一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
BioX2014-2(2014-6)

高度なメンタルローテーションを利用した画像 CAPTCHA の提案

藤田 真浩† 池谷 勇樹† 可児 潤也† 米山 裕太† 西垣 正勝†

†静岡大学大学院情報学研究科 〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: †nisigaki@inf.shizuoka.ac.jp

あらまし 人間の高度な認知処理を利用した CAPTCHA の一つとして，メンタルローテーションを利用した画像 CAPTCHA が知られている．メンタルローテーションは，一つの視点から写された 2 次元物体や 3 次元物体を頭の中で回転させ，異なる視点から写された形姿を認識する能力である．しかし，既存のメンタルローテーション CAPTCHA は，パターンマッチングや機械学習を用いたマルウェアの攻撃によって突破されている．そこで，「異なる形状を持つ二つのオブジェクト間における部位の対応関係を問う」という問題形式を用いることで，より攻撃耐性が高いメンタルローテーション CAPTCHA を提案する．本稿では，提案方式のプロトタイプシステムの実装を行うとともに，提案方式に関する有用性を検討した．

キーワード CAPTCHA，メンタルローテーション，3DCG

A Proposal of Imaged-based CAPTCHA using sophisticated Mental Rotation

Masahiro FUJITA† Yuki IKEYA† Junya KANI† Yuta YONEYAMA†

and Masakatsu NISHIGAKI†

†Graduate school of Informatics, Shizuoka University

3-5-1, Johoku, Naka-ku, Hamamatsu, Shizuoka, 432-8011 Japan

E-mail: †nisigaki@inf.shizuoka.ac.jp

Abstract As one of the advanced Completely Automated Public Turing tests to tell Computers and Humans Apart (CAPTCHAs), the CAPTCHAs using mental rotation has been proposed. Mental rotation is an advanced human-cognitive-processing ability to rotate mental representations of one single 2D/3D object. However, as have already been reported, the mental rotation CAPTCHA can be overcome by pattern matching and/or machine learning. Therefore, we propose to enhance the mental rotation CAPTCHA, which we call “sophisticated mental rotation CAPTCHA”. The sophisticated mental rotation CAPTCHA system asks users to click the location on an object, which is corresponding to the position where the marker in another distinct object is located. In this paper, we have developed a prototype system of the CAPTCHA and have examined the effectiveness of the CAPTCHA.

Keyword CAPTCHA，Mental Rotation，3DCG

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 サービス提供サイト等に対する DoS（Denial of Service，サービス不能）攻撃が定常的に行われている．このような攻撃を防ぐためにはマルウェアによる 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．この要求を実現する技術の一つである CAPTCHA は，人間には容易に正解できるがコンピュータには正解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と判定する技術である[1].

現在，多くの Web サービス提供サイトでは文字判読型の CAPTCHA（図 1）がマルウェアの攻撃を防ぐ典型的な手法として広く採用されている．しかし，文字判読 CAPTCHA は OCR（自動文字読取）を備えたマルウェアによって破ることが可能であると指摘されている[2][3]．この問題に対し研究者たちは，マルウェアが依然として模倣が困難な「人間の高度な認知能力」を利用することでマルウェアの正解困難な CAPTCHA を実現しようと試みてきた[4]．この流れの中で，人間が有する「メンタルローテーション」の能力を利用した YUNiTi's CAPTCHA [5]（図 2）が提案された．メンタルローテーションは，一つの視点から写された 2 次元物体や 3 次元物体を頭の中で回転させ，異なる視点か

This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.
Copyright ©2014 by IEICE

ら写された形姿を認識する能力であり，人間が有する空間認識能力の一つである．

3 次元の空間認識はコンピュータが苦手とする分野の一つであり，YUNiTi's CAPTCHA はマルウェアが正解困難である理想的な CAPTCHA の一つとして注目を集めた[6]．しかしその後，YUNiTi's CAPTCHA にもテンプレートマッチングを用いた攻撃に対する脆弱性が存在することが報告された[7]．これは，YUNiTi's CAPTCHA が，「複数の候補画像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」という単純なメンタルローテーションタスクを用いていることに起因する．

そこで本稿では，メンタルローテーションタスクを「異なる形状を持つ二つのオブジェクト間における部位の対応関係を問う」という課題へと昇華し，この課題を用いることで，より攻撃耐性が高いメンタルローテーション CAPTCHA（以下，本稿では「高度なメンタルローテーション（sophisticated mental rotation）を利用した画像 CAPTCHA」と呼ぶ）を提案する．人間であれば，二つの 3 次元オブジェクトが同一のモデルであるか否かを判定するだけでなく，二つのオブジェクト間の部位の対応関係を認識することが可能である．さらに，異なる形状を持つオブジェクト間でも，両者の間に意味的な対応関係がとれればメンタルローテーションを行うことが可能である．前述のとおり，メンタルローテーションタスクは一般に，提示された二つのオブジェクトが同一のモデルであるか否かを判定する課題であるが，本稿では「異なる形状を持つ二つのオブジェクト間」でメンタルローテーションを行い，「部位の対応関係を認識する」という二つの点で，メンタルローテーションタスクの概念を拡張していることに注意されたい．

本稿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するとともに，メンタルローテーションを利用した既存の 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 章で提案方式に関する考察を示す．最後に 5 章でまとめと今後の課題を述べる．

2. メンタルローテーション

人間は空間認識能力が優れている．このため，3 次元オブジェクトが写っている 2 次元画像から，そのオブジェクトの 3 次元形状を容易に推測することができる[8]．また，人間は，一つの視点から写された 2 次元物体や 3 次元物体を頭の中で回転させ，異なる視点から写された形姿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る[9][10]．したがって，人間であれば 3 次元オブジェクトを撮影した 2 枚の画像を見たとき，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 次元モデルが同一のモデルであるか否かを判定することができる．

図 1 文字判読 CAPTCHA の認証画面例

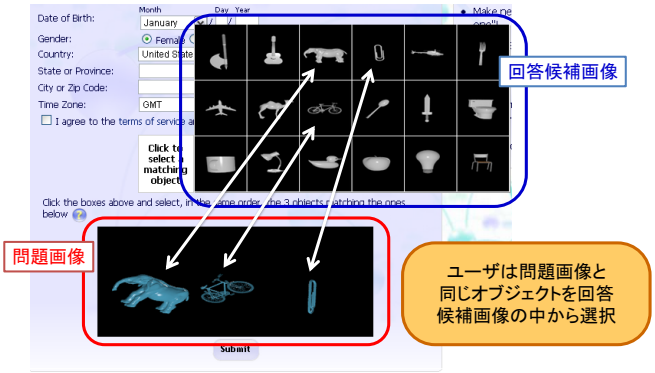

図 2 YUNiTi's CAPTCHA の認証画面例

メンタルローテーションを利用した CAPTCHA としては YUNiTi's CAPTCHA[5]が提案されている．YUNiTi's CAPTCHA は「複数の候補画像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」というメンタルローテーションタスクを利用している．具体的には，3 次元オブジェクトを 3 個並べた出題画像を提示し，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選択できたユーザを人間と判定する．出題画像は，候補画像のオブジェクトを別の角度から映した画像である．

しかし，YUNiTi's CAPTCHA はテンプレートマッチングを用いた攻撃に対する脆弱性が存在することが報告されている[7]．YUNiTi's CAPTCHA で用いられる 3 次元オブジェクト（候補画像）は毎回の問題で同一であり，18 体しかない．したがって，それぞれのオブジェクトに対して数百回分程度の出題画像を参照画像群として保存してしまえば，新たな出題画像に対し，自動プログラムがその出題画像と各オブジェクトの参照画像との類似度を計算することによってどのオブジェクトが写された画像であるかを識別可能である．

3. 高度なメンタルローテーションを利用した画像 CAPTCHA

3.1. コンセプト

本稿では，メンタルローテーションタスクを「異なる形状を持つ二つのオブジェクト間における部位の対応関係を問う」という課題へと昇華し，この課題を用

いることで，より攻撃耐性が高いメンタルローテーション CAPTCHA を提案する．高度なメンタルローテーションを利用した画像 CAPTCHA の認証画面例を図 3（type-α），図 4（type-β）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）の二枚の画像から構成される．ユーザは，出題画像に写された 3 次元オブジェクトの任意の 1 部位に付加されたマーカ（赤色の球）が回答画像ではどこに当たるのかを回答する．

一般にメンタルローテーションタスクでは，提示された二つのオブジェクトが同一のモデルであるか否か判定することをユーザに求める．しかし，2 章に示したとおり，メンタルローテーションタスクを使用した YUNiTi's CAPTCHA はパターンマッチングによって突破可能である．そこで，提案方式ではメンタルローテーションタスクに二つのタスクを追加することで，マルウェアが突破困難な出題形式を持つ CAPTCHA へと昇華している．

一点目が，「部位の対応関係を認識する」というタスクである．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．本方式は「複数の候補画像の中から一番近い画像を選ぶ」という出題形式でないため，単なるテンプレートマッチングでは突破不可能であり，テンプレートマッチングに対する攻撃耐性が向上していることが期待される．

二点目が，「異なる形状を持つ二つのオブジェクト間」でのメンタルローテーションである．マルウェアは 3 次元画像認識の技術を利用可能であるため，一つの 3 次元オブジェクトを異なる二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が復元される可能性がある[11]．そこで，出題画像と回答画像で，形状が異なるオブジェクトを使用することで立体復元技術への耐性を持たせている．人間であれば，出題画像と回答画像とで一部の形状が異なっていたとしても，同一のオブジェクトと認識し，メンタルローテーションを行うことが可能である（type-α）．更には，完全に形状が異なっていても，それらが“意味的に同じ”であれば[1]，その意味を理解した上で，メンタルローテーションを行うことが可能である（type-β）．

3.2. 手順

提案方式の認証画面作成手順は図 5，6 に示すとおりである．なお，システムには大量の 3 次元オブジェクトの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．

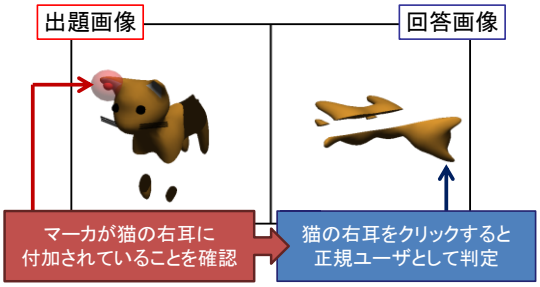

図 3　高度なメンタルローテーションを利用した画像 CAPTCHA の認証画面例（type-α）

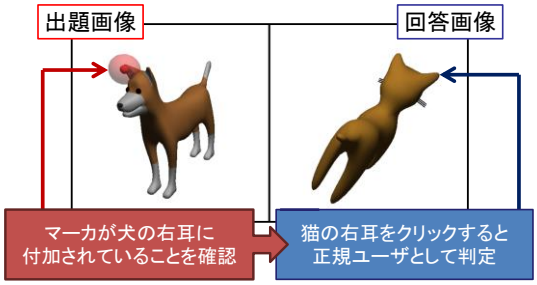

図 4　高度なメンタルローテーションを利用した画像 CAPTCHA の認証画面例（type-β）

① システムは，出題画像に利用する 3 次元オブジェクト（以下，出題用オブジェクト）のモデルをランダムに選ぶ．
② type-α の場合，システムは出題用オブジェクトに任意の加工を施す．
③ システムは，出題画像の視点を任意に選ぶ．
④ システムは，回答画像に利用する 3 次元オブジェクト（以下，回答用オブジェクト）を，①で選んだ出題用オブジェクトに任意の加工を施すことで生成する（type-α）．あるいは，①で選んだオブジェクトと意味的に同じオブジェクトをランダムに選ぶ（type-β）．
⑤ システムは，回答画像の視点を任意に選ぶ．
⑥ 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．
⑦ システムは，⑤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．なお，type-β の場合，“意味的に同じ部位”を求めることに注意されたい．
⑧ システムは，出題画像（マーカが描画されている）と回答画像（マーカは付与されているが描画されていない）を表示する．
⑨ ユーザは，回答画像において「出題画像内のマーカが付与された部位（⑥で選ばれた部位であり，

[1] たとえば，図 4 では，出題画像と問題画像のオブジェクトが「四足歩行動物」という意味で共通している．

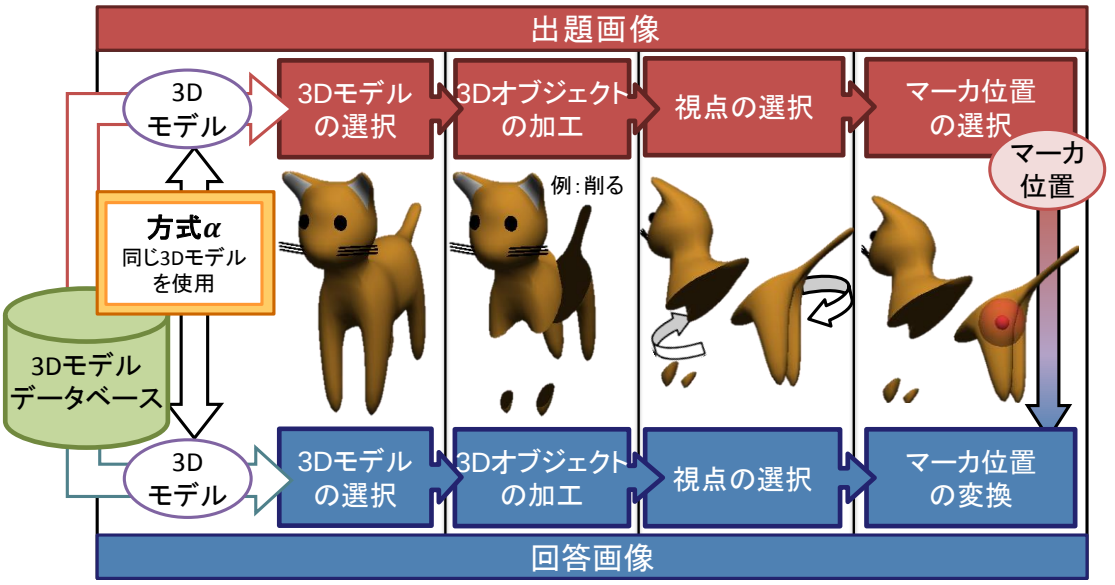

図 5　type-α の自動生成手順

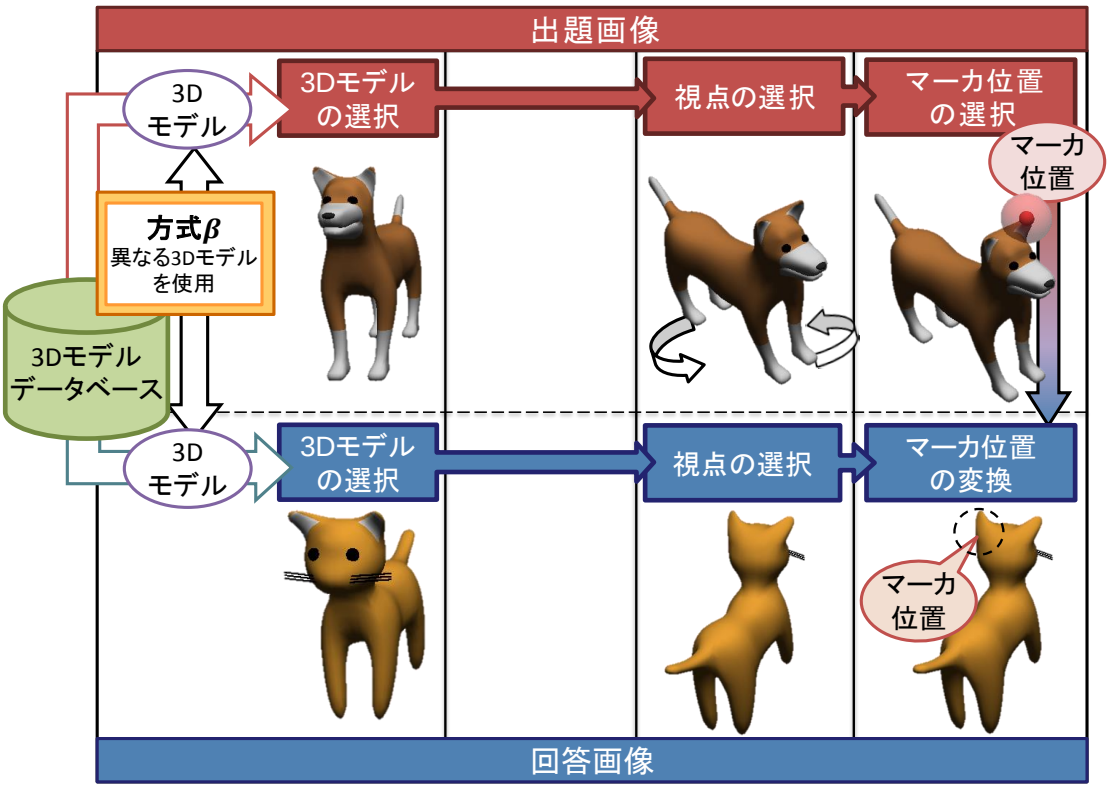

図 6　type-β の自動生成手順

⑦で求めた部位)」を回答する.

⑩　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.

type-α の手順②④において，オブジェクトの加工方法は種々の方法が考えられる．図 3 や図 5 では，オブジェクトを x,y,z 方向に任意の倍率で拡大･縮小するとともに，オブジェクトの一部を削っている．また，type-β の手順⑦「意味的に同じ部位を求める」ためには，登録されている 3 次元モデルそれぞれに対して，モデル間における部位同士の対応関係を記したデータベースがシステム内に管理されている必要がある．

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，マーカの位置を同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）自身がユーザの回答に対する正解判定を行う」ことを可能としている．加えて，システムに大量の 3 次元モデルを登録しておき，使用するモデル，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. プロトタイプシステムの実装

提案方式（type-α）のプロトタイプシステムを実装することで，3.2 節に示した手順で CAPTCHA システムが実現できることを確認した．図 7 にプロトタイプシステムの認証画面例を示す．ユーザは，出題画像中のマーカ部位（赤い球）が回答画像上のどの位置にあるかを同定し，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（ディスプレイ上の座標）と正解部位の位置（ディスプレイ上の座標）の距離が閾値以下であれば認証成功とした．図 7 の例では，出題画像においてマーカが猫の右耳を示していることがわかるため，回答画像における猫の右耳をクリックすれば正解となる．

実装において留意した点を以下に述べる．

3.3.1. 加工方法

オブジェクトを加工（3.2 節の手順②④）する目的の一つが，立体復元技術に対する耐性の向上である．その加工方法には種々の方法が考えられる．今回の実装では，問題画像，回答画像のオブジェクト各々に対して，x,y,z 方向にそれぞれ任意の倍率で拡大・縮小をするとともに，オブジェクトの一部を削るという加工を行った．削る割合については，オブジェクトに外接する直方体体積の約 3 割を削ることとした．

3.3.2. マーカ

回答オブジェクトにおける「出題用オブジェクトのマーカとして選択された頂点」に対応する頂点の座標が正解座標となる．ここで，頂点座標は 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部位がディスプレイ上ではどの座標にあたるかを計算した上で，その 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 30 ピクセルを円の半径とした．出題画像内に描画するマーカは，ユーザにマーカの中心を意識してもらえるように，大きな半透明の球の中に小さな不透明の球が入っている形状とした．

また，3.3.1 節に示したとおり，今回の実装ではオブジェクトに対して「一部を削る」という加工を施している．したがって，マーカ位置は，オブジェクトにおいて削られた部分が選択されないよう考慮した．

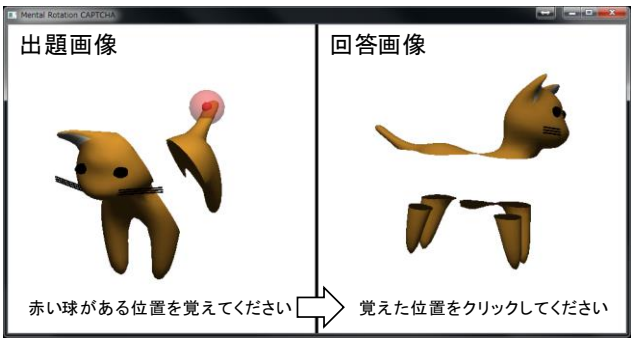

図 7　プロトタイプシステムの認証画面例

3.3.3. 視点

回答画像の視点の選択（3.2 節の手順⑤）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正解箇所を解読できる危険性が生じる．したがって，今回の実装では，回答画像の視点が問題画像の視点から Y 軸（垂直軸）を中心として 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した．

4. 考察

2 章で述べたように，YUNiTi's CAPTCHA は，人間が有する「3 次元物体の認識能力」を利用している点，出題画像の自動生成を達成している点で秀逸な画像 CAPTCHA である．しかし，「複数の候補画像の中から一番近い画像を選ぶ」という形の質問形態となっているため，テンプレートマッチングに対する脆弱性を残していた．

YUNiTi's CAPTCHA においては，正解画像の候補となる 3 次元オブジェクトの数が 18 種類しかないため，この問題が非常に顕著となる．正解画像の候補となる 3 次元オブジェクトの数が少ない場合，過去の出題画像をアーカイブすることによって，オブジェクトごとに「数百の異なる視点から描画されたオブジェクトの画像」を収集することは，非現実的とはいえない．YUNiTi's CAPTCHA の出題画像はいずれかのオブジェクトを任意の視点から描画した画像であるため，前もってあらゆる視点からの画像をアーカイブしておけば，アーカイブ画像群の中に出題画像に近い画像が必ず存在するため，テンプレートマッチングによって正解オブジェクトを特定でき得る[7].

提案方式では，3.2 節に示したとおり，生成され選る画像の組み合わせは「モデル数×視点数×部位数」となり，YUNiTi's CAPTCHA の「モデル数×視点数」と比較して，過去の画像をすべてアーカイブし終える

までの手間が倍増している．加えて，提案方式は「複数の候補画像の中から一番近い画像を選ぶ」という質問形態をとっておらず，毎回の認証画面には出題画像と回答画像の 2 枚が表示されるだけである．毎回の認証で画像が少ない分，不正者は過去の画像をすべてアーカイブするためにより多くの手間を要することになる．これらの点で，提案方式はテンプレートマッチングに対する耐性が向上していると期待される．

また，3.1 節で説明したように，立体認識技術を利用した攻撃については，出題画像と回答画像とで異なる形状を持つ 3 次元オブジェクトを利用することで対処を行っている．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．特に機械学習を利用した解読などについては，今後，早急に分析を行う必要がある．

5. まとめと今後の課題

本稿では，「異なる形状を持つ二つのオブジェクト間における部位の対応関係を問う」という問題形式を用いることで，より攻撃耐性が高いメンタルローテーション CAPTCHA を提案した．提案方式では，従来のメンタルローテーションタスク（二つのオブジェクトが同一か否か判定する）に，「異なる形状を持つ二つのオブジェクト間」でメンタルローテーションを行い，「部位の対応関係を認識する」という二つのタスクを加えている．これにより，パターンマッチングにより高い耐性を有するメンタルローテーション CAPTCHA が実現された．

今後は，今回実装したプロトタイプシステムを利用して，type-α のユーザビリティに関する評価を行っていく．また，type-β についても開発・評価をしていきたい．

謝辞

静岡大学大学院情報学研究科漁田武雄教授には，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．本稿で使用した 3 次元モデルは「メタセコ素材！（http://sakura.hippy.jp/meta/）」で公開されている素材を利用させていただきました．御礼申し上げます．

文　献

[1] The Official CAPTCHA Site, 入手先〈http://www.captcha.net〉（参照 2014-04-30）.

[2] J.Yan, A.S.E.Ahmad, Breaking Visual CAPTCHAs with Naïve Pattern Recognition Algorithms, 2007 Computer Security Applications Conference, pp. 279-291, 2007.

[3] J.Elson, J.Douceur, J.Howela, J.Saul, Asirra: a CAPTCHA that exploit interest-aligned manual image categorization, 2007 ACM CSS, pp.366-374, 2007.

[4] K Chellapilla, K Larson, P Simard, M Czerwinski, “Computers beat humans at single character recognition in reading-based Human Interaction, Proofs(HIPs), 2nd Conference on Email and Anti-Spam (CEAS), 2005.

[5] YUNiTi.com, 入手先〈http://www.yuniti.com/〉（参照 2014-04-30）.

[6] 3D-based Captchas become reality, CNET, 入手先〈 http://www.cnet.com/news/3d-based-captchas-become-reality/〉（参照 2014-04-30）.

[7] How they'll break the 3D CAPTCHA, 入手先〈http://technobabblepro.blogspot.jp/2009/04/how-theyll-break-3d-captcha.html〉（参照 2014-04-30）.

[8] Tom Stafford, Matt Webb, Mind Hacks, Oreilly & Associates Inc., 2004.

[9] R. Shepard, L. Cooper, Mental images and their transformations, MIT Press, Cambridge, MA, 1982.

[10] R. Shepard, J. Metzler, Mental rotation of three dimensional objects, Science, New Series, Vol.171, No.3972, pp.701-703, 1971.

[11] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, U.K., 2000.