

不鮮明化画像が実現するスキーマを利用した画像認証方式の改良

著者	山本 匠, 原田 篤史, 漁田 武雄, 西垣 正勝
雑誌名	日本セキュリティ・マネジメント学会誌
巻	23
号	3
ページ	17-29
発行年	2009-12
出版者	日本セキュリティ・マネジメント学会
権利	(C)日本セキュリティ・マネジメント学会
URL	http://hdl.handle.net/10297/00028920

[研究論文]

不鮮明化画像が実現するスキーマを利用した 画像認証方式の改良

*An Enhanced User Authentication System
Using Schema of Visual Memory Achieved by Unclear Image*

静岡大学創造科学技術大学院 日本学術振興会特別研究員 (DC)	山本 匠
Graduate School of Science and Technology, Shizuoka University	Takumi YAMAMOTO
Research Fellow of the Japan Society for the Promotion of Science (DC)	
三菱電機株式会社	原田 篤史
Mitsubishi Electric Corporation	Atsushi HARADA
静岡大学情報学部	漁田 武雄
Faculty of Informatics, Shizuoka University	Takeo ISARIDA
静岡大学創造科学技術大学院 科学技術振興機構, CREST	西垣 正勝
Graduate School of Science and Technology, Shizuoka University	
Japan Science Technology and Agency, CREST	Masakatsu NISHIGAKI

要 旨

近年、人間の画像認識能力の高さを利用して記憶負荷を軽減させる画像認証方式が注目されている。しかしながら、画像認証方式は毎回の認証時にパス画像がディスプレイ上に表示されるため、認証時の覗き見攻撃に対して脆弱であった。この問題に対し、覗き見をする攻撃者にとってパス画像の記憶が困難となるように、モザイク化等の不鮮明化処理を施した一見すると無意味に見える画像（不鮮明化画像）をパス画像として使用する「画像記憶のスキーマを利用したユーザ認証方式」が提案されている。しかし、不鮮明な画像であっても、同じパス画像を毎回の認証で用いる限り、攻撃者にそれを覚えられる可能性が残る。この問題に対しては、正規ユーザに m 枚のパス画像を記憶させた上で、1回の本人認証にあたって n 枚 ($m > n$) のパス画像を用いて認証を行うという運用（以降、 $m \cdot n$ 対策と呼ぶ）が考えられるが、パス画像の増加にともなうユーザの負荷増大を緩和する工夫なくしてはその導入は難しい。また、一般の画像認証方式においては、囲画像（認証画面にパス画像と共に表示される複数の画像）の用意およびその頻繁な更新が難しく、不鮮明化画像を用いた方式においても、この問題は未解決のままであった。そこで本論文では、不鮮明な画像を利用する画像認証方式だからこそ実現可能な方法で、両問題の解決を図る。 $m \cdot n$ 対策導入時のユーザの負荷に対しては「パス画像を思い出すにあたっての手がかりとなる言語情報を認証時にヒントとして提示する方法」を、囲画像の用意に対しては「不鮮明化画像を加工することによって自動的に囲画像を生成する方法」を用いて本方式の改良を行う。本論文では改良方式のプロトタイプシステムを実装し、比較実験により改良方式の有効性を確認する。

キーワード

画像認証 覗き見攻撃 スキーマ 不鮮明化画像 言語手がかり 囲画像

1. はじめに

近年、人間の画像認識能力の高さを利用して記憶負荷を軽減させる画像認証方式[1-4]が注目されている。しかし、画像認証方式は毎回の認証時にパス画像が画面上に表示されるため、認証時の覗き見攻撃に対して脆弱であった。この問題に対し、覗き見をする攻撃者にとってパス画像の記憶が困難となるように、モザイク化等の不鮮明化処理を施した一見無意味な画像をパス画像として使用する「画像記憶のスキーマを利用したユーザ認証方式」（以下、基本方式と呼ぶ）が提案されている[5]。正規ユーザのみオリジナル画像を見せ、スキーマ（オリジナル画像と不鮮明化画像の間の認知構造的なリンク）[6]を学習させることにより、正規ユーザは不鮮明化画像を有意味な画像として認識できるようになり、パス画像を容易に記憶することができる。人間は画像の記憶に優れてはいるものの、それは有意味な画像を記憶する場合に限ってのことであり、無意味に見える画像を記憶することはやはり難しい[7,8]。ゆえに、他人のパス画像（不鮮明化画像）を覗き見て記憶することは、攻撃者にとって困難な作業となる。

しかし、基本方式では、毎回の認証におけるパス画像は常に同じものが使われる方式となっているため、攻撃者が覗き見た認証画面の中のパス画像がなりすましの際の認証画面にも必ず表示されることになる。不鮮明な画像であっても、同じパス画像を毎回の認証で用いる限り、攻撃者にそれを覚えられる可能性が残る。

この問題に対しては、正規ユーザに m 枚のパス画像を記憶させた上で、1回の本人認証にあたって n 枚 ($m > n$) のパス画像を用いて認証を行うという運用（以降 $m \cdot n$ 対策と呼ぶ）を導入することで、ある程度解決することが可能と考えられている。しかし、正規ユーザに一回の認証に必要なパス画像の枚数よりも多くのパス画像を記憶させることは、正規ユーザの負荷の増大につながる。そのため、 $m \cdot n$ 対策の導入に対してはユーザ負荷増大の緩和対策が必須となる[9]。

そこで本論文では、パス画像を思い出すにあたっての手がかりとなる言語情報を認証時に提示することにより、 $m \cdot n$ 対策の導入に際してのユーザ負荷の軽減を図る。本論文では本改良方式（基本方式に $m \cdot n$ 対策を導入し、さらに言語手がかりを認証時に提示するという改良を加えた方式）のことを、RVC（Recognition with Verbal Cue）方式と呼ぶことにする。RVC方式では、スキーマを有する正規ユーザは、手がかり情報によって認証時の再認および想起の促進が期待される。一方、スキーマを持たない攻撃者はこの手がかりを正規ユーザと同等には活用できないと考えられる。

また、一般の画像認証方式においては、囲画像（認証画面にパス画像と共に表示される複数の画像）の用意およびその頻繁な更新が難しく、基本方式においても、本問題は未解決のままであった。適切な囲画像を潤沢に用意することができなければ、認証システムの安全性の低下やユーザの認証時における認識負荷の増大につながる。

そこで本論文では、不鮮明化画像の特長に着目し、従来の写真や絵（オリジナル画像）を利用する画像認証方式では実現不可能な方法で、囲画像を自動生成する方式を提案する。本論文では本改良方式のことを、ADG（Automatic Decoy image Generation）方式と呼ぶことにする。

本論文では、覗き見攻撃耐性の強化および囲画像の用意という2つの異なる課題に対し、RVC方式とADG方式という2つの異なる改良方式による解決を試みる。本論文は両方式をそれぞれ独立した観点から検討した段階での報告となるが、本研究の最終目標は、両方式を併用した場合に相乗効果が発揮されるような方式へと改良を進めることにある。そこで、本論文の最後で、現時点の両方式を同時に導入した場合の状況について触れ、今後の研究の方向性と課題を確認する。

2. 基本方式のコンセプト

画像認証方式にとって覗き見攻撃が脅威となるの

は、正規ユーザのみならず覗き見攻撃者にとっても画像の記憶は容易であるからである。そこで基本方式では、覗き見をする攻撃者にとってパス画像の認識が困難となるように、モザイク化等の不鮮明化処理を施した一見無意味な画像（図 1 右）をパス画像として使用する。人間は、無意味に見える（意味を言語化できない）画像を記憶することはやはり難しい[7,8]。ゆえに、他人のパス画像（不鮮明化画像）を覗き見て記憶することは、攻撃者にとって困難な作業となる。

ただし、不鮮明化画像は、それ単体だけでは正規ユーザにも認識・記憶が困難である。そこで正規ユーザにのみオリジナル画像（図 1 左）を見せ、不鮮明化画像といっしょに記憶してもらう。不鮮明化画像にはオリジナル画像の特徴が残されているため、正規ユーザは不鮮明化画像を有意義な画像として認識できるようになり、パス画像を容易に記憶することができる。これは、不鮮明なパス画像に対する「スキーマ[6]」を正規ユーザに学習させていることに相当する。ここでスキーマとは、人間が外界からの情報を知覚した際に無意識のうちに蓄積している「その情報をどのように認識・記憶したかという知識構造」を意味する認知心理学用語である。

スキーマを認証に利用することで、不鮮明化処理を施したパス画像であっても正規ユーザは容易にこれを記憶でき、一方、スキーマを学習していない覗き見攻撃者には他人のパス画像を記憶することが困難であるという認証方式が実現できる。

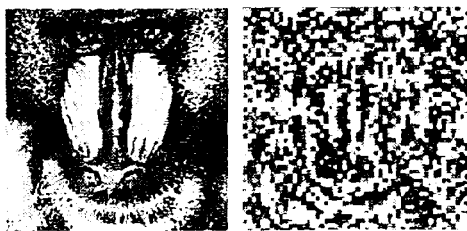


図 1 画像の不鮮明化処理

3. RVC 方式:覗き見攻撃耐性の強化

3.1 基本方式における課題

既存研究[5]において、基本方式は、既存の画像認

証方式（オリジナル画像をパス画像として利用する方式）と比べ、正規ユーザの認証成功率を高く維持したまま、攻撃耐性についても有望な結果を残している。しかし、基本方式では、毎回の認証におけるパス画像は常に同じものが使われる方式となっているため、攻撃者が覗き見た認証画面の中のパス画像がなりすましの際の認証画面にも必ず表示されることになる。不鮮明な画像であっても、同じパス画像を毎回の認証で用いる限り、攻撃者にそれを覚えられ可能性が残る。

この問題に対しては、正規ユーザに m 枚のパス画像を記憶させた上で、1 回の本人認証にあたって n 枚 ($m > n$) のパス画像を用いて認証を行うという運用 ($m \cdot n$ 対策) を導入することで、ある程度解決することが可能と考えられている。しかし、正規ユーザに一回の認証に必要なパス画像の枚数よりも多くのパス画像を記憶させることは、正規ユーザの負荷の増大につながる。そのため、 $m \cdot n$ 対策の導入に対してはユーザ負荷増大の緩和対策が必須となる。

著者らは既に、動画から複数のパス画像を抽出した上で、ユーザに動画のストーリーと複数のパス画像を合わせて記憶してもらうことで $m \cdot n$ 対策導入時のユーザの負荷を軽減する方式を提案している[9]。本論文においては、既存研究[9]とは別のアプローチによってユーザの負荷軽減を図る。具体的には、パス画像を思い出すにあたっての手がかりとなる言語情報を認証時に提示する（RVC 方式: Recognition with Verbal Cue）ことにより、基本方式に $m \cdot n$ 対策を導入したときのユーザ負荷の増大を抑制する。RVC 方式では、スキーマを有する正規ユーザは、手がかり情報によって認証時の再認および想起の促進が期待される。一方、スキーマを持たない攻撃者はこの手がかりを正規ユーザと同等には活用できないと考えられる。

3.2 RVC 方式

RVC 方式では、覗き見攻撃耐性強化のために、基本方式に $m \cdot n$ 対策が導入される。その上で、パス

画像の想起を促進するための手がかりを認証画面に提示する。手がかりには、パス画像に対応するオリジナル画像の意味を言葉で表現した文を用いる。手がかり情報が提示されない場合には、正規ユーザは、記憶している複数のパス画像のスキーマ全てを想起しなければ、スキーマに対応する画像を認証画面の中から選択することができなかつた。一方、手がかり情報を提示することで、言語手がかりで指定されたパス画像に対するスキーマのみを想起すればよくなり、ユーザの負荷が緩和される。想起が容易になれば、想起に要する時間の短縮や、想起ミスも少なくなると期待され、認証時間の短縮や認証成功率の向上が見込まれる。

ここで、画面に提示される手がかりは正規ユーザだけでなく、覗き見攻撃者にも与えられることになる。しかし、既存研究[5]の3.4節の実験から、不鮮明化画像であれば、覗き見攻撃者にパス画像の内容を言葉で伝えた場合であってもパス画像の推測成功率を低下させることができるという結果が得られている。この不鮮明化画像の特長から、認証時にパス画像に対する手がかり情報を言葉で与えたとしても、攻撃者にはその情報を有効に活用できないことが予想できる。

3.3 RVC方式の認証手順

登録時には、 m 枚のパス画像のそれぞれが、それに対する手がかり情報といっしょに提示される。正規ユーザはパス画像（不鮮明化画像）とそのオリジナル画像を見てスキーマを獲得する際に、オリジナル画像の想起の手がかりとなる言語情報も記憶することになる。登録画面例を図2および図3に示す。

認証時には、 m 枚のパス画像の内、 n 枚 ($m > n$)のパス画像が用いられる。毎回の認証ごとに n 枚のパス画像は選び直され、その都度のパス画像 n 枚を使って認証が行われる。現在認証画面に表示されているパス画像に対する手がかり情報が提示されることにより、正規ユーザにのみ効果的にパス画像の想起・再認が促され、パス画像を容易に選択すること

が可能となっている。図4に、認証画面例を示す。攻撃者が正規ユーザの認証を覗き見たとしても、次の認証で同じパス画像が現れるとは限らず、覗き見によるパス画像の推測は困難になると期待される。

なお、図5に示すように、オリジナル画像を用いた従来の画像認証方式に手がかり情報を与えた場合は、攻撃者に答え（パス画像）を教えていることと等しく、認証方式として成立し得ない。本改良が、不鮮明な画像だからこそ実現する、画期的な方式であることに注目されたい。

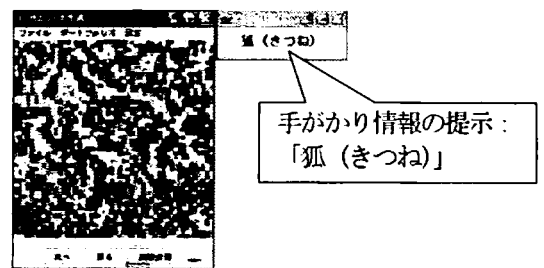


図2 登録画面(不鮮明化画像表示時)

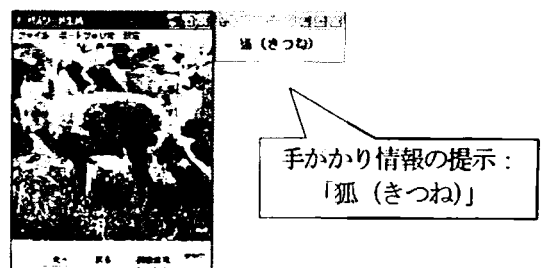


図3 登録画面(オリジナル画像表示時)

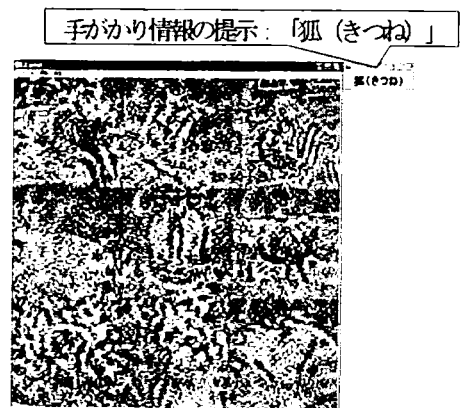


図4 認証画面



図 5 認証画面(オリジナル画像)

3.4 検証実験

本方式の有効性を、基本方式と RVC 方式との比較実験を通じて評価する。被験者は本学情報系学部学生 10 名である。

3.4.1 本人認証実験

m-n 対策の導入（記憶すべきパス画像の枚数の増加）によって増大されるユーザの負荷が、手がかりの提示によってどの程度抑えることができるのかについて、本人認証率に関する基本方式との比較実験を通じて検証する。

● 実験方法

本実験システム（RVC 方式）では、正規ユーザが記憶すべきパス画像の枚数を 10 枚とし、9 択の認証フェーズ（認証画面中にパス画像 1 枚と罫画像 8 枚が提示される）を 4 ターン行って認証可否の判定を行う。すなわち、 $m=10$ 、 $n=4$ である。認証に使用されるパス画像は、認証の都度、10 枚のパス画像セットの中から 4 枚がランダムに選択される。パス画像の手がかりは、パス画像に写っている動物の種類の名前（例：犬、馬など）とした。また、比較の

ために、本実験システムから手がかりの提示を除去したシステム（基本方式に m-n 対策のみを導入した方式。以下、比較方式 1 と呼ぶ）も構築して、同様の実験を行う。

このような実験システムを使用した理由は、基本方式の本人認証実験（既存研究[5]の 4.1 節で行われた実験）の結果と比較するためである。基本方式では、被験者（正規ユーザ）が 4 枚のパス画像を覚え、9 択×4 ターンの認証を行っている。RVC 方式（ $m=10$ 、 $n=4$ 、手がかり=有）、比較方式 1（ $m=10$ 、 $n=4$ 、手がかり=無）、基本方式（ $m=4$ 、 $n=4$ 、手がかり=無）の実験の本人認証率を比較することにより、手がかり情報の有無によるユーザの負荷の違いをパフォーマンスの尺度として調べることができる。

パス画像登録後、1 日後と 8 日後に、各被験者につき 5 回ずつ認証を行ってもらう。なおパス画像登録後、被験者は認証実験以外の場でパス画像やオリジナル画像を確認することはできない。本論文の実験で使用した画像は、様々な種類の動物が写っている背景つきの写真画像 90 枚である。

● 実験結果

実験結果を表 1 に示した。基本方式の結果は、既存研究[5]の 4.1 節の実験結果の再掲である。表中、「認証成功率」は、各認証試行において認証に成功した（1 回の認証において、4 ターンのパス画像選択全てに成功した）割合である。また、ターンごとのパス画像選択にかかった回答時間の平均を「ターンごとの平均回答時間」として記した。

1 日後、8 日後とも、RVC 方式の本人認証率は、

表 1 本人認証実験の結果

	基本方式[5] ($m=4$, $n=4$, 手がかり=無)		比較方式 1 ($m=10$, $n=4$, 手がかり=無)		RVC 方式 ($m=10$, $n=4$, 手がかり=有)	
	1 日後	8 日後	1 日後	8 日後	1 日後	8 日後
成功率	50/50 (100%)	49/50 (98%)	45/50 (90%)	41/50 (82%)	50/50 (100%)	49/50 (98%)
ターンごとの 平均回答時間 [秒]	8.19	7.10	10.85	16.48	7.30	7.18

基本方式と同様、ほぼ 100%を維持している。比較方式 1 (手がかり無) の認証率が低下している事実より、手がかり情報を用いることで、認証画面に表示されているパス画像に対するスキーマの想起が促進され、認証成功率の低下が抑えられたのだと推測できる。

平均回答時間についても、RVC 方式と基本方式はほぼ同等であり、比較方式 1 (手がかり無) ではその増加が確認できる。よって、手がかり情報により想起すべきスキーマが絞られ、囲画像の中からパス画像を見つける作業が容易になったのだと考えられる。

3.4.2 覗き見攻撃によるなりすまし実験

基本方式に $m \cdot n$ 対策を導入すれば、覗き見攻撃耐性が向上することは容易に想像ができる (攻撃者が正規ユーザの認証作業を覗き見たとしても、次の認証で同じパス画像が現れるとは限らないため)。しかし、RVC 方式では、 $m \cdot n$ 対策とともに、言語手がかりを提示するという改良を基本方式に加えている。既存研究[5]の 3.4 節の結果からもわかるとおり、攻撃者は言語手掛かりを十分活用することはできないと考えられるが、言語手がかりの提示が覗き見攻撃の脅威をどれほど増加させてしまうかについて実験により確認する必要がある。

● 実験方法

本実験システム (RVC 方式) では、正規ユーザが記憶すべきパス画像の枚数を 10 枚とし、2 択の認証フェーズ (認証画面にパス画像 1 枚と囲画像 1 枚が提示される。その際、現在認証画面に表示されているパス画像に対する言語手がかりも表示する) を 1 ターン行って認証可否の判定を行う。すなわち、 $m = 10$, $n = 1$ である。認証に使用されるパス画像は、認証の都度、10 枚のパス画像セットの中からランダムに選択される。2 択システムとした理由は、覗き見攻撃者に非常に有利な条件であっても本方式が有効であるかを測るためである。また、この設定は基本方式における覗き見攻撃の実験 (既存研究[5]の 3.3 節で行われた実験) の設定と同一である (基本方式

と RVC 方式はパス画像の枚数 m と手がかり情報提示の有無が異なるだけであり、RVC 方式は $m = 10$, $n = 1$, 手がかり = 有, 基本方式は $m = 1$, $n = 1$, 手がかり = 無) ので、両者をそのまま比較することができるというメリットもある。

覗き見攻撃耐性に対する $m \cdot n$ 対策の影響と手がかり情報提示の影響を個別に評価するためには、基本方式に $m \cdot n$ 対策のみを導入した方式 (3.4.1 節の実験における比較方式 1. 本実験においては $m = 10$, $n = 1$, 手がかり = 無)、および、基本方式に手がかり情報提示のみを導入した方式 (以下、比較方式 2 と呼ぶ。本実験においては $m = 1$, $n = 1$, 手がかり = 有) を対象とした覗き見攻撃実験についても実施すべきである。しかし、3.4.1 節の実験結果より、比較方式 1 は、本人認証成功率および認証時間において RVC 方式および基本方式に匹敵するパフォーマンスが得られないことが判明しているため、ここでは比較方式 1 に対する実験は割愛した。また、比較方式 2 ($m = 1$, $n = 1$, 手がかり = 有) の覗き見攻撃成功率は、「基本方式 ($m = 1$, $n = 1$, 手がかり = 無) の覗き見攻撃成功率 $P1$ 」と「パス画像の内容を言葉で伝えた際の基本方式に対するパス画像の推測成功率 $P2$ (既存研究[5]の 3.4 節の実験結果より 74%)」を用いて $1 - (1 - P1) \times (1 - P2)$ によって試算できることから、ここでの実験は省略した。

本実験では、実験者 (正規ユーザ) が認証フェーズにおける 2 択の選択を 1 ターン行って認証を通過する認証画面を、各被験者 (攻撃者) が間近から覗き見し、その直後に、正規ユーザへのなりすましを試みる。各被験者につき、同じパス画像セットについて 5 回ずつ認証試行を行ってもらった。i 回目の認証試行で覗き見したパス画像が、i+1 回目以降の認証試行で登場するケースが起り得るので、認証試行の回数を重ねるほど被験者は有利になっていく。既存研究[5]を参考に、各認証試行において、被験者の覗き見時間は 5 秒に設定した。

● 実験結果

実験の結果を表 2 に示した。基本方式の結果は、

既存研究[5]の 3.3 節の実験結果の再掲である。表中、「成功率」は 10 人の各被験者につき 5 回ずつ行った認証試行の全体の成功率（なりすまし成功率）を表し、「平均時間」は一回のパス画像選択に要した回答時間の平均値である。

表 2 覗き見攻撃実験の結果

	基本方式	RVC 方式
成功率	46/50 (92%)	39/50 (78%)
平均回答時間[秒]	2.66	5.17

比較方式 2 のなりすまし成功率が約 98%（基本方式のなりすまし成功率 $P1=92\%$ 、パス画像の内容を言葉で伝えた際のパス画像推測成功率 $P2=74\%$ より $1-(1-P1)\times(1-P2)$ を算出）と試算される。よって、手がかり情報提示の導入は覗き見攻撃耐性をわずかに劣化させてしまうものの、 $m\cdot n$ 対策の効果によって RVC 方式のなりすまし成功率が基本方式のそれよりも 14% 低く抑えられたことがわかる。

なお、手がかり情報の提示によって覗き見攻撃者にパス画像に関する情報がいくらか漏れてしまうという問題に対しては、今後、手がかり情報やその提示方法を工夫していくことによって改善が可能であると考えている。

例えば、「きつねの左前足をクリックしてください」などといったように言語手がかりを詳細化するような方法が考えられる。スキーマを持たない攻撃者にとって、画像中の細かい情報（向き、姿勢、各部位の位置など）まで推測することは難度が高いと考えられる。一方、不鮮明化画像の意味（スキーマ）を知っている正規ユーザにとっては、指示された場所（部位）をクリックすることは容易である。この方法は、ユーザの記憶負荷を大きく増加させることなく、パス画像選択の総当たり数を増やすことを可能にするというメリットもある。さらに、手がかりにより指定する部位を認証の度に变化させるようにすれば、ある認証フェーズで「左前足」をクリックしている瞬間を覗き見られたとしても、次回の認証においては例えば「尻尾をクリックしてくださ

い」という指示に変わるため、リプレイ攻撃防止効果を得ることもできると考えられる[10]。

4. ADG 方式: 囲画像の自動生成

4.1 画像認証における囲画像の問題

画像認証方式において、パス画像を隠すために利用される囲画像（認証画面にパス画像と共に表示される複数の画像）を適切に用意することも重要な手続きの 1 つである。

毎回の認証で常に同じ囲画像のセットを利用してしまうと、攻撃者が認証画面中の画像一枚一枚に当たりをつけ、「その画像を選択して認証に失敗したならば、その画像はパス画像ではない」というように、パス画像の候補が徐々に絞られていく問題（exhaustive-attack）がある。しかし、逆に、認証の都度、すべての囲画像を一新するようにすると、攻撃者が覗き見を繰り返すことによって、毎回の認証画面に必ず表示される画像がパス画像であると知られてしまう（intersection-attack）。

以上より、ある一定枚数の囲画像は前回の認証から引き継ぎ、残りの囲画像は正規ユーザが見たことの無い全く新しい画像を用いるという折衷案が適切と考えられる。しかし、認証の都度、一定枚数の全く新しい囲画像を準備するにあたっては、以下の問題を考慮しなければならない。

- (1) ネットワークを介して毎回囲画像をダウンロードする場合
 - (a) アクセス集中によるサーバ負荷および通信帯域消費の観点から、通信はできる限り抑えることが望ましい。
 - (b) 誰でもサーバから囲画像をダウンロードできるとした場合、攻撃者も囲画像の情報を用いて、他人のパス画像を絞り込むことが可能である。
- (2) 製品の工場出荷時に、あらかじめ大量の囲画像を記憶領域に保存しておく場合
 - (a) ユーザ数が多い場合、製品ごとに異なる「大量の囲画像」を用意することは困難である。

- (b) すべてのユーザの製品に保存する囲画像が同じであった場合には、攻撃者は、自身が購入した製品に含まれている囲画像情報を用いて他人のパス画像を絞り込むことが可能である。
- (3) ユーザが撮影した写真を利用する場合
 - (a) ユーザ自身が撮影した写真を囲画像とすると、パス画像と囲画像のどちらに対しても再認が引き起こされ、ユーザがパス画像の選択の際に混同する[2]。

そこで、本論文では、あらかじめ大量に囲画像を用意したり、ネットワークを介して自動的に囲画像を取得したりする方法とは別のアプローチにより新しい囲画像（正規ユーザにとって馴染みの無い画像）を取得する方法を検討する。本論文では、不鮮明化画像の特長に着目し、従来の写真や絵（オリジナル画像）を利用する画像認証方式では実現不可能な方法で、囲画像を生成する方式（ADG方式：Automatic Decoy image Generation）を提案する。

4.2 不鮮明化画像の特長を利用した囲画像の生成

はじめに、図6の不鮮明化画像を見てもらいたい。図6の不鮮明化画像は図1の不鮮明化画像を時計回りに90度回転（今後特に断りが無い限り、時計回りを回転方向の基準とする）させた画像である。



図6 不鮮明化画像の加工例

既に2章で図1のオリジナル画像とそれに対応する不鮮明化画像を見ているにも関わらず、図6の不鮮明化画像の意味（何が映っていて、どのような状態になっているかなど）を類推することは難しかったのではないだろうか。このように、不鮮明化画像にある細工を加えた場合、あたかも加工前の元の画像とは全く無関係な画像のように知覚される。

すなわち、正規ユーザが記憶しているパス画像や正規ユーザにとって馴染みの深い画像からでも、正規ユーザがパス画像との混乱をきたすことの無い囲画像を生成することが可能だと考えられる。一方、図7のようにオリジナル画像に対して同様の加工を行った場合、明らかに加工された画像だと認識できてしまい、これを囲画像として利用することはできないことに注意されたい。

画像の加工により囲画像を生成する方法（以下、囲画像生成法と呼ぶ）を基本方式に導入するという改良を加えることにより、ADG方式は従来の画像認証方式における囲画像の用意に関する問題を解決することができると思われる。

4.3 ADG方式における囲画像の生成手順

囲画像はパス画像を紛れさせるために用いられるものであるため、囲画像（囲画像生成法により作成された不鮮明化画像。以下、加工不鮮明化画像と呼ぶ）とパス画像（オリジナル画像を不鮮明化処理することにより得られる不鮮明化画像。以下、自然不鮮明化画像と呼ぶ）は両者の区別がつかないようにしていないといけな。すなわち加工不鮮明化画像は、自然不鮮明化画像らしさを十分保持している必要がある。

本論文では、動物を被写体とした写真を実験に用いている。そのため、動物の身体全体が写っている画像であれば、画像の下半分に足があり、画像の上半分に頭部があり、画像の中心に胴体があるという構造を持つものが多い。また、動物の顔のアップが写っている画像であれば、上半分に目があり、下半分に口があるという構造を持つものが多い。著者らが行った事前調査から、実際に多くの被験者が、これらの構造に注目することによって加工不鮮明化画像と自然不鮮明化画像の識別を試みていた。そこで今回は、上記の構造を崩さない不鮮明化画像を「自然不鮮明化画像らしさを有する画像」と考えることとする。



図 7 図 6 に対するオリジナル画像

これを考慮すると、例えば図 6 に示した「回転」は、(写真の撮り方にもよるが) 一般に画像の構造を崩すことになるため、画像を作成するための加工には適さないと考えられる。そこで以下では、動物の写真を前提とした上で、「自然不鮮明化画像らしさを有する画像」の作成が比較的期待できると考えられる 3 種類の画像生成法を示す。

1) 画像生成法 1

画像生成法 1 は、図 8 のように 2 枚のオリジナル画像 A と B のそれぞれ上半分と下半分をつなげる方法である。2 枚のオリジナル画像を組み合わせた後に、不鮮明化処理を施して加工不鮮明化画像を得る。上下の画像の境界部分の不整合を整えるために、境界部分にはグラデーションフィルタを適応する。本手法で作成した加工不鮮明化画像の例を、不鮮明化処理前の画像とともに図 9 に示す。

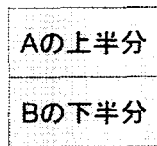


図 8 画像生成法 1



図 9 画像生成法 1 による加工不鮮明化画像例

2) 画像生成法 2

図 6 に示した「回転」は、画像の雰囲気を変化させるには効果的だと考えられる。しかし、前述のとおり、「回転」単体だけでは、自然不鮮明化画像らしさを大きく崩すと考えられる。そこで「回転」により画像の雰囲気を大きく変化させた後に、

自然不鮮明化画像らしさを補完する方法を考える。具体的には、図 10 のように正立したオリジナル画像 B に回転したオリジナル画像 A を同じ割合で重ね合わせる方法である。2 枚のオリジナル画像を重ね合わせた後に、不鮮明化処理を施して加工不鮮明化画像を得る。オリジナル画像 A の回転角度は 90 度、180 度、270 度の 3 種類である。回転したオリジナル画像 A により不自然さが增大するが、正立したオリジナル画像 B を重ね合わせることで「自然不鮮明化画像らしさ」を補うことが可能だと考えられる。本手法で作成した加工不鮮明化画像の例を、不鮮明化処理前の画像とともに図 11 に示す。

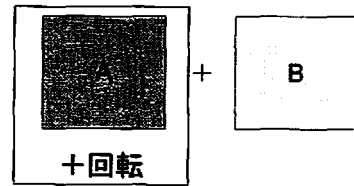


図 10 画像生成法 2

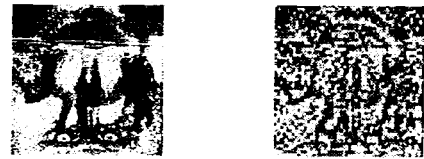


図 11 画像生成法 2 による加工不鮮明化画像例

3) 画像生成法 3

画像生成法 3 は、図 12 のように画像生成法 1 と画像生成法 2 を併用した方式である。すなわち、画像生成法 1 の要領で 2 種類のオリジナル画像を作成し、その 2 種類を画像生成法 2 の要領で重ね合わせる。最後に不鮮明化処理を施して加工不鮮明化画像を得る。本手法で作成した加工不鮮明化画像の例を、不鮮明化処理前の画像とともに図 13 に示す。

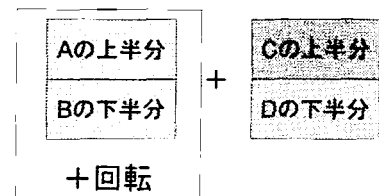


図 12 画像生成法 3



図 13 四画像生成法 3 による加工不鮮明化画像例

なお、図 9、図 11、図 13 の加工不鮮明化画像の作成に用いたオリジナル画像を図 14 に示す。

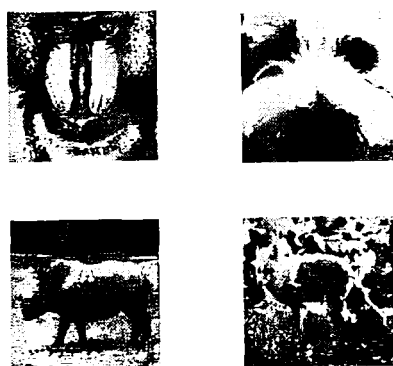


図 14 加工不鮮明化画像の作成に用いたオリジナル画像

4.4 検証実験

ADC 方式の有効性を確かめるために実証実験を行う。被験者は本学情報系学部学生 10 名である。

4.4.1 識別実験

四画像生成方法 1~3 により作成される加工不鮮明化画像が自然不鮮明化画像らしさをどの程度保持しているのか識別実験によって評価する。

● 実験方法

本実験システムは、被験者に、2 択の認証画面を提示する。2 枚の内、どちらか 1 枚が自然不鮮明化画像であり、もう 1 枚が加工不鮮明化画像である（両画像の位置は毎回ランダムに変わる）。被験者は 2 枚の画像の中で自分が直感的に自然不鮮明化画像だと思うものを選択する。四画像生成法 1 に対する識別実験を例に採り、具体的な実験手順を以下に示す。四画像生成法 2 および 3 の識別実験も同様の手順で実施される。

- (i) 5 枚のオリジナル画像 1~5 を用意する。
- (ii) 全てのオリジナル画像 i を不鮮明化し自然不鮮明化画像 $pass(i)$ ($i=1\sim5$) を作成する。
- (iii) 全てのオリジナル画像 i に対し、 i 以外の 4 枚のオリジナル画像を用い四画像生成法 1 により作成され得る全ての加工不鮮明化画像の画像セット $decoy(\wedge i,1)$ ($i=1\sim5$) を生成する^{*}。
- (iv) 1~5 の中から一つの数字 k をランダムに選ぶ。
- (v) $decoy(\wedge k,1)$ の中から任意に 1 枚の加工不鮮明化画像を選び、これと自然不鮮明化画像 $pass(k)$ による 2 択の識別実験を行う。
- (vi) k を変え、(iv)、(v) の識別実験を繰り返す。ただし、一度使用した k は選ばれない。識別実験を 5 回繰り返した時点で、5 枚全ての自然不鮮明化画像が尽くされ、実験 1 セットが終了となる。
- (vii) オリジナル画像 1~5 を一新し、(ii) ~ (vi) を計 4 セット繰り返す。すなわち、各被験者は四画像生成法 1 の識別実験につき 2 択の選択を 20 回繰り返す。

被験者は識別実験を始める前に、各四画像生成法において加工不鮮明化画像がどのように作成されているのかを図を用いて詳細に説明される。

● 実験結果

実験の結果を表 3 に示した。表中、「識別成功率」は 10 人の各被験者につき 20 回ずつ行った各四画像生成法に対する識別試行の全体の成功率（自然不鮮明化画像を正しくを選択できた割合）を表す。

表 3 識別実験の結果

	四画像生成法 1	四画像生成法 2	四画像生成法 3
識別成功率	117/200 (58.5%)	121/200 (60.5%)	115/200 (57.5%)

^{*} 例えば $decoy(\wedge 4,1)$ は、オリジナル画像 4 以外のオリジナル画像 1, 2, 3, 5 を用いて四画像生成法 1 により生成され得る加工不鮮明化画像の全てを表す。すなわち、オリジナル画像 A の上半分とオリジナル画像 B の下半分の組合せによって得られる加工不鮮明化画像を $d(A,B)$ と表すとすると、 $decoy(\wedge 4,1)=\{d(1,2), d(1,3), d(1,5), d(2,1), d(2,3), d(2,5), d(3,1), d(3,2), d(3,5), d(5,1), d(5,2), d(5,3)\}$ である。

全ての生成法において、被験者は 60%前後の割合で自然不鮮明化画像と加工不鮮明化画像との違いを認識していることが見てとれる。しかし、識別が容易な 2 択システムにおいても識別率を 60%程度に抑えることができていることから、囲画像生成法 1~3 によって作成された加工不鮮明化画像は概ね自然不鮮明化画像らしい画像となっていると考えてよいと判断できる。

ただし、識別が 50%で成功するケース（被験者が完全に当て推量で回答する場合を意味する）を帰無仮説として、各生成法における識別成功率に対して t 検定を行った結果、囲画像生成法 1~3 それぞれの有意確率は $p=0.0634$, $p=0.0109$, $p=0.0119$ となり、囲画像生成法 1 以外の 2 つの生成法において $p<0.05$ で有意差（5%有意）が見られた。よって、少なくとも囲画像生成法 2,3 によって生成された加工不鮮明化画像は、攻撃者にパス画と囲画像とを切り分けるヒントを幾分与えてしまっていることがわかる。これは、自然不鮮明化画像と加工不鮮明化画像との差を利用した推測攻撃が ADG 方式の脅威となる可能性を意味する。

現行の囲画像生成法の質が不十分であった原因として、「自然不鮮明化画像らしさを有する画像」の定義についての検討が不十分であったことが挙げられる。本論文では、4.3 節で述べているように「動物の身体全体が写っている画像」と「動物の顔のアップが写っている画像」の 2 パターンの画像にしか焦点を当てていない。これらのパターンに該当する画像は少なくないが、例えば「動物の身体全体が写っている画像」の中でも、「複数の動物の身体全体が写っている画像」もあれば、「一頭の動物の身体全体しか写っていない画像」もあるだろう。また、画像中の部位（顔や体）の位置は似ていても、空間周波数が異なる場合は、画像の印象が変わってくる。今後は、画像のパターンを細分化した上で、それぞれのパターンに応じた加工をしてやることによって、囲画像の自然不鮮明化画像らしさを維持し、推測攻撃に対する耐性を改善していく必要がある。

4.4.2 本人認証実験

本節では、パス画像から生成される囲画像を使用しても本人認証率が劣化することがないかを本人認証実験により確認する。

正規ユーザにとって馴染みのない画像（未知画像）から囲画像を生成した場合、その囲画像もまた正規ユーザにとっては馴染みのない画像であることが一般的である。よって、あらかじめ未知画像を少数枚用意しておき、それら未知画像から囲画像を生成してやれば、（実験により確かめるまでもなく）正規ユーザがパス画像と囲画像を混同することはないだろう。そこで本節では、あえて正規ユーザにとって馴染みの深い画像（パス画像）から生成された囲画像を用いて、4.3 で提案した囲画像生成法 1~3 によって生成された囲画像が正規ユーザの認証にどの程度影響を与えるか調査する。

● 実験方法

囲画像の用意の方法を除けば、本実験システム（ADG 方式）の設定は既存研究[5]の基本方式における本人認証実験と全く同一である。すなわち、正規ユーザが記憶するパス画像は 4 枚であり、9 択の認証フェーズ（認証画面中にパス画像 1 枚と囲画像 8 枚が提示される）を 4 ターン行って 1 回の認証とするシステムを用いる（ $m=4$, $n=4$ ）。

ターン毎に 4 枚のパス画像の中から 1 枚がランダムに重複無く選ばれ認証画面に表示される。パス画像と共に表示される 8 枚の囲画像は、現在表示されているパス画像以外の 3 枚のパス画像のオリジナル画像から、4.4.1 節と同じ方法で生成される。ただし、4.4.1 節の実験では、囲画像生成法 j ($j=1\sim 3$) ごとに囲画像セット($\text{decoy}(\wedge i, j)$)を用意したが、本実験では、3 つの囲画像セットを 1 つにまとめた囲画像セット($\{\text{decoy}(\wedge i, 1) + \text{decoy}(\wedge i, 2) + \text{decoy}(\wedge i, 3)\}$)の中から各 8 枚の囲画像を選出する。

パス画像登録の後、1 日後と 8 日後に、各被験者につき 5 回ずつ認証を行ってもらう。なおパス画像登録後、被験者は認証実験以外の場でパス画像やオリジナル画像を確認することはできない。

● 実験結果

実験結果を表 4 に示した。基本方式の結果は、既存研究[5]の 4.1 節の実験結果の再掲である。表中の用語は 3.4 節と同じである。

表 4 本人認証実験の結果

認証実施日	基本方式		ADG 方式	
	1日後	8日後	1日後	8日後
成功率	50/50 (100%)	49/50 (98%)	46/50 (92%)	47/50 (94%)
ターン毎の 平均回答 時間[秒]	8.19	7.10	20.60	17.673

実験結果から、たとえ、本人がスキーマを有しているパス画像（本人にとって馴染みの深い画像）から囲画像を生成したとしても、正規ユーザは自分のパス画像を高い確率で認識できていることが確認できる。

しかし、基本方式よりも成功率が若干低下していること、および、回答に倍以上の時間を要していることを考えると、パス画像を元に囲画像生成法 1~3 を用いて生成された加工不鮮明化画像を囲画像として用いることは、正規ユーザの認識負荷の増加につながってしまったことがわかる。

これは、4.3 節の囲画像生成法 1~3 によって生成された囲画像の中に、パス画像の特徴がある程度再認可能な状態で残っていたことが原因だと考えられる。この対策としては、囲画像生成法のアルゴリズムにパス画像の特徴が大きく崩れるような処理を組み込むことが有効だと考えられる。例えば、歪曲処理を使えば、エッジの連続性を失うことなく、画像全体に変化を持たせることができる。また、本節の冒頭で述べたように、実際の運用時においては、パス画像から囲画像を生成するのではなく、正規ユーザにとって馴染みのない画像（未知画像）から囲画像を生成してやれば、囲画像が正規ユーザの認証に悪影響を与えることはないだろう。

5. おわりに

本論文では、画像記憶のスキーマを利用した認証方式（基本方式）における覗き見攻撃耐性の強化および囲画像の用意という 2 つの異なる課題に対し、RVC 方式と ADG 方式という 2 つの異なる改良方式を提案し解決を試みた。

RVC 方式では、基本方式に比べ利便性（パス画像の想起・再認の容易さ）を低下させることなく覗き見攻撃耐性の強化を達成することができた。ただし、言語手がかりを利用した推測攻撃の脅威が増大していることから、言語手がかりまたはその提示方法の改良が必要である。

ADG 方式においては、正規ユーザにとって馴染みの深い画像（パス画像）から加工不鮮明化画像（囲画像）を生成したとしても、認証が機能することを示すことができた。しかし、認証に要する時間の増加が認められた上に、自然不鮮明化画像と加工不鮮明化画像の差を利用した推測攻撃の脅威が増大する可能性があることがわかった。囲画像生成法の改善が必要である。

本論文は両方式をそれぞれ独立した観点から検討した段階での報告となるが、本研究の最終目標は、両方式を併用することによって、安全性と利便性が両立した画像認証方式を実現することにある。現時点の両方式を併用した場合、ADG 方式における囲画像生成方式がまだ不十分のため、正規ユーザの認証時間を増加させてしまうことになり、RVC 方式における言語手がかりの提示によるユーザ負荷低減効果が大きく損なわれてしまう。よって、本研究の今後のステップとしては、まずは、囲画像生成法を改良し、自然不鮮明化画像と区別することが非常に困難な加工不鮮明化画像を生成する方法を実現することが急務となるだろう。

謝辞

本研究は科研費 (No.20-6290) の研究助成を受けている。また、本研究は一部、(財)セコム科学技術振興財団の研究助成を受けている。

参考文献

- [1] T. Pering, M. Sundar, J. Light, and R. Want: Photographic Authentication through Untrusted Terminals, IEEE Pervasive Computing, Vol.2, No.1, pp.30-36, (Jan 2003).
- [2] R. Dhamija, A. Perrig: Deja Vu: A User Study Using Images for Authentication, 9th USENIX Security Symposium, pp.45-58, 2002.
- [3] Real User Corporation: PassFace , <http://www.realuser.com/>(2009年9月 確認).
- [4] 高田哲司, 小池英樹: あわせ絵 画像登録と利用通知を用いた正候補選択方式による画像認証方式の強化法, 情報処理学会論文誌, Vol.44, No.8, pp.2002-2012, 2002.
- [5] 原田篤史, 漁田武雄, 水野忠則, 西垣正勝: 画像記憶のスキーマを利用したユーザ認証システム, 情報処理学会論文誌, Vol.46, No.8, pp.1997-2013, 2005.
- [6] W. F. Brewer: Schemata, In R. A. Wilson & F. C. Keil (Eds.), MIT Encyclopedia of the Cognitive Sciences, pp.729-730, 1999.
- [7] 太田信夫, 多鹿秀継 編著: 記憶研究の最前線, 北大路書房, 2001.
- [8] 松川順子: ランダム図形の命名作用と再認, 心理学研究, 54, pp.62-65, 1983.
- [9] 山本匠, 原田篤史, 漁田武雄, 西垣正勝: 画像記憶のスキーマを利用した認証方式の改良: ストーリーの利用による記憶負荷の削減, 日本セキュリティ・マネジメント学会誌, Vol.23, No.3, 2009.
- [10] 山本匠, 漁田武雄, 西垣正勝: 不鮮明化画像を利用した暗示・応答型画像認証方式の提案, 情報処理学会論文誌, Vol.50, No.9, 2009.

(受付日: 2009年6月12日)

(受理日: 2009年10月30日)

著者略歴

山本 匠 (やまもと・たくみ) 2006年静岡
大学情報学部情報科学科卒業, 2007年9月同大学大
学院修士課程修了, 現在, 同創造科学技術大学院博
士課程, 日本学術振興会特別研究員 (DC), 情報セ
キュリティに関する研究に従事.

原田篤史 (はらだ・あつし) 2001年静岡
大学情報学部情報科学科卒業, 2003年同大学大
学院修士課程修了, 2006年同博士課程修了, 同年 三
菱電機株式会社 情報技術総合研究所入社, 情報セ
キュリティに関する研究に従事.

漁田武雄 (いさりだ・たけお) 1950年生.
1976年広島大学大学院教育学研究科博士課程後期
中退, 同年広島大学教育学部助手, 1988年国立特
殊教育総合研究所研究員, 1982年静岡大学教養部
講師, 現在, 静岡大学情報学部情報社会学科教授,
文学博士, 人間の記憶の文脈依存機構の解明に関
する研究に従事, 著書等としては「目撃証言と文脈依
存記憶」(現代のエスプリ 350, 目撃者の証言: 法
律と心理学の架け橋 至文堂)等がある, 日本心理
学会会員, 日本認知心理学会会員, 日本基礎心理
学会会員, アメリカ心理学会国際会員.

西垣正勝 (にしがき・まさかつ) 1990年
静岡大学工学部光電機械工学科卒業, 1992年同大
学院修士課程修了, 1995年同博士課程修了, 日本
学術振興会特別研究員 (PD) を経て, 1996年静岡
大学情報学部助手, 1999年同講師, 2001年同助教
授, 2006年より同創造科学技術大学院助教授,
2007年より准教授, 博士 (工学), 情報セキュリ
ティ, ニューラルネットワーク, 回路シミュレー
ション等に関する研究に従事.