolf 攻撃に対する安全性評価を目的とした人工虹彩の生成法について,”SCIS2013, 2D2-3, 2013.
- 丹 寛之, 井沼 学, 大塚 玲, 今井 秀樹, “虹彩照合アルゴリズムに対するウルフ攻撃,”SCIS2014, 3E5-2, 2014.

参考文献

[1] M.Une, A.Otsuka and H.Imai, “Wolf Attack probability: A Theoretical Security Measure in Biometrics Based Authentication Systems,”IEICE TRANS on info and Sys, Vol.E91-D, No.5, pp.1380-1389, 2008.

[2] J.Daugman, “How Iris Recognition Works,”IEEE Circuits and Sys for Video Tech, Vol.14, pp21-30, 2004.

[3] 小島由大, 繁富利恵, 美添一樹, 井沼学, 大塚玲, 今井秀樹, “虹彩認証におけるウルフ攻撃確率の理論的考察,”SCIS2008, 2008.