情報処理学会論文誌 Vol.57 No.9 1954–1964 (Sep. 2016)

Locimetric型メンタルローテーションCAPTCHA

藤田 真浩[1] 池谷 勇樹[2] 可児 潤也[2] 西垣 正勝[1,a)]

受付日 2015年12月2日, 採録日 2016年6月2日

概要：人間の高度な認知処理を利用したCAPTCHAの1つとして，Cognometric型のメンタルローテーションタスクを利用したYUNiTi CAPTCHAが提案されている．メンタルローテーションとは，1つの視点から写された2次元オブジェクトや3次元オブジェクトを頭の中で回転させ，異なる視点から写された姿形を認識する能力である．しかし，Cognometric型のYUNiTi CAPTCHAは，候補画像の中から出題画像と類似した画像を選ぶという戦略によってマルウェアに突破される可能性がある．この攻撃に対する耐性を高めるためには，正解オブジェクトと類似した囮オブジェクトを候補画像の中に多数含めておくことが肝要になるが，類似したオブジェクトの混入は，人間の正答率を低下させてしまう．そこで，Locimetric型の出題形式を採用することによって，安全性（類似画像選択攻撃に対する耐性）と利便性（人間にとってより正解容易）を備えたメンタルローテーションCAPTCHAを提案する．本論文では，提案方式を実装し，利便性に関する基礎実験を行うとともに，安全性に関する検討を行った．その結果，提案方式の利便性の低下は妥当な範囲に抑えられつつも，攻撃耐性が大きく高められたことが示された．

キーワード：CAPTCHA，メンタルローテーション，Cognometric，Locimetric

A Locimetric-based Mental-rotation CAPTCHA

Masahiro Fujita[1] Yuki Ikeya[2] Junya Kani[2] Masakatsu Nishigaki[1,a)]

Received: December 2, 2015, Accepted: June 2, 2016

Abstract: Mental-rotation is an advanced human-cognitive-processing ability to rotate mental representations of one single 2D/3D object. The YUNiTi CAPTCHA is a Completely Automated Public Turing tests to tell Computers and Humans Apart, in which a "cognometric" mental-rotation task is performed. Here, a challenge in the YUNiTi CAPTCHA is that cognometric task can be overcome by malware that simply choose the most similar image to the question image among the candidate images. An effective measure against it is to use similar decoy objects in candidate images as many as possible. However, the similar objects become highly confusable distractors for humans. To cope with this issue, this paper proposes to apply a "locimetric" mental-rotation task in the mental-rotation CAPTCHA. We developed an experimental system to confirm the usability of the proposed CAPTCHA and obtained acceptable performance. We also discussed its security.

Keywords: CAPTCHA, mental rotation, cognometric, locimetric

1. はじめに

自動プログラム（マルウェア）によって，メールアカウントの不正取得やブログへのスパムコメント書き込みといったWebサービスの不正利用が定常的に行われている．マルウェアによるWebサービスの不正利用と，人間による正規のサービス利用とを識別する技術が必要不可欠である．この要求を満たす技術の1つとしてCAPTCHA（Completely Automated Public Turing test to tell Computers and Humans Apart）が現在広く使われている．CAPTCHAは，人間には容易に正解できるが，機械には正答困難である問題をユーザに出題することで，正解できたユーザを人間だ

[1] 静岡大学創造科学技術大学院
Graduate School of Science and Technology, Shizuoka University, Hamamatsu, Shizuoka 432–8011, Japan
[2] 静岡大学大学院情報学研究科
Graduate School of Informatics, Shizuoka University, Hamamatsu, Shizuoka 432–8011, Japan
[a)] nisigaki@inf.shizuoka.ac.jp

© 2016 Information Processing Society of Japan



図 1 文字判読型 CAPTCHA の例
Fig. 1 Example of text-based CAPTCHA.

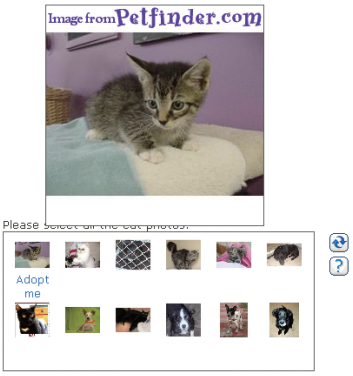


図 2 Asirra の例
Fig. 2 Example of Asirra.

と判定するチューリングテストである [1].

現在，CAPTCHA の基本形態である文字判読型 CAPTCHA が，多くの Web サービス提供サイトで利用されている（**図 1**）．文字列に歪み・ノイズを付加した形でユーザに提示し，ユーザが正しく判読できた場合は人間として，できなかった場合はマルウェアとして判別する．しかし，文字判読型 CAPTCHA は OCR（自動文字読取）を備えたマルウェアによって突破可能であることが指摘されている [2], [3]．提示する文字列の変形やノイズを増やすことで CAPTCHA の攻撃耐性を高めることはできるが，正規ユーザである人間の正答率が下がり，利便性が低下してしまう．

この問題に対し，Asirra などの画像識別型 CAPTCHA が提案された [4]．Asirra の認証画面例を**図 2** に示す．Asirra はユーザに猫画像と犬画像を合計 12 枚提示し，犬と猫を正しく識別できたユーザを人間と判別する．画像の認識は文字列の認識よりもはるかに難しい問題だと考えられていたため，マルウェアによる正答は困難であると期待されていた．しかし，機械学習を用いたプログラムによって Asirra が突破可能であるという研究報告がされた [5]．マルウェア耐性の高い CAPTCHA を実現するためには，マルウェアには依然として模倣が困難な「人間の高度な認知能力」を利用して，マルウェアが正答困難な CAPTCHA を実現する必要がある [6], [7], [8].

この課題を解決する興味深いアプローチとして，人間が有する「メンタルローテーション」の能力を巧みに利用した YUNiTi CAPTCHA が提案されている（**図 3**）[9]．メンタルローテーションとは，1 つの視点から写された 2 次元オブジェクトや 3 次元オブジェクトを頭の中で回転させ，異なる視点から写された姿形を認識する能力であり，人間が有する空間認識能力の 1 つである．3 次元の空間認識は機

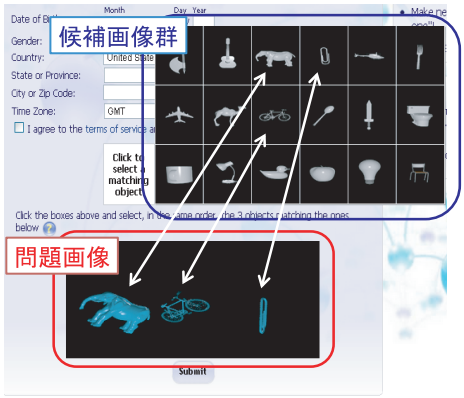


図 3 YUNiTi CAPTCHA の認証画面例
Fig. 3 Example of YUNiTi CAPTCHA.

械が苦手とする分野の 1 つであり，YUNiTi CAPTCHA はマルウェアが正答困難である理想的な CAPTCHA の 1 つとして注目を集めた [10]．YUNiTi CAPTCHA は，出題画像に写されている 3 次元オブジェクトと同一の 3 次元オブジェクトを候補画像群の中から正しく選択できたユーザを人間として判別する．出題画像と候補画像では 3 次元オブジェクトの向きが 3 次元的に異なっており，Cognometric 型（囮オブジェクトの中に含まれる正解オブジェクトを選択する方式）の出題形式となっている．

しかし，YUNiTi CAPTCHA のような Cognometric 形式の CAPTCHA の場合，マルウェアは，正解画像を推定するために「候補画像群の中から出題画像に類似した画像を選ぶ」という戦略をとることが可能である．この攻撃に対する耐性を高めるためには，正解オブジェクトと類似した囮オブジェクトを候補画像の中に多数混入することが肝要になる．しかし，類似したオブジェクトの混入は，人間の正答率の低下に直結する．

そこで本論文では，Locimetric 型（単一の 3 次元オブジェクトの中の特定部位を選択する方式）の出題形式を採用した，安全性（類似画像選択攻撃への耐性を有する）と利便性（人間にとって正解容易である）を備えた新たなメンタルローテーション CAPTCHA を提案する．

本論文の構成は次のとおりである．2 章では，メンタルローテーションについて説明した後，メンタルローテーションを用いた既存 CAPTCHA について述べる．3 章で提案方式についての詳細を述べた後，4 章で利便性に関する基礎実験を行う，5 章で提案方式の安全性に関する議論を示し，6 章で自動生成について議論する．7 章では関連研究との比較をし，最後に 8 章でまとめと今後の課題を述べる．

2. YUNiTi CAPTCHA

人間は空間認識能力が優れているため，2 次元画像から 3 次元形状を比較的容易に推測することができる [11]．また，人間は 1 つの視点から写された 2 次元オブジェクトや 3 次

© 2016 Information Processing Society of Japan

元オブジェクトを頭の中で回転させ，異なる視点から写された姿形を認識することが可能である．この能力は「メンタルローテーション」と呼ばれ，人間の高度な認知処理の一種として知られている [12], [13]．すなわち人間は，ある3次元オブジェクトが異なる視点から写された2枚の画像を見たときに，一方の画像に写っている3次元オブジェクトを頭の中で回転させ，もう一方と比較することで，2枚の画像に写っている3次元オブジェクトが同一であるか否かを判定することができる．

3次元オブジェクトのメンタルローテーションを利用したCAPTCHAとしてYUNiTi CAPTCHAが提案されている [9], [10]．YUNiTi CAPTCHAの認証画面例を図3に示す．YUNiTi CAPTCHAでは，「候補画像群の中から出題画像と同じ3次元オブジェクトが写された画像を選ぶ」というCognometric型メンタルローテーションタスクが採用されている．3問の出題画像が1度に提示され，それぞれのオブジェクトが何であるかを18個の候補画像の中から正しく選択できたユーザを人間と判定する．出題画像は毎回異なる視点から3次元オブジェクトを写した画像となっている．候補画像群の撮影方向は不変であり，つねに同一の候補画像群が表示される．

しかし，YUNiTi CAPTCHAのようなCognometric型のメンタルローテーションCAPTCHAの場合は，姿形の異なる複数のオブジェクトの中に出題画像と同一のオブジェクトが1体だけ混入する形態となっているため，「候補画像群の中から出題画像に最も類似した画像を選択する」という単純な戦略によってマルウェアにも正解画像が求められてしまう危険がある．この攻撃に対する耐性を高めるためには，正解オブジェクトと類似した囮オブジェクトを候補画像の中に多数含めておくことが肝要になる．しかし，Cognometric形式のYUNiTi CAPTCHAにおいては，類似したオブジェクトの混入は人間の正答率の低下に直結する．

3. 提案方式

3.1 コンセプト

本論文では，「単一の3次元オブジェクトの中の特定部位を選択する」というLocimetric型のメンタルローテーションタスクを採用した，新たなメンタルローテーションCAPTCHAを提案する．提案方式の認証画面例を図4に示す．認証画面は，「出題画像」（左側の画像）および「回答画像」（右側の画像）の2枚の画像から構成される．2枚の画像は同一の3次元オブジェクトを異なる視点から描画した後に線画化した画像であり，出題画像にのみマーカ（灰色の球）が表示されている．ユーザは，出題画像におけるマーカ部位が回答画像ではどこにあたるのかを回答する．人間であれば，出題画像の3次元オブジェクトを頭の中で回転させ，回答画像の3次元オブジェクトと比較することによって，回答画像における正解部位（出題画像のマーカ部位に対応する部位）を認識可能である．なお，画像生成に使用する3次元オブジェクトの種類，大きさ，視点，マーカ位置は問題生成のたびに変更する．

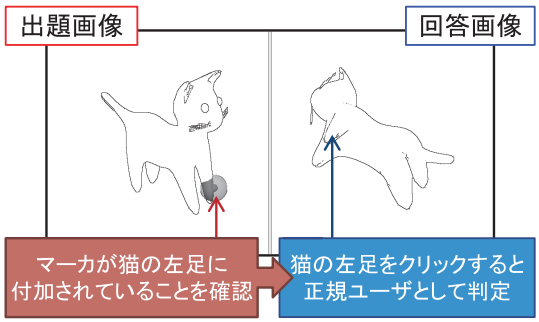


図4 提案CAPTCHAの認証画面
Fig. 4 Authentication window for proposed CAPTCHA.

単一のオブジェクトによって構成されるLocimetric型のメンタルローテーションタスクであれば，マルウェアは「最も似ている画像を探す」という戦略がとれなくなる．この結果，安全性（類似画像選択攻撃への耐性を有する）と利便性（人間にとって正解が容易である）を備えたメンタルローテーションCAPTCHAが実現されることが期待される．

Locimetric型メンタルローテーションタスクの場合は，同一の3次元オブジェクトを異なる視点から写した2枚の2次元画像（出題画像と回答画像）が提示される形になる．したがって，マルウェアはパターンマッチングや立体認識の技術を利用し，出題画像と回答画像の間の部位の対応を同定する攻撃を試みるであろう．

パターンマッチングは，領域ベースマッチングと特徴ベースマッチングに大別される [14], [15]．領域ベースマッチングは画像中の部分領域どうしをマッチングする方式であり，2次元画像の大きな変形に対する耐性が概して乏しい [14], [17]．特徴ベースマッチングは，画像中の特徴点（局所記述子）どうしをマッチングする方式であり，2次元画像の拡大・縮小・回転に対する堅牢性がある [14]．しかし，特徴ベースマッチングも，被写体の向きが3次元的に大きく異なる場合には特徴点の対応付けが難しくなる [17], [18]．そこで，提案方式では，出題画像と回答画像の間で，3次元オブジェクトをX軸，Y軸それぞれに対して45度以上の視点の回転を加えることで対策を行う．なお，視点の回転角度は，出題ごとに毎回ランダムに選ばれる．また，オブジェクトのスケールについても，出題ごとに毎回ランダムに変更する．

立体認識は，1つの3次元オブジェクトを異なる2つの視点から撮影した2枚の画像から，その3次元オブジェクトの立体形状を同定する技術である [16]．この攻撃に対し，提案方式では，前述のスケール変換に加えて，画像を線画化した状態で出題することで対策を行っている．色や

© 2016 Information Processing Society of Japan

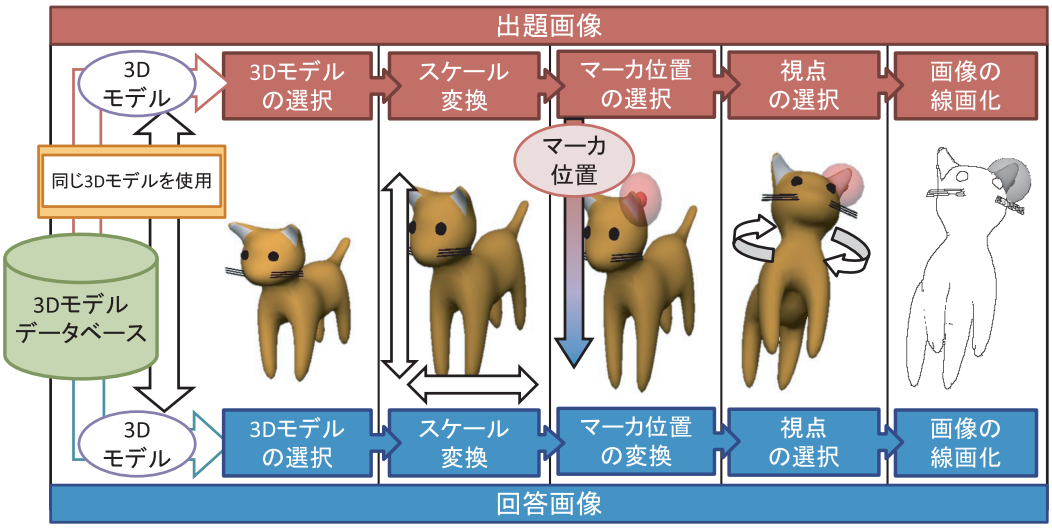


図 5 提案 CAPTCHA の自動生成手順
Fig. 5 Automatic generation procedure in proposed CAPTCHA.

影といった奥行きを知る手がかりとなる情報が取り除かれる分，3 次元画像認識の難度が高まることが期待される．

Cognometric 方式の場合は，候補画像の 3 次元オブジェクトの中に出題画像と同一の 3 次元オブジェクトが 1 体だけ存在する．上述のスケール変換，回転角度の下限，線画化などを適用しても，同一の 3 次元オブジェクトどうしの画像間の類似度は，異なる 3 次元オブジェクトどうしの画像間の類似度と比べると，概して高いといえる．すなわち，マルウェアが「出題画像と最も類似した画像を探す」という戦略をとることができる Cognometric 方式の場合は，スケール変換，回転角度の下限，線画化などの対策だけでは十分な対策効果が見込めない可能性が高いことに注意されたい．この点について著者らが検証を行った結果を付録に記す．

3.2 手順

提案方式の認証画面作成手順を図 5 に示す．なお，システムには大量のオブジェクトの 3 次元モデルが登録されていることを前提とする．以下に，手順の詳細を示す．

①システムは，出題画像と回答画像に利用する 3 次元モデルを任意に選ぶ．
②システムは，① で選んだ 3 次元モデルにランダムにスケール変換を施し，出題用オブジェクトを生成する．
③同様に，システムは，① で選んだ 3 次元モデルにランダムにスケール変換を施し，回答用オブジェクトを生成する．
④システムは，出題用オブジェクトに対してマーカの部位をランダムに選ぶ．
⑤システムは，④ で選んだマーカの位置に対応する回答用オブジェクトの部位を求める．
⑥システムは，出題用オブジェクトの視点をマーカが視認できる範囲で任意に選ぶ．
⑦同時に，システムは，回答用オブジェクトの視点を，⑥ で選ばれた視点から X 軸・Y 軸ともに 45 度以上異なる範囲から任意に選ぶ．
⑧システムは，出題画像を描画したうえで線画化する．ただし，マーカ部分は線画化せずにグレースケール変換を行う．
⑨同時に，システムは，回答画像を描画したうえで線画化する．回答画像にもマーカは付加されているが，マーカ自体は描画されない．
⑩システムは，出題画像と回答画像を表示する．
⑪ユーザは，回答画像において「出題画像内のマーカ部位（④ で選ばれた部位であり，⑤ で求めた部位）」を回答する．
⑫システムは，正答できたユーザを人間，正答できなかった人間をマルウェアと判定する．

提案方式においては，回答画像にはマーカが描画されていない．したがって，マルウェアは出題画像と回答画像の情報だけを用いて，回答画像におけるマーカ部位を特定しなければならない．これに対し，システムは自動生成の過程で回答画像におけるマーカの位置を知っている．これが「落とし戸」となり，システム（機械）が「マルウェア（機械）には認識できない問題」を自動生成し，かつ，システム（機械）自身が回答に対する正解判定を可能としている．

システムに大量の 3 次元モデルを登録しておき，使用する 3 次元モデル，伸縮率，マーカの位置，視点の位置を，認証のたびにランダムに選ぶことで，ほぼ無数の問題を自動生成することが可能である．

3.3 実装

3.3.1 仕様

提案方式の基礎実験を行うため，実験システムの実装を

© 2016 Information Processing Society of Japan

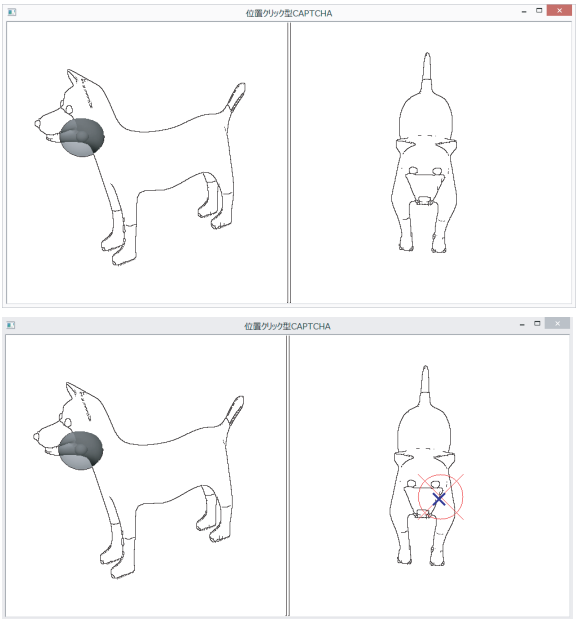


図 6 実験システムの認証画面例（上：回答待機時，下：回答後）
Fig. 6 Authentication window for experiment system (Top: before user's click, Bottom: after user's click).

行った．**図 6** に実験システムの認証画面例を示す．図 6（上）はユーザからの回答を待機している状態の画面であり，図 6（下）はユーザによる回答後の画面である．ユーザは，出題画像中のマーカ部位（灰色の球）が回答画像上のどの位置となるかを同定し，マウスクリックによって回答する．ユーザがクリックした箇所と正解部位の位置の距離が閾値以下であれば認証成功とした．図 6 の例では，出題画像においてマーカが犬の口元に付いているため，回答画像における犬の口元をクリックすれば正解となる．ユーザのクリック後，図 6（下）に示したようにクリック位置（×印）と正解範囲（○印）を表示するとともに，認証の成否と所要時間をユーザに知らせた．

3.3.2 画像生成に関する制約

提案方式は，画像の生成にあたって，問題画像のサイズ，マーカの大きさ，ならびに，視点についていくつかの制約が存在する．これらの制約に関するパラメータについては，システム実装にあたって予備実験を行い，経験的に適切な値を定めた．以下に，それぞれの詳細について述べる．

(1) 画像サイズ

出題画像と回答画像の画像サイズは，縦 500 画素 × 横 500 画素とした．左上が $(0, 0)$ 画素，右下が $(499, 499)$ 画素である．

(2) スケール変換

出題画像および回答画像の 3 次元オブジェクトをスケール変換（3.2 節の手順 ②③）する目的の 1 つが，マルウェアに対する解読耐性向上である．今回の実装では，X 軸，Y 軸，Z 軸に対してそれぞれ独立に任意の倍率で伸縮を行った．ただし，オブジェクトが大きくなりすぎ（認証画面からはみ出しすぎ）たり，小さくなりすぎたりしないように，スケール変換の倍率の範囲は 1.0 から 1.5 の間に制限した．

(3) マーカ

回答用オブジェクトにおけるマーカ部位の中心が正解座標となる．ここで，正解座標は 3 次元データであるのに対し，ユーザによるマウスクリックは（ディスプレイ上の座標情報として得られるため）2 次元データである．このため，3 次元オブジェクト上の正解座標がディスプレイ上ではどの座標にあたるかを計算したうえで，2 次元正解座標とクリックされた座標との距離によって正解判定を行っている．正解範囲は，正解座標を中心とした円の内部であり，今回の実装では 40 画素を半径とした．

(4) 視点

出題画像の視点の選択（3.2 節の手順 ⑥）において，マーカが視認できなくなる視点が選ばれた場合，正答困難な問題が生成されてしまう．このため，今回の実装では，出題画像の視点は，マーカが付加されている部位が手前側に表示されるような制約を追加した．

回答画像の視点の選択（3.2 節の手順 ⑦）において，「出題画像の視点」に近い視点が選ばれた場合，出題画像と回答画像が類似した画像となる確率が高まり，両方の画像を比較することでマルウェアがマーカ部位を解読できる危険性が生じる．3.1 節で述べたように，今回の実装では，回答画像の視点は，出題画像の視点から X 軸および Y 軸に対して 45 度以上離れた角度の中からランダムに選ばれるような制約を追加することによって，この問題に対応している．ただし，回答画像中の正解座標が認証画面からはみ出てしまう場合は，視点の再選択を行った．

(5) 画像の線画化

出題画像および回答画像を線画化した状態でユーザに提示する目的の 1 つが，マルウェアに対する解読耐性の向上である．今回の実装では，マーカの視認性に配慮し，マーカ部分はグレースケール変換を行った．

4. 利便性に関する評価実験

4.1 目的

YUNiTi CAPTCHA を再現した CAPTCHA（以下，YUNiTi 型 CAPTCHA），および，類似モデルを含む YUNiTi 型 CAPTCHA の実験システムについて実装する．提案方式とこれらの CAPTCHA を正答率と回答時間の観点から比較することで，正規ユーザ（人間）にとって「提案方式が，類似モデルを含む YUNiTi 型 CAPTCHA と同程度以上に正解可能であること」を確認する．

4.2 諸元

本実験の被験者は情報系大学生 20 名である．各被験者に，YUNiTi 型 CAPTCHA と提案方式をそれぞれ 5 問連続して解いてもらった．被験者は，先に YUNiTi 型 CAPTCHA

© 2016 Information Processing Society of Japan

を解いた後に提案方式を解くグループ α と，逆の順番で CAPTCHA を解くグループ β に分かれて実験を行った．なお，本番の回答にとりかかる前に，どちらの CAPTCHA も被験者が満足するまで練習を行うことを許した．2 種類の CAPTCHA の回答後，さらに，すべての被験者に，候補画像群に「類似モデルが含まれる場合の YUNiTi 型 CAPTCHA」を 5 問連続して解いてもらった．ただし，グループ α もグループ β も，被験者は，類似モデルを含む YUNiTi 型 CAPTCHA の実験の前に，類似モデルを含まない YUNiTi 型 CAPTCHA の実験を済ませている．よって，類似モデルを含む YUNiTi 型 CAPTCHA の実験においては練習のフェーズは割愛した．

4.2.1 提案方式

提案方式の実験システムは 3.3 節で実装したシステムである．本実験では，10 種類の 3 次元モデル（A～J）を使用する．今回使用した 3 次元モデルは，すべて動物（哺乳類，鳥類，爬虫類）で統一した．練習では，モデル A～E をランダムに使って問題生成する（被験者は，練習を繰り返すうちに，同じモデルに関する問題を複数回目にすることがありうる）．本番では，モデル F～J をランダムな順序で 1 回ずつ使って問題を 5 問生成する（被験者は，5 種類のモデルに関する問題を 1 回ずつ目にする）．スケール変換および視点については，3.3 節で説明した制約の下，毎回ランダムに選ばれる．なお，提案方式における出題画像と回答画像の視点の選ばれ方は，被験者には知らせていない．今回の実験では，回答（クリック位置），回答の正否，所要時間を記録した．各回答の結果は被験者に毎回表示した．

4.2.2 YUNiTi 型 CAPTCHA

YUNiTi 型 CAPTCHA の認証画面例を**図 7** に示す．オリジナルの YUNiTi CAPTCHA は 18 枚の候補画像群の中から 3 種類の正解画像を回答させる形態である．しかし，5 種類のモデルのみを利用する提案方式と実験条件を一致させるために，「5 枚の候補画像群（図 7 における上段の 5 枚の画像）の中から 1 枚の出題画像（図 7 における左下の 1 枚の画像）に相当する画像を同定するタスク」を 1 問として扱う．使用する 3 次元モデルも，提案方式の実験システムと同じ 10 種類のモデル（A～J）である．練習では，モデル A～E を候補画像群として用い，その中から正解となるモデルをランダムに選んで問題を生成する．本番では，モデル F～J を候補画像群として用い，その中から正解となるモデルをランダムな順序で 1 回ずつ使って問題を 5 問生成する．各候補画像と出題画像の画像サイズは，どちらも縦 160 ピクセル × 横 160 ピクセルである．視点については毎回ランダムに選ばれるが，オリジナルの YUNiTi CAPTCHA の仕様に合わせてスケール変換は行っていない．表示される画像はすべてグレースケール画像である．今回の実験では，回答（選択した画像），回答の正否，所要時間を記録した．各回答の結果は被験者に毎回表示した．

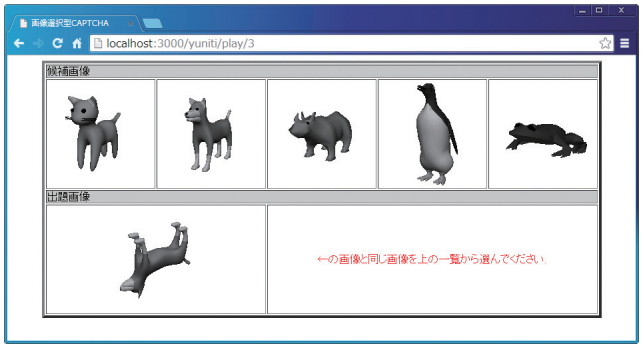


図 7 YUNiTi 型 CAPTCHA の認証画面例
Fig. 7 Authentication window for YUNiTi-type CAPTCHA.

4.2.3 類似モデルを含む YUNiTi 型 CAPTCHA

YUNiTi 型 CAPTCHA は Cognometric 型メンタルローテーションタスクであるため，攻撃耐性を向上させるためには，正解オブジェクトと類似したオブジェクトを候補画像群の中に複数含めておくことが肝要になる．この状況をシミュレートするための実験システムが「類似モデルを含む YUNiTi 型 CAPTCHA」である．具体的には，相異なる 3 体の 3 次元モデル A，B，C と 2 体の類似した 3 次元モデル K，L（**図 8**）を 5 枚の候補画像群として用いる形で図 7 の実験システムを運用している．類似モデルを含む YUNiTi 型 CAPTCHA の実験においても問題を 5 問生成するが，モデル K または L のいずれかが必ず正解となるように全問題を生成した．極論すると，類似したオブジェクトを選択する攻撃に対する耐性を備えるためには，すべての候補画像を正解オブジェクトと類似したオブジェクトにしなければならないが，今回は被験者の利便性についても配慮して「5 枚の候補画像群の中に回答画像と類似する画像が 2 枚混在する」という実験設定としている．その他の実験条件は類似モデルを含まない YUNiTi 型 CAPTCHA と同一である．

4.3 実験結果

提案方式，YUNiTi 型 CAPTCHA，類似モデルを含む YUNiTi 型 CAPTCHA の実験結果を**表 1** に示す．

表 1 より，全被験者の平均正答率（被験者 20 人が各 5 問ずつ行った全 100 試行の成功確率）は，提案方式では 80%，類似モデルを含まない YUNiTi 型 CAPTCHA では 100%，類似モデルを含む YUNiTi 型 CAPTCHA では 68%であ



図 8 類似モデルの候補画像
Fig. 8 Candidate images for similar 3D models.

© 2016 Information Processing Society of Japan

表 1 実験結果
Table 1 Experimental results.

被験者	提案方式		YUNiTi 型 CAPTCHA			
			類似モデルなし		類似モデルあり	
	正答率	平均時間[s]	正答率	平均時間[s]	正答率	平均時間[s]
1	5/5	6.7	5/5	2.0	3/5	11.9
2	4/5	6.6	5/5	1.9	3/5	7.2
3	3/5	5.6	5/5	1.4	5/5	7.1
4	4/5	3.0	5/5	1.5	4/5	3.0
5	4/5	2.2	5/5	1.3	4/5	2.3
6	4/5	2.3	5/5	2.1	3/5	2.8
7	4/5	3.2	5/5	1.6	3/5	3.6
8	5/5	9.6	5/5	2.8	3/5	9.9
9	4/5	3.6	5/5	2.1	3/5	4.5
10	5/5	3.3	5/5	3.2	4/5	3.8
11	5/5	8.3	5/5	2.8	4/5	9.3
12	5/5	2.9	5/5	2.3	2/5	4.5
13	1/5	6.7	5/5	2.6	2/5	4.6
14	4/5	5.4	5/5	3.1	5/5	3.4
15	5/5	10.2	5/5	2.4	3/5	4.9
16	2/5	3.6	5/5	2.6	3/5	3.5
17	3/5	6.5	5/5	3.8	4/5	6.5
18	4/5	3.2	5/5	2.4	4/5	3.4
19	4/5	6.2	5/5	2.8	3/5	5.5
20	5/5	3.6	5/5	2.2	3/5	4.5
平均	80%	5.1	100%	2.3	68%	5.3

る．全被験者の 1 問あたりの平均所要時間は，提案方式では 5.1 秒，類似モデルを含まない YUNiTi 型 CAPTCHA では 2.3 秒，類似モデルを含む YUNiTi 型 CAPTCHA では 5.3 秒である．この結果から，提案方式が採用している Locimetric 型のメンタルローテーションタスクは，類似モデルを含まない状況であれば，YUNiTi CAPTCHA で用いられている Cognometric 型のメンタルローテーションタスクよりも難しいことが分かる．しかし，類似モデルが含まれた状況においては，Cognometric 型よりも Locimetric 型メンタルローテーションタスクのほうが正答率が高く，所要時間についてはどちらもほぼ同じ結果となっている．

類似画像選択攻撃に対する攻撃耐性を有することを要件とした場合，提案方式と類似オブジェクトを含む YUNiTi 型 CAPTCHA の比較となるが，今回の実験から，正規ユーザ（人間）にとって「提案方式が，類似モデルを含む YUNiTi 型 CAPTCHA と同程度以上に正解可能であること」が確認できた．さらに，類似モデルを含む YUNiTi 型 CAPTCHA に対して，提案方式と同様にスケール変換や線画化といった攻撃耐性向上策を講じた場合には，正答率

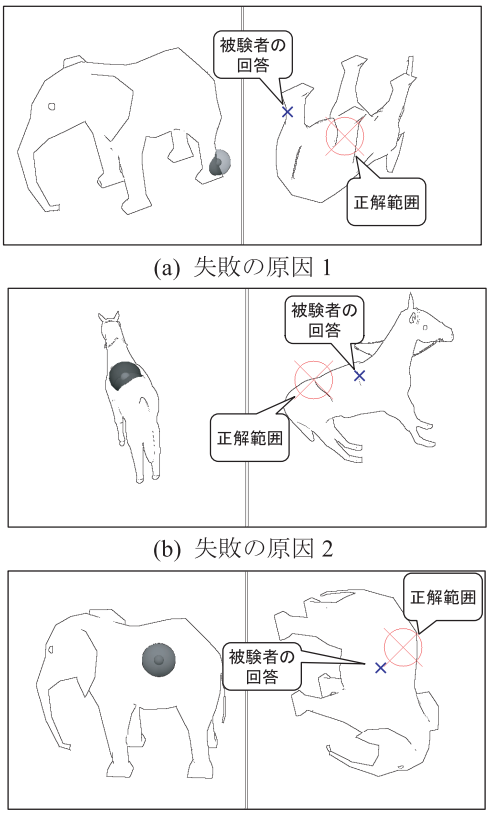


図 9 失敗の原因
Fig. 9 Reasons of failures.

の低下や所要時間の増加が予想される．また，提案方式において，被験者が認証に失敗した試行には，以下の 3 つの原因が見られた．これらの原因に対策を行うことで，提案方式の正答率を向上させる余地は多分にあると思われる．

1 つ目の原因は 3 次元オブジェクトの左右の誤認識である．具体的には，動物の左後足にマーカが付加されている出題画像に対して，左右を混同して回答画像の右後足をクリックしてしまった例があった（**図 9** (a) が実際の失敗例）．ユーザが左右を意識してメンタルローテーションを行うようになれば，この間違いは少なくなることが期待できる．

2 つ目の原因は奥行きが認識しにくい視点の存在である．具体的には，出題画像において，3 次元オブジェクトの真正面や真後ろからの視点が選ばれた場合に，マーカが表示されている位置の奥行方向の認識が困難であった試行がいくつか見られた（図 9 (b) が実際の失敗例）．この問題に対しては，真正面や真後ろの視点を選ばないように，出題画像の視点に新たな制限を追加することで対策可能である．ただし，不適切な視点は 3 次元オブジェクトの形状に応じて異なる可能性もあるため，さらなる調査を行う必要がある．

3 つ目の原因は正解部位が隠れてしまう視点の存在である（図 9 (c) に実際の失敗例を示す）．具体的には，回答画像において，3 次元オブジェクトの正解部位とは反対側の視点が選択された場合，クリックすべき位置が隠れてし

© 2016 Information Processing Society of Japan

まった試行がいくつか見られた．正解部位が必ず見える視点を選択することは可能であるが，その制約が攻撃に有利に働く可能性があるかもしれない．この対策については，今後模索する必要がある．

5. 安全性に関する考察

5.1 ブルートフォース攻撃

今回の提案方式の実装では，縦 500 ピクセル × 横 500 ピクセルの回答画像中に様々な動物が表示される．このうち，ブルートフォースの攻撃対象となるのは実際に動物が描画されたエリアの面積であり，3.3 節で実装したシステムにおいてこの大きさを実測したところ，平均しておよそ 50,000 平方ピクセルであった．これに対し，正解判定の閾値を半径 40 ピクセルの円（面積は約 5,000 平方ピクセル）と設定したため，単純計算すると総当たり数は約 10 通りとなる．

ただし，評価実験（4 章）の結果を分析したところ，もし正解判定の閾値を半径 35 ピクセルの円（面積は約 4,000 平方ピクセル）に設定していたとしても，正答率は低下せずに 80%であったことが確認できた．この場合，総当たり数は約 12 通りとなる．さらに，もし半径 30 ピクセルの円（面積は約 3,000 平方ピクセル）に設定したならば，総当たり数は約 17 通りとなり，正答率は 72%であった．

4.3 節で述べた被験者の失敗原因への対策は「ユーザにマーカの位置をより正確に伝えること」にも貢献すると期待できるため，将来的には正答率を維持したまま正解判定の閾値の円を小さくすることができるであろう．このため，提案方式も，少なくとも，オリジナルの YUNiTi CAPTCHA（18 種類の候補画像から 1 枚を選択）と同程度の総当たり数（1 問あたり 18 通り）は確保できる見込みが高い．

それ以上の総当たり攻撃対策については，単純には，（オリジナルの YUNiTi CAPTCHA が 3 問 1 組の問題形式を採用しているように）1 回あたりの問題数を増やす方法が考えられる．しかし，問題数の増加は利便性の減少とトレードオフとなる．利便性を維持したまま提案方式の総当たり攻撃耐性を向上させる工夫を検討する必要がある．

5.2 パターンマッチング攻撃および 3 次元形状復元攻撃

Locimetric 型メンタルローテーションタスクの場合は，出題画像と回答画像は同一の 3 次元オブジェクトを異なる視点から写した 2 枚の画像であるため，マルウェアはパターンマッチングや立体認識の技術を利用し，出題画像のオブジェクトと回答画像のオブジェクトの部位の対応を同定する攻撃を試みるであろう．3.1 節で説明したように，提案方式では，領域ベースマッチングや特徴ベースマッチングを利用した攻撃については 3 次元オブジェクトのスケール変換および視点の変更を行うことで，立体認識技術を利用した攻撃についてはスケール変換および線画化を行うことで，それぞれ対策を行っている．

また，パターンマッチングや立体認識の技術は，マルウェアだけでなく提案方式の強化のためにも活用することができる．すなわち，生成された出題画像と回答画像に対してパターンマッチングや立体認識を適用し，マルウェアによる解読の恐れがある画像については，システムがこれを事前に破棄することが可能である．これにより，パターンマッチング攻撃や 3 次元形状復元攻撃を高い確率で無効化することが期待できる．

5.3 その他の攻撃

提案方式に対する攻撃手法のうち，典型的なものについては，前節までに考察した．しかし，マルウェアによる攻撃手法は多様であり，提案方式の解読耐性が理論的に証明されているわけではない．特に，線画からの立体復元 [19], [20] やデータベース攻撃（攻撃者があらゆる三次元モデルを入手したうえで，その知識を使って正解部位を推測する攻撃）については提案方式の深刻な脅威となりうる可能性がある．立体復元およびデータベース攻撃は YUNiTi CAPTCHA にも共通の脅威であるため，YUNiTi との比較に焦点を当てた本論文においては詳細な検討を割愛するが，今後さらなる分析を行う必要がある．

6. 自動生成に関する考察

提案方式では，3.2 節に示した手順のとおり，3 次元コンピュータグラフィックス技術を利用して毎回新しい出題画像を容易に生成することができる．3 次元モデルを利用した Web サービスは近年急激に増加しており，将来的には大量の 3 次元モデルが世の中に出回ることが予想される．したがって，Web 上から収集した多数の 3 次元モデルをシステムに登録しておき，使用するオブジェクト，ならびに，オブジェクトのパラメータ（サイズや回転角度）を変更することによって，出題画像を無数に生成することが可能となる．ただし，ボールや丸椅子のような 3 次元モデルや，透明な部分を持つ 3 次元モデルは，マーカの部位が特定できず，回答不能となるため利用することができない．しかし，このような 3 次元モデルは，モデルの頂点データや色データから識別することが可能であるため，提案方式に適したモデルを自動収集することは十分に現実的である．

7. 関連研究

SKETCHA [21] は，2 次元のメンタルローテーションを利用したメンタルローテーション CAPTCHA である．3 次元モデルを 2 次元画像へ投影し，その 2 次元画像に線画化と回転（0，90，180，270 度）を施したうえでユーザに提示する．提示画像をユーザが 1 回クリックするごとに，2 次元画像が 90 度回転し，画像を直立状態（0 度の回転）

© 2016 Information Processing Society of Japan

に戻すことができたユーザを正規ユーザとして判定する．Cognometric 型（囮オブジェクトの中に含まれる正解オブジェクトを選択する方式）の出題形式となっていないため類似した画像を選択する攻撃に耐性を有する点，線画化を利用することでマルウェアに対する解読耐性を高めている点は，提案方式と同等のアドバンテージである．一方で，90 度単位の回転であるため 1 問あたりの総当たり数が小さい（4 通り）点，問題画像の自動生成のためにはすべての 3 次元モデルに対して「どちらが上か」という情報を付与する必要がある点については，提案方式（や YUNiTi CAPTCHA）と比べて改良の余地が残っている．

田中らは，YUNiTi CAPTCHA の問題画像や候補画像に写るオブジェクトが視点によっては認識しにくいことを指摘している [22]．問題画像や候補画像に写るオブジェクトを複数方向から連続して撮影し，アニメーション化することによって，この問題の解決を試みている．アニメーション化によって，複数の視点からオブジェクトを写した画像を表示させることが可能となる．これは，正規ユーザのユーザビリティ向上の観点からは興味深いアプローチであるが，マルウェアに対しても解読のための情報をより多く与えてしまうため，類似した画像を選択する攻撃による解読に対する耐性の低下が懸念される．

8. まとめと今後の課題

本論文では，Locimetric 型（単一の 3 次元オブジェクトの中の特定部位を選択する方式）のメンタルローテーションタスクを採用した，新たなメンタルローテーション CAPTCHA を提案した．Locimetric 方式を採用することによって，安全性（類似画像選択攻撃に対する耐性）と利便性（人間にとってより正解容易）を有するメンタルローテーション CAPTCHA が実現される．提案方式の実装，利便性に関する評価実験，攻撃耐性に関する考察を行い，本方式の有用性を示した．

今後の課題として，視点の選択範囲の検討（マーカ位置の認識がより容易となるような視点を選択することによって，正規ユーザの正答率が向上する），正解判定範囲の検討（範囲が大きいほど正規ユーザの正答率は向上するが，ブルートフォース攻撃に対して脆弱となる），提案方式の攻撃耐性に関する理論的評価などがあげられる．今回の評価実験の結果を参考にして，これらの検討を進めていきたい．また，出題画像生成時にパターンマッチングや立体認識技術を活用し，マルウェアに対して脆弱な出題画像をあらかじめ除去する方法についても検討予定である．

謝辞 静岡産業大学漁田武雄教授には，メンタルローテーションに関してご教授いただきました．本研究は JSPS 科研費 JP25280046 の助成を受けました．本論文の評価実験で使用した 3 次元オブジェクトは，メタセコ素材!（http://sakura.hippy.jp/meta/）ならびに TurboSquid（http://www.turbosquid.com/）などで公開されている無料素材を利用させていただきました．御礼申し上げます．

参考文献

[1] The Official CAPTCHA Site, available from ⟨http://www.captcha.net⟩ (accessed 2014-12-04).
[2] Yan, J. and El Ahmad, A.S.: Breaking Visual CAPTCHAs with Naïve Pattern Recognition Algorithms, *Proc. ACSAC2007*, pp.279–291 (2007).
[3] Elson, J., Douceur, J., Howela, J., et al.: Asirra: A CAPTCHA that exploit interest-aligned manual image categorization, *Proc. ACM CCS 2007*, pp.366–374 (2007).
[4] ASIRRA – Microsoft Research, available from ⟨http://research.microsoft.com/en-us/um/redmond/projects/asirra/⟩ (accessed 2014-12-04).
[5] Golle, P.: Machine Learning Attacks Against the ASIRRA CAPTCHA, *Proc. ACM CCS 2008*, pp.535–542 (2008).
[6] Chellapilla, K., Larson, K., Simard, P.Y., et al.: Computers beat humans at single character recognition in reading-based Human Interaction Proofs (HIPs), *Proc. 2nd Conference on Email and Anti-Spam* (2005).
[7] Yamamoto, T., Tygar, J.D. and Nishigaki, M.: Captcha using strangeness in machine translation, *Proc. AINA 2010*, pp.430–437 (2010).
[8] Yamamoto, T., Suzuki, T. and Nishigaki, M.: A Proposal of Four-panel cartoon CAPTCHA, *Proc. AINA 2011*, pp.159–166 (2011).
[9] YUNiTi, available from ⟨http://www.yuniti.com/⟩ (accessed 2014-12-04).
[10] CNET: 3D-based Captchas become reality, available from ⟨http://www.cnet.com/news/3d-based-captchas-become-reality/⟩ (accessed 2014-12-04).
[11] Stafford, T. and Webb, M.: *Mind hacks: Tips & tricks for using your brain*, O'Reilly Media, Inc. (2004).
[12] Shepard, R.N. and Cooper, L.A.: *Mental images and their transformations*, The MIT Press (1986).
[13] Shepard, R.N. and Metzler, J.: Mental rotation of three dimensional objects, *Science*, *New Series*, Vol.171, No.3972, pp.701–703 (1971).
[14] 伊藤康一，高橋 徹，青木孝文：高精度な画像マッチング手法の検討，第 25 回信号処理シンポジウム，pp.547–552 (2010).
[15] 藤吉弘亘，安倍 満：局所勾配特徴抽出技術：SIFT 以降のアプローチ，精密工学会誌，Vol.77, No.12, pp.1109–1116 (2011).
[16] Hartley, R. and Zisserman, A.: *Multiple view geometry in computer vision*, Cambridge University Press (2003).
[17] 熊沢逸夫：コンピュータビジョンの基礎となる対応点問題をめぐって，映像情報メディア学会誌，Vol.60, No.3, pp.313–320 (2006).
[18] 金沢 靖，金谷健一：2 画像間の特徴点対応の自動探索–シーンに関する知識を上手に使う，画像ラボ，Vol.15, No.11, pp.20–23 (2004).
[19] 小林孝至，西村 治，角所 考，淡誠一郎，北橋忠宏：単一手書き線画に基づく大まかな 3 次元形状伝達のための立体復元，電気学会論文誌 C，Vol.116, No.9, pp.998–1006 (1996).
[20] 五十嵐健夫：スケッチインタフェースの研究動向，コンピュータソフトウェア，Vol.23, No.4 (2007).
[21] Ross, S.A., Halderman, J.A. and Finkelstein, A.: Sketcha: A captcha based on line drawings of 3D models,

© 2016 Information Processing Society of Japan

Proc. WWW 2010, pp.821–830 (2010).
[22] 田中知樹，児玉英一郎，王 家宏，高田豊雄：物体認識能力に着目した三次元物体アニメーション CAPTCHA の提案，情報処理学会第 77 回全国大会，6X-02 (2015).

付 録

A.1 スケール変換・回転角度の下限・線画化の対策を施した YUNiTi CAPTCHA の類似画像選択攻撃に対する耐性

A.1.1 目的

3.1 節で，“マルウェアが「出題画像と最も類似した画像を探す」という戦略をとることができる Cognometric 方式の場合は，スケール変換，回転角度の下限，線画化などの対策だけでは十分な対策効果が見込めない可能性が高い”と述べた．この点について著者らが検証を行った結果を記す．

A.1.2 実験諸元

5 体の 3 次元モデル（3.3 節で実装したシステムで使用したモデル A～E）を利用して，候補画像 5 枚（S1～S5）と問題画像 5 枚（T1～T5）を作成した．候補画像 S1～S5 は，モデル A～E をそれぞれ x 軸に対して反時計回りに 10 度，y 軸に対して時計回りに 30 度回転したうえで，線画化（3.3.2 項 (5)）を適用することによって生成されている（**図 A·1** 右）．問題画像 T1～T5 は，モデル A～E をそれぞれ x 軸に対して反時計回りに 10 度，y 軸に対して反時計回りに 15 度回転した（これによって，候補画像 Si と問題画像 Ti は，それぞれ y 軸に対して 45 度回転した画像となる）うえで，x，y，z 軸方向にそれぞれランダムに 1.0～1.5 倍のスケール変換（3.3.2 項 (2)）と線画化（3.3.2 項 (5)）を適用することによって生成されている（図 A·1 左）．

「候補画像 S1～S5 の中から各問題画像 Ti（i = 1～5）に一番近い画像を選択する」というタスクを構成することによって，YUNiTi にスケール変換，回転角度の下限（45 度），線画化を適用した CAPTCHA をシミュレートすることができる．各問題画像 Ti（i = 1～5）に対して，パターンマッチングを用いて，候補画像 Sj（j = 1～5）の中から問題画像 Ti に一番近い画像を選択する．Ti に対して Si が選ばれる確度が高ければ，「YUNiTi にスケール変換，回転角度の下限，線画化を適用した CAPTCHA」は類似画像を選択するパターンマッチング攻撃に脆弱であるということになる．

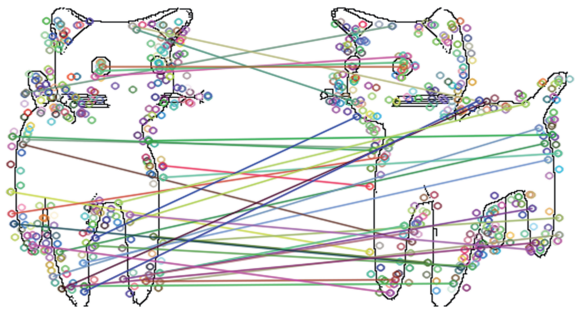


図 A·1 パターンマッチング（SURF）の結果の例
Fig. A·1 Result of pattern matching (SURF).

A.1.3 パターンマッチング手順

今回の調査では，以下の手順によって，候補画像 S1～S5 の中から問題画像 Ti に一番近い画像を選ぶ．

1. 問題画像 Ti に写っているオブジェクトのサイズに合わせて，候補画像 S1～S5 の縮尺を正規化する．具体的には，オブジェクトの x 方向の長さと y 方向の長さのうち，大きいほうの長さをオブジェクトのサイズととらえ，Ti と Sj のオブジェクトのサイズが等しくなるように，S1～S5 を拡大縮小する，
2. 領域ベースマッチングによって Ti に近い Sj（j = 1～5）を選別する．領域ベースマッチングには種々の方法が考えられるが，今回は最もシンプルな方法の 1 つである「面積比を求める」方法を利用した．具体的には，Ti の面積と Sj（j = 1～5）の面積を比較して，その差が θ_1 以下であれば「一致」と判定する．以下，手順 2 で一致と判定された候補画像を Sj*（j* = 1～5）と記す．
3. Sj*（j* = 1～5）に対し，特徴ベースマッチングによって Ti に最も近い Sj を選出する．今回は SURF [15] を採用した．具体的には，Ti と Sj*（j* = 1～5）のそれぞれの画像ペアに対して SURF を適用し，距離関数の値が閾値 θ_2 以下である特徴点対を抽出する．特徴点対の総数が最大となった画像 Sj を Ti に最も近い候補画像として選出する．

なお，閾値 θ_1，θ_2 は予備実験を通じて経験的に決定した．今回の各閾値の値は次のとおりである．$\theta_1 = 12500$，$\theta_2 = 0.17$.

A.1.4 実験結果

A.1.3 の手順により，各問題画像 Ti（i = 1～5）に対し，候補画像 Sj（j = 1～5）の中から一番近い画像を選択した結果，全問正解となる画像が選出された．この結果から，「YUNiTi にスケール変換，回転角度の下限，線画化を適用した CAPTCHA」は類似画像を選択するパターンマッチング攻撃によって突破されるケースがあることが確認された．

ただし，A.1.3 の手順 3 で実行した SURF によって抽出された特徴点対を確認したところ，問題画像のオブジェクトと候補画像のオブジェクトの各部位間の対応については，SURF はこれを確実には発見できていないことが認められた．実際の例として，問題画像 T1 と候補画像 S1 の SURF マッチングの結果を可視化した画像を図 A·1 に示

© 2016 Information Processing Society of Japan

す．図 A·1 左が T1，右が S1 であり，対応がとれた特徴点対が線で結ばれている．対応がとれた特徴点対の総数自体は多いため，A.1.3 の手順 3 のルールによって S1 が（T1 に一番近い画像として）正しく選出されたものの，部位間のマッチングとしては誤った結果が得られている箇所が少なくないことが分かる．これは，Locimetric 型（問題画像と候補画像の対応部位を解答する形式）の CAPTCHA である提案方式が，同じ部位を選択するパターンマッチング攻撃に耐性を有することを示す結果にもなっていることに留意されたい．



藤田 真浩 （学生会員）

2013 年 3 月静岡大学情報学部情報科学科卒業．2015 年 3 月同大学院修士課程修了．現在，同創造科学技術大学院博士後期課程．情報セキュリティ，ヒューマンインタフェースに関する研究に従事．



池谷 勇樹

2013 年 3 月静岡大学情報学部情報科学科卒業．2015 年 3 月同大学院修士課程修了．同年富士通株式会社入社．在学中，情報セキュリティに関する研究に従事．



可児 潤也

2012 年 3 月静岡大学情報学部情報科学科卒業．2014 年 3 月同大学院修士課程修了．同年株式会社富士通研究所入社．在学中，情報セキュリティに関する研究に従事．



西垣 正勝 （正会員）

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．同講師，助教授の後，2006 年より同創造科学技術大学院助教授．2007 年同准教授，2010 年同教授．博士（工学）．情報セキュリティ全般，特にヒューマニクスセキュリティ，メディアセキュリティ，ネットワークセキュリティ等に関する研究に従事．2013～2014 年情報処理学会コンピュータセキュリティ研究会主査．2015 年より電子情報通信学会バイオメトリクス研究専門委員会委員長．

© 2016 Information Processing Society of Japan