Vol.2017-DPS-170 No.4
Vol.2017-CSEC-76 No.4
2017/3/2

物体のサイズ感を利用した3DCG画像CAPTCHAの評価

西原 大貴[1,a)]　新井 イスマイル[2]

概要：近年，Web サービスに対し問題となっている，アカウントの大量取得など自動プログラムを用いた機械攻撃を防ぐ技術の一つとして，CAPTCHA と呼ばれる人間か機械かを識別するテストが利用されている．人間特有の「常識からの逸脱を認識する能力」として，2 つの 3 次元 (3D) オブジェクトのめり込みを検出できる能力に着目した既存研究の 3DCG 画像 CAPTCHA は，CAPTCHA に要求される 3 要件「利便性」「安全性」「自動生成性」を満たすとされたが，輪郭抽出技術の応用などによって機械が解読できる可能性がある．これに対し，本研究では，特定の 3D オブジェクトを拡大・縮小し，そのサイズ感が周囲と異なる場合に違和感を覚える人間の能力に着目した CAPTCHA を提案している．本稿では，以前の検証結果 [1] を踏まえ，使用するオブジェクトを選定し再検証を行った．その結果，物体数 4 体の時に，被験者の正解率は，以前の手法 0.65 や既存手法の 0.67 に比べ，提案手法は 0.76 に向上した．

キーワード：ネットワークセキュリティ，CAPTCHA，サイズ感，3 次元コンピュータグラフィックス

1. はじめに

近年, 自動プログラムを用いてアカウントを大量に取得するなど，Web サービスに対する機械攻撃が問題となっている．これを防ぐ技術の一つとして, CAPTCHA(Completely Automated Public Turing test to tell Computers and Humans Apart) と呼ばれる人間か機械かを識別するチューリングテストが利用されている．それらのうち文字判別型 CAPTCHA(**図 1**[*1]) は，歪みやノイズをかけた文字列画像の文字列を読み取らせる CAPTCHA であり，現在では広く利用されている．しかし，近年では OCR(Optical Character Recognition) 技術の発展などにより，機械攻撃によって破られる可能性が高まってきた．すなわち，CAPTCHA は，人間にとって解読しやすいこと (利便性) の他に，機械攻撃耐性 (安全性) が確保されている必要がある．一方で，CAPTCHA には，出題が自動生成可能である (自動生成性) という要求も存在する [2]，[3]．これを満たさない場合，出題の総数は有限となり，データベースを参照する機械攻撃が予測される．従って，CAPTCHA にはこれらの 3 要件が要求される．

これを満たす既存研究として藤田らは，常識的な形状をした異なる 2 つの 3 次元オブジェクトをマージしてめり込ませることで生成した非現実オブジェクトをユーザに選択させる 3DCG 画像 CAPTCHA 手法 (以下，非現実画像 CAPTCHA) を提案した [2]．しかしながら，輪郭抽出技術を応用した機械攻撃により破られる可能性が考えられる．これに対し，我々は「サイズ感」に着目し，拡大・縮小された非常識な大きさの物体をユーザに選択させることで，輪郭抽出技術などにより各オブジェクトの形状や名称が限定されたとしても容易には解読されないと期待できる手法を提案している．

図 1　文字判別型 CAPTCHA

本稿では，提案手法の利便性について，以前の検証結果 [1] を踏まえ，使用に適する物体としてサイズ感が固定できる物体を用いて再検証した結果，物体数 4 体の時，以前の手法，(再検証した) 提案手法，既存手法それぞれの正解率は，0.65，0.76，0.67 であった．すなわち，提案手法の正解率は，以前の手法から改善され，また，既存手法に対しても優位性が示された．

以下，2 章で関連研究として，文字判別型 CAPTCHA や画像を用いた CAPTCHA を紹介する．3 章にて，本研究の提案手法の詳細を説明する．4 章では，提案手法を実装して実験を行い，5 章でその結果と考察を述べる．6 章では本研究のまとめと今後の課題を述べる．

1　明石工業高等専門学校 電気情報工学科
2　奈良先端科学技術大学院大学 総合情報基盤センター
a)　e1227@s.akashi.ac.jp

*1　https://auth.sso.biglobe.ne.jp/mail/

Vol.2017-DPS-170 No.4
Vol.2017-CSEC-76 No.4
2017/3/2

表 1 既存手法の 3 要件に対する評価

	安全性	利便性	自動生成性
文字判別型	△	○	○
Assira	×	○	△
4 コマ漫画	×	△	×
2 枚画像	○	△	○
非現実画像	△	○	○

2. 関連研究

文字判別型 CAPTCHA は，図 1 のように歪みやノイズがかけられた文字列画像の文字列をユーザが読み取り，テキストとして入力するものである．自動生成が可能であり，かつ機械攻撃耐性に優れていたため，現在に至るまで多くの Web サービスで利用されてきた．しかし，近年では OCR 技術の発展により，機械攻撃によって破られつつある．これの対策として，安全性を高めるためには，出題画像の歪みやノイズを強くすればよいが，同時に人間にとっても解読しにくくなり，利便性の低下を伴う．すなわち，一般的に安全性と利便性の関係はトレードオフであり，安全性を確保しつつも，より利便性の高い CAPTCHA 手法が必要とされる．この問題を解決するため，以下に挙げるような様々な手法が提案されてきた．文字判別型 CAPTCHA を含めた各々の手法について，著者が 3 要件への評価を行い，高い順に ○△× で表したものを**表 1** に示す．

2.1 Assira

Assira[4] は，12 枚の犬と猫の画像から，猫をすべて選択できたユーザが人間であるとする CAPTCHA である．猫の絵を認知する能力は人間の高度な認知能力であり，機械による突破は難しいと考えられていた．しかし，2 クラスの分類を得意とする機械学習判別機を用いた攻撃が有効であるとされた [5] ため，安全性に問題がある．

2.2 4 コマ漫画 CAPTCHA

4 コマ漫画 CAPTCHA[6] は，人間特有の最も高度な認知処理である「ユーモアを解する能力」に着目し，ランダムに並べ替えられた 4 コマ漫画の各コマを，正しい順序に並べ替えさせる手法を用いた．機械はユーモアの理解が困難で，正攻法による突破が簡単ではない．しかし，並べ替え総数が少なく総当たり攻撃 (ブルートフォースアタック) に脆弱であるため安全性に問題があり，また起承転結が明解な 4 コマ漫画の自動生成が難しいという問題が残る．

2.3 2 枚の画像を重ね合わせた CAPTCHA

小林らが提案した，2 枚の画像を重ね合わせた画像の認識能力を問う CAPTCHA[3] は，**図 2** のように，重ねられた元の 2 枚の画像が何であるかを 10 種類の大分類に分けられた合計 100 個の選択肢から選択する方式である．従って，答えは ${}_{100}C_2 = 4950$ 通り存在し，藤田らが 4096 通り確保できれば十分であるとした [2] ことを鑑みれば，機械攻撃耐性は高い．また重ねられた 2 枚の画像を自動で分離することは困難であるため，安全性が保たれていると言える．しかし，この「2 枚画像」方式は，検証の結果，人間の回答時間が平均 27.2 秒であり，一般的な文字判別型 CAPTCHA が 10 秒から 18 秒 [3] であることに比べて長くなるという課題を指摘しており，利便性に欠ける．

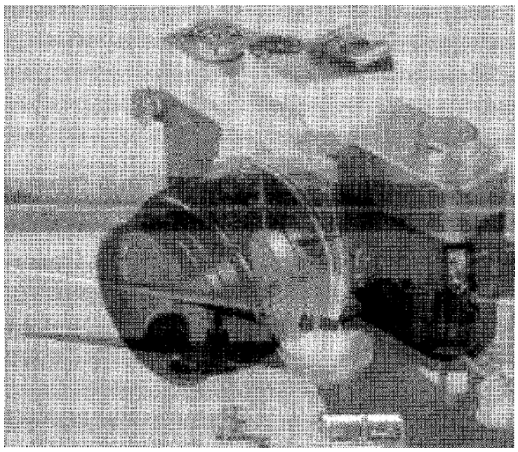


図 2 2 枚の画像を重ね合わせた CAPTCHA

図 3 非現実画像 CAPTCHA

2.4 非現実画像 CAPTCHA

藤田らは，「常識からの逸脱を認識する能力」が人間特有の高度な認知能力であることに着目し，2 体の 3D オブジェクト同士をめり込ませて生成した新しいオブジェクト (非現実オブジェクト) をユーザに選択させる非現実画像 CAPTCHA を提案した．具体的には，**図 3** に示すような画像をユーザに出題し，複数の 3D オブジェクトの中に配置された 1 体の非現実オブジェクトをクリックさせる．これは，3DCG を用いることで無数の出題を自動生成でき，また，常識を持つ人間は容易に正解できるが，機械は人間の常識を備えることが困難なため，通常と非現実のオブ

図 4 提案手法による CAPTCHA 画像のイメージ

ジェクトを見分け難く，CAPTCHA の 3 要件を満たすとした．特に，安全性の検証としてオブジェクト同士の境界線が，マージされてできためり込み部分であるのか，あるいはめり込んではいないが遮蔽関係にあるのかを機械学習により検出する攻撃手法や，その他総当たり攻撃にも耐性を持ちうるとされた．具体的には，機械学習を用いた手法では，あらかじめ入手した大量の出題画像から，「一部を切り出した画像」と「その部分に正解オブジェクト (めり込んでいる部分) が存在するか否か」という教師用データセットを用いて機械学習を行うことで，画像中に「めり込みが含まれるか否か」を判定する分類器を作り，めり込んだオブジェクトを検出する攻撃手法を実装した．その後，この手法では画像中の「めり込んだ部分」と「遮蔽関係」を検出できるかどうかを検証した結果，正解率は 69.6%であることから，「遮蔽関係」と「めり込み」の区別は機械にとって困難であると結論付けた．また，総当たり攻撃耐性の検証では，CAPTCHA の有するべき総当たり数が 4096 通りであるとし，機械が画像解析によって出題画像中のすべてのオブジェクトを抽出できた場合を考えれば，物体数 N に対して，総当たり数は N となるため，4 体のオブジェクトが描画された出題画像を 6 枚出題し，全て正解できたユーザを人間とみなせば $4^6 = 4096$ 通り確保できるとした．

しかしながら，遮蔽関係あるいはめり込みである境界部分について，両者の区別を試みる方法として，輪郭抽出技術などによって境界部分の特徴を取り出し，機械学習を用いて遮蔽関係かめりこみであるかを検出するなど，その他の攻撃手法により，めり込んだオブジェクトが検出できる可能性が考えられる．

3. 物体のサイズ感を利用した手法の提案

藤田らが，常識からの逸脱を認識する人間特有の能力として，非現実オブジェクトを用いたことに対し，本研究では，物体の常識的なサイズ感をユーザに識別させる CAPTCHA 手法を提案する．

3.1 概要

提案手法では，**図 4** に示すように，「背景」3D オブジェクトを基準として，複数の「物体」3D オブジェクトを配置した画像を出題し，その中から全オブジェクトに対して非常識な大きさの「正解」オブジェクト (この例では，テーブル上の横転した白いコップ) を選択できたユーザを人間とみなす．

背景および配置する物体として用いる 3D モデルは，実世界でのサイズ情報と共に，予めデータベースに大量に登録されているとする．その中から背景を 1 つ選択したのちに，その背景に対して大きすぎず，かつ小さすぎない物体を無作為に選択する．これは，例えば閉ざされた室内空間に，車など本来は屋外にあるような大きなオブジェクトが配置されると通常より大きく見え，また，大きな空間に小さいオブジェクトが配置されると見えにくくなると著者が感じたためである．

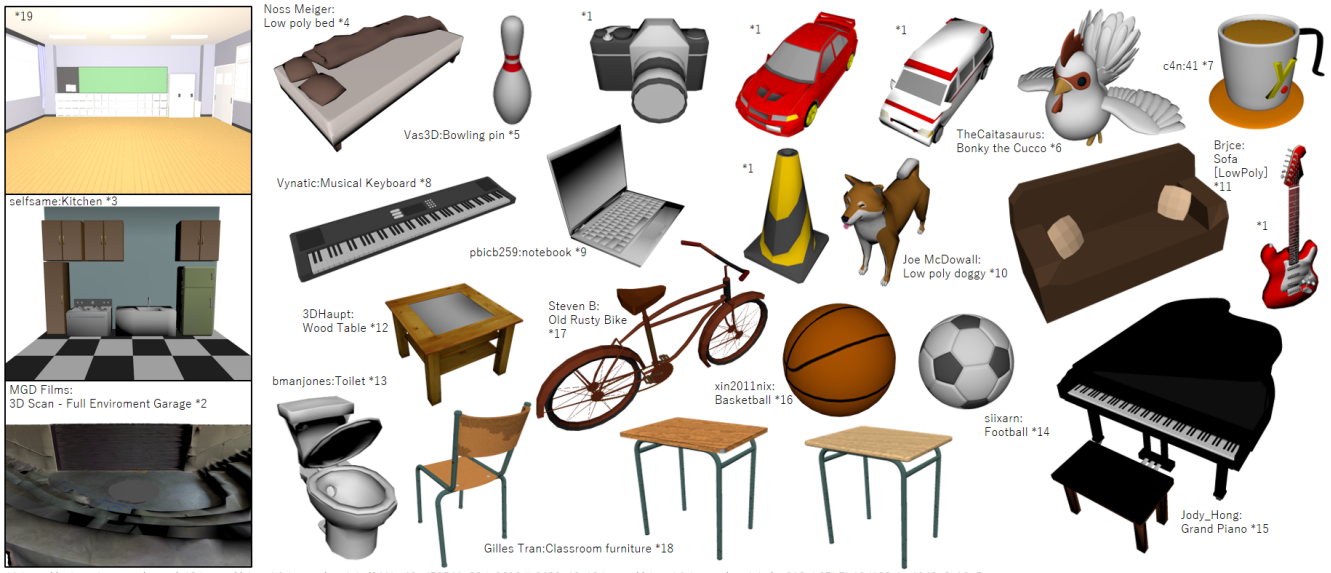
なお，データベースに登録する物体として，人間が常識的な大きさを把握できないものは除外する．以前の検証 [1] により，鉢植えなどのように，物体の特性として大きさが一意に定まらないものや，ライフル銃などのように大きさが固定されていたとしても一般的に馴染みがないものは，サイズ感を捉えにくいという結果を得たからである．

また，出題生成の際は，物体は背景に対して宙に浮かせず，互いに重ならないよう任意の位置に配置し，出題の 3DCG 画像を描画するためのカメラ位置は，配置した物体が全て映る範囲内で，無作為に定める．

3.2 期待される提案手法の有効性

本提案手法では，ユーザは解読のためにサイズ感を捉える，すなわち出題された 2 次元画像に映る 3 次元空間を想像し，配置された物体の大きさを背景やその他の物体から相対的に認識する必要がある．そのため，解読する際は，背景が示す場所や状況 [*2]，物体の 3 次元空間内の位置関係を理解し考慮する必要がある．出題画像には，背景を含め複数のオブジェクトが映っているため，機械が個々のオブジェクトを認識することは容易ではないが，近い将来には，機械が輪郭抽出や機械学習などにより，配置された個々の物体の正体をおおむね解明することで，その物体の常識的な大きさを検索エンジンやデータベースから参照できる可能性も考えられる．しかし，提案手法では，背景の場所や，背景と物体の位置関係を解読できない限り，機械による突破は容易ではないと思われる．例えば，学校教室内と体育館内では置かれる物体が同じであっても出題の 2 次元画像として描画される大きさは異なる．また，同じ背景内

[*2] 将来的には，背景の状況に応じた解読をユーザに求める方法も検討している．例えば，道路上に配置された自動車と，机上に置かれたミニチュアカーでは，その大きさは全く違うが，いずれも人間にとっては違和感がなく常識的な大きさである一方で，機械は常識を備えることが難しく，安全性の向上が期待できる．



*1 http://sakura.hippy.jp/meta/, *2 https://sketchfab.com/models/341be40ad79541c88de9612db6423c43, *3 https://sketchfab.com/models/ea318ab27b7b46d188eba4040a2c12e5,
*4 https://sketchfab.com/models/d3825852d4264688af04c6e6a48dab84, *5 https://sketchfab.com/models/513488fe22524a9d93c550cf982ed2e9, *6 https://sketchfab.com/models/49a44042c7af4675bf47eb2e77fd47f0,
*7 https://sketchfab.com/models/eb89708a80be4468be445a67b8097133, *8 https://sketchfab.com/models/ae3dee84c9e54f8e88e90df0b45eda5c, *9 https://sketchfab.com/models/fa19ccc1fece40fc84870f6671d2c929,
*10 https://sketchfab.com/models/1c8c763518ab4751bfcddf0b6a34011a, *11 https://sketchfab.com/models/9dc6696eef71420aa4114d3afe95e644, *12 https://sketchfab.com/models/ceK99OSKpP8DUbBuJJ2S0kw4mYq,
*13 https://sketchfab.com/models/e44bee96002a4f56b8b4a194c8745d72, *14 https://sketchfab.com/models/e82e717d86fb42b0bca1ff6c2ef74c28, *15 https://sketchfab.com/models/ae725dcd33174a718e6293573e0a70a1,
*16 https://sketchfab.com/models/536d687777f0473dbc79f00ea4a17917, *17 https://sketchfab.com/models/fb2e16d216ca4fda948a073ad5ee26a5, *18 http://www.oyonale.com/modeles.php?lang=en&page=29,
*19 http://www003.upp.so-net.ne.jp/kakomiki/page/3ddata.html

図 5 実験で使用した背景 (左) および物体 (右) の 3D オブジェクト

であっても，カメラの位置や向きにより，遠くまで映るのかあるいは手元近くのみが映るかによって，3 次元空間での奥行きの度合いが異なるため，物体の位置関係の解読は容易ではない．背景との関連を 3 次元的に把握できない場合であっても，大きさを把握できた複数の物体同士で 2 次元画像に投影された大きさを比較することにより，解答を推測できる可能性が考えられる．しかし，出題画像には奥行きがあるため，手前と奥に配置された物体では 3 次元空間で同じ大きさであっても投影された 2 次元画像での描画画素数は異なるため，この手法では破ることができないと思われる．これらより，高い安全性が期待できる．

また，近い将来，3D モデルが Web 上に大量に出回ると予測されるため，出題に用いる 3D モデルを大量に取得できると思われるが，有限であることに変わりはない．しかし，使用するモデルの組み合わせや物体の配置場所，カメラ位置は無作為に決定され，ほぼ無数の出題が自動生成できるため，自動生成性を有すると言える．

一方で，人間は，出題画像から 3 次元空間上の様子を想像でき，背景や配置された物体の奥行きや位置関係を推測し，常識的なサイズ感を瞬時に把握することができる．そのため，出題に対する解読の負担が小さくなり，利便性の確保が期待できる．

以上のことから，提案手法は CAPTCHA に要求される 3 要件を満たすと期待する．

4. 提案手法の実装と実験内容

4.1 実装の条件

図 5 に示すような，3 種類の背景 (学校教室，台所，ガレージ) およびサイズ感が固定できる 22 種類の物体 (ベッド，ボーリングピン，カメラ，スポーツセダン，救急車，鶏，コーヒーカップ，キーボード (楽器)，ノート PC，ロードコーン，柴犬，ソファ，エレキギター，リビングテーブル，自転車，バスケットボール，サッカーボール，グランドピアノ，トイレ，学校用椅子，学校用机 2 種) の 3D オブジェクトを Web 上から収集した．これらを用い，3.1 に従ってシステムを実装し，出題画像を 640×480 画素で生成した．

表 2 各実験の内容

実験	出題枚数	正解の倍率	物体数
(1) 以前の手法	26	0.5～0.75, 1.5～2	4
(2) 今回の手法	24	0.5, 1.5	4, 8, 12, 16
(3) 既存の手法	24	-	4, 8, 12, 16

4.2 実験内容

3.2 で，提案手法は自動生成性を定性的に満たすと判断したが，利便性および安全性については，検証が必要であるため，下記に挙げる実験を行う．

明石高専の機械工学科または電気情報工学科に属する 15 人の被験者に下記の 3 つの出題画像群 (実験 (1) 以前の手法，実験 (2) 今回の手法，実験 (3) 既存の手法) のそれぞれに対して画像上の正解だと思う座標をクリックしてもらい，各出題画像の回答時間と正解率を記録した．この時の各実験の内容を**表 2** にも示す．

(1) 以前の検証 [1] と同じ 26 枚の画像 (正解オブジェクトの倍率は，0.5～0.75 または 1.5～2 の範囲で乱数値であり，配置する物体数は実装の簡略化のため 4 体のみで

あった).

(2) 4.1 で示した今回の手法において，各背景 (3 種類)，正解オブジェクトの倍率は，0.5 または 1.5 倍 (2 種類)，配置する物体数 4，8，12，16 体 (4 種類) の全通りを組み合わせた 24 枚．正解オブジェクトの倍率は，上記 2 種類の固定値とした．これは，正解オブジェクトの倍率を固定しない場合，同じ倍率でその他の条件 (背景の種類や物体数) が異なる出題画像の実験結果について比較しにくいと考えたためである．配置する物体数は，以前の検証と同じ 4 体に加え，比較のため既存手法の検証と同じ条件にした．

(3) 既存手法である非現実画像 CAPTCHA[2] を再現し，生成した 24 枚．配置する物体数 4，8，12，16 の 4 種類について各 6 枚ずつ生成した．この物体数は，既存研究での検証条件と同じである．この時 3D モデルは，前項 (2) と同じものを使用した．

システムのインタフェースに慣れるため，各検証の前に被験者自身が十分と思うまで練習することを許した．しかし，実際の運用上では，出題画像で配置されるオブジェクト自体はユーザにとって初見であることが推定される．この状況に近づけるため，練習では，検証本番時に使用しない 3D モデルのみを使用した．

また，提案手法の総当たり攻撃への耐性の指標を調べるため，(2) の今回の手法で用いた出題画像について，正解オブジェクトの描画画素数を調べ，出題画像全体の画素数 ($640 \times 480 = 307200$ 画素) に対する割合 (正解の描画割合) を算出した．

4.3 仮説

4.2 で挙げた 3 つの実験 ((1) 以前の手法，(2) 今回の手法，(3) 既存の手法) に対して，それぞれ仮説を述べる．以前の手法 (1) では，以前の検証 [1] と同様に，使用する物体のサイズ感が一意に定まらず，利便性 (被験者の正解率や回答時間) が低いと思われる．一方で，以前の検証 [1] を元に，被験者がサイズ感を捉えやすい物体を使用した今回の手法 (2) では，利便性が改善され，特に正解率は，以前の検証時から分かったこととして適切な物体を選別した場合に得られるとした正解率 0.9 程度へ，向上すると期待される．加えて，提案手法は，既存手法と同様に人間特有の常識からの逸脱を認識する能力に着目した手法であるため，以前の手法 (2) の結果は，既存手法の再現である (3) と同等の利便性が得られると期待する．

また，藤田らの既存研究では，物体数の増加とともに，回答時間が長くなった一方で，正解率に大きな変化は見られなかった．したがって，今回の提案手法 (2) の結果も同様なものであると予測する．

表 3 各物体数に対する平均正解率と平均回答時間

	平均正解率			平均回答時間 [s]		
	以前	今回	既存	以前	今回	既存
4 体	0.65	0.76	0.67	4.60	4.67	3.54
8 体	-	0.27	0.62	-	7.11	3.99
12 体	-	0.43	0.68	-	10.43	3.54
16 体	-	0.26	0.76	-	9.47	4.42

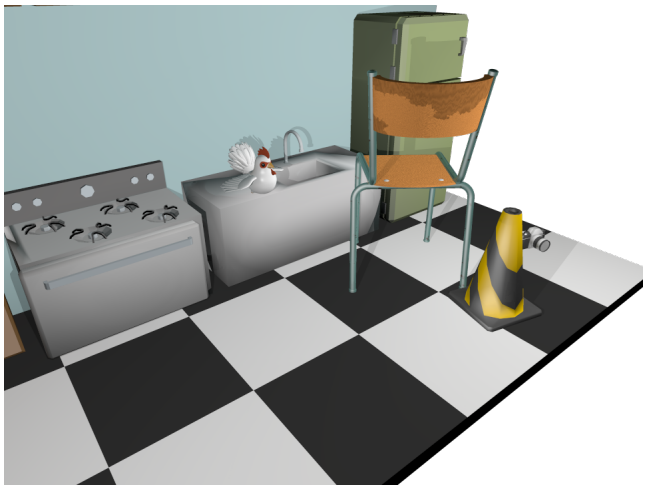


図 6 正解率が高かった成功出題例 (正解率 1，椅子が正解)

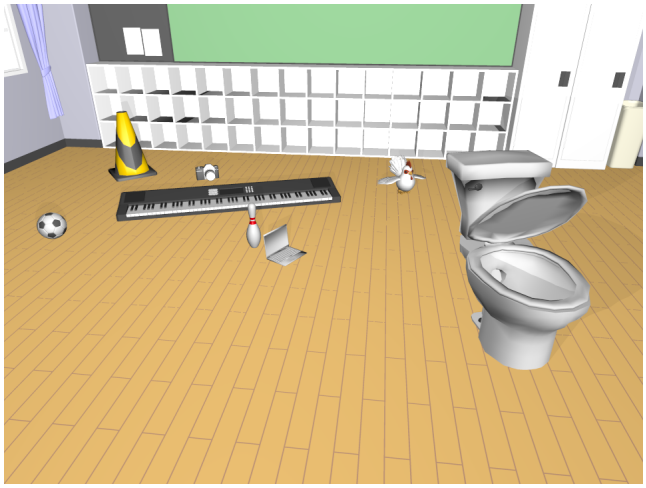


図 7 正解率が低かった失敗出題例 (正解率 0.07，キーボードが正解)

5. 結果と考察

5.1 利便性

実験の結果として，各物体数に対する，各手法毎の平均正解率および平均回答時間を**表 3** に示す．

表 3 によると，物体数 4 体の場合，正解率が以前手法の 0.65 から今回手法の 0.76 へと改善が見られ，回答時間はほぼ同等であった．**図 6** は，物体数 4 体での出題のうち，正解率が高かった成功出題例 (正解オブジェクトは椅子) である．また正解率は，既存手法の 0.65 を上回ったことが分かる．しかし，以前の検証 [1] で目標とした正解率 0.9 程度には達していない．被験者が正解できなかった画像を個々

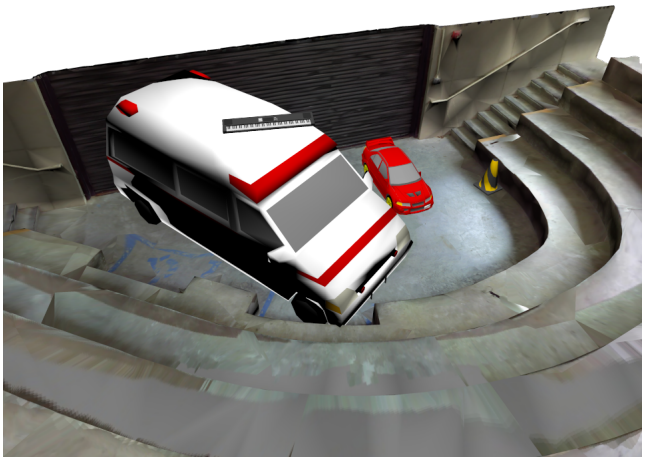


図 8 正解率が低かった失敗出題例 (正解率 0.47，スポーツセダンが正解)

に分析し，不正解であった原因とその対策について下記に述べる．

1 つ目の原因は，サイズ感を捉えにくいオブジェクトを使用していたことである．ロードコーンやソファは筆者の先入観により，サイズ感が固定されると考えたために選択したが，被験者にとっては曖昧だったと思われる例が含まれていた．対策として，将来的には，物体の実世界での大きさ情報を Web 上で検索し，ショッピングサイトや百科事典などから自動収集する予定であるが，その際に複数の文献でサイズ感が一致するような物体に限定すれば良い．ところが，キーボードはサイズ感が明確であり，一般的な知名度も低くないと思われるにも関わらず，そのサイズ感を捉えて正解できた例が少なかった．**図 7** はその一例である (正解オブジェクトはキーボード)．バスケットボールやサッカーボールにも同様の傾向が見られた．音楽やこれらの球技を趣味としない人には分かりにくい，家庭に置かれていることが少ないなど，様々な原因を推測できるが，個々の物体の特性が要因であるため，根本的な対策が難しい．解決のためには，システムを運用する中で，学習によって正解率が低くなるオブジェクトを排除するなどの対策を講ずる方法が考えられる．

2 つ目の原因は，被験者が特定の 2 物体のみで大きさを比較してしまったことである．正解オブジェクトの近くに配置された物体のみと大きさを比較し，誤った方を選択してしまった例がいくつか見られた．**図 8** はその一例である (正解オブジェクトはスポーツセダン)．大きさが大きく異なる物体同士が近くに配置されると比較しにくいことが主な原因だと思われる．今後の課題として，物体の大きさと配置による結果の違いを検証したい．

3 つ目の原因は，正解オブジェクトが小さすぎたことである．出題画像は，3 次元空間上に配置した物体を 2 次元画像へ投影して生成される．その物体自体が背景に対して相対的に小さい，あるいはカメラ位置が物体から遠いため

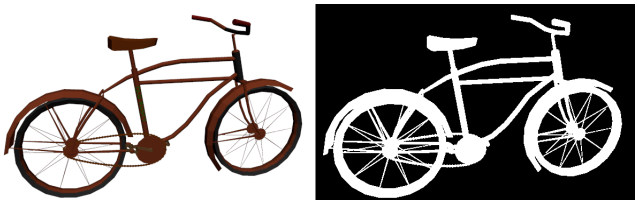


図 9 自転車の 3D オブジェクト (左) とその正解マスク画像 (右)

表 4 誤クリックを正解とした時の平均正解率 (今回の手法)

	修正前	修正後
4 体	0.76	0.76
8 体	0.27	0.41
12 体	0.43	0.46
16 体	0.26	0.41

に出題画像での描画画素数が少なくなってしまった物体は，3 次元空間上で大きさを変更しても，投影後の 2 次元画像での画素数の変化量は小さい．このため，被験者はその差異を認知できなかった可能性が考えられる．対策として，物体の倍率を変更する前後について 2 次元画像での変化画素数の最小閾値を決めれば良い．

4 つ目の原因は，被験者の誤クリックである．被験者は正解を認識していると思われるが，上記 3 つ目の原因により正解オブジェクトが小さいためにクリック位置がずれてしまったケースや，自転車など物体の形状特性として物体描画部分 (正解画素数) が小さくクリックしにくかったケースが見られた．自転車の 3D オブジェクトの様子と，それに対し，自転車をクリックしたとみなす範囲を示した正解マスク画像 (白い部分をクリックできれば自転車をクリックしたと判断される) を**図 9** に示す．この図のように，自転車のフレーム部分は細く，特に車輪の中の中空部分をクリックしてしまった例が含まれていた．このようなケースを，著者が目視で確認し正解と見なした場合，今回の手法の正解率は**表 4** の通りとなる．解決のためには，物体が描画されている部分の付近も正解の範囲とすることや，自転車などのように 2 次元画像に投影された場合にクリックしにくいようなモデルの場合は，細い箇所を太くする，車輪などのように中空部分も正解画素とみなす，などの対策が挙げられる．

その他，実験の結果から考えられる内容を下記に述べていく．

表 3 や表 4 から分かることとして，既存研究は正解率や回答時間に大きな変化がほとんど見られず，物体数との相関が低い．一方で提案手法では，物体数 4 体の時に，その他の場合に比べ正解率が高く回答時間が短かった．しかし，オブエジェクト数が増える (8 体以上になる) と，4 体の場合と比べて正解率が $\frac{1}{2}$ 程度に減少し，また回答時間が 1.5 ～ 2 倍以上に増加しており，提案手法は物体数 4 付近が最適であると思われる．したがって，3 体や 5 体とした

表 5 既存手法の再現実験と既存研究論文中の結果の比較

	平均正解率		平均回答時間 [s]	
	再現実験	論文中の値	再現実験	論文中の値
4 体	0.67	0.98	3.54	2.2
8 体	0.62	0.92	3.99	3.2
12 体	0.68	0.90	3.54	4.2
16 体	0.76	0.94	4.42	5.5

表 6 物体数と正解の倍率に対する正解率および回答時間

	平均正解率		平均回答時間 [s]	
	0.5 倍	1.5 倍	0.5 倍	1.5 倍
4 体	0.64	0.87	4.82	4.52
8 体	0.40	0.13	5.89	8.34
12 体	0.56	0.31	8.36	12.5
16 体	0.42	0.09	9.00	9.95

場合，それらがさらに改善される可能性があるため，今後これらについても検証していく．

なお，表 3 に示した既存の手法による実験 (再現実験) の結果は，**表 5** に示すとおり，既存研究の論文中 [2] の値と大きく異なった．この原因として，被験者の慣れと，使用したオブジェクトの違いという 2 つが挙げられる．既存の論文中では，被験者が十分だと思うまで実験本番と同じオブジェクトを用いて練習させたことで，被験者が本番に望む前に，出題画像に含まれるオブジェクトの形状を十分に確認し慣れていた一方で，4.2 で述べた通り我々が行った再現実験では，練習に用いたオブジェクトは本番で一切使用していない．これにより，被験者の正解しやすさに影響したと考えられる．他方の原因である，使用したオブジェクトの違いとして，再現実験では既存手法に向かないオブジェクトが使用されていたことが考えられる．既存の論文中でも指摘されているが，マージしてめり込ませる 2 つのオブジェクトの組み合わせにより，その形状によってはめり込み部分がほとんどできず，被験者が正解しにくい場合がある．実験で使用したオブジェクトは，提案手法に最適であるものを選択したため，既存手法には不向きであり，被験者の正解率や回答時間を悪化させた可能性が考えられる．

表 6 は，提案手法である今回の手法について，出題画像の物体数と正解オブジェクトの倍率に対する正解率および回答時間の平均を示している．これによると，正解オブジェクトの倍率が 0.5 倍の時は，1.5 倍の時と比べ，ほぼ全ての物体数で正解率，回答時間ともに優れている．0.5 倍への縮小時と同等の結果を得られる 1.5 以上の拡大倍率について，模索する必要があるが，正解オブジェクトを大きくしすぎると，出題画像に投影される正解画素数が増え，後述する総当たり攻撃に対して脆弱になる．これを避けるため，正解オブジェクトの倍率として，拡大を行う出題に比べ縮小する出題を確率的に多く生成するなどの対策が考えられる．

表 7 は，今回の手法について，背景に対する平均正解率と平均回答時間を示している．これによれば，いずれの背景に対しても，正解率はほぼ同等であり，多少の違いはあるが回答時間も大きく異ならない．したがって，提案手法の結果は使用する背景に対する依存はほとんどないと思われる．

表 7 背景に対する平均正解率と平均回答時間

背景	平均正解率	平均回答時間 [s]
教室	0.44	7.97
台所	0.41	8.36
ガレージ	0.43	7.43

5.2 総当たり攻撃耐性

藤田らの既存研究 (非現実画像 CAPTCHA)[2] では，機械が画像解析によって出題画像中の物体を抽出できた場合，描画された物体数 N に対して 1/N の確率で突破できたと仮定すれば，出題画像 1 枚あたり N 通りの総当たり数しか有さないと論じている．しかし，提案手法は，既存研究とは異なり背景があるため，配置された物体を正確に抽出することは容易ではないと考えられる．また背景には，物体を配置せずとも，背景を特徴付ける物体が複数存在している．例えば，今回の実験で用いた背景である教室には，教卓，黒板，窓などが，台所には，コンロや冷蔵庫が置かれている．したがって，出題画像中のオブジェクト数を正確に把握することが難しい．これらを踏まえ，本稿では提案手法の総当たり攻撃耐性の指標として，出題画像中に含まれる正解の描画割合を基準として議論する．

提案手法の利便性の実験 (2) で用いた出題画像について，正解の描画割合は，平均すると，1.796%となった．これは，機械が出題画像中の 1 画素をランダムにクリックする総当たり攻撃を行えば，目安としておおよそ 1.796%の確率で突破できることを示している．すなわち，実験で用いた提案手法の出題画像は $\frac{1}{1.796/100} \fallingdotseq 55.68$ 通りの総当たり数を有すると言える．藤田らは論文中で CAPTCHA の有するべき総当たり数が 4096 通りであるとした [2] ことを踏まえれば，出題画像 1 枚のみでは総当たり数が低い．したがって，既存研究と同様に，複数枚出題し全て正解できたユーザを人間と見なす方法が対策として考えられる．この場合，出題数 M に対し，1 枚あたりの総当たり数の N 乗の総当たり数が確保される．例えば，提案手法で 2 枚の出題とすれば，$55.68^2 \fallingdotseq 3100$ 通りとなり，3 枚とすれば $55.68^3 \fallingdotseq 172622$ 通りが確保できる．2 枚の出題では，総当たり数がやや不足するが，出題画像生成時に，正解の描画割合を 1.56%程度に制約すれば，総当たり数は $(\frac{1}{1.56/100})^2 \fallingdotseq 4109$ 通りが確保される．ただし，5.1 で述べたように，物体を小さくすると，倍率を変化させた時に出題の 2 次元画像での変化量が少なく，人間が正解オブジェクトを発見しにくくなる恐れを指摘している．そのため，利便性を確保しつつ正解の

表 8 物体数 4 体において複数枚出題した時の正解率と回答時間

出題数 M	正解率	回答時間 [s]
1	0.76	4.67
2	0.58	9.34
3	0.44	14.01

描画割合を削減できる閾値を今後検証して行く必要がある．

しかし，出題数 M を増加させれば，安全性 (総当たり攻撃耐性) が向上する一方で，利便性 (人間の正解率および回答時間) が低下すると思われ，目安として正解率は M 乗，回答時間は M 倍となる．例えば，物体数 4 体の場合について，出題数 M に対する利便性は**表 8** の通りとなる．この表によれば，3 枚出題しても，回答時間は 14.01 秒であった．文字列判別型 CAPTCHA が，小林らは 10 秒から 18 秒程度 [3]，藤田らは 12 秒程度 [2] としていることを踏まえれば，3 枚の出題時でも十分であると言える．しかし，正解率はいずれの枚数でも目標値 0.9 を満たせず，1 枚あたりの正解率の向上が今後の課題となる．

6. おわりに

本稿では，物体のサイズ感を利用した 3DCG 画像 CAPTCHA 手法について，以前の検証結果から得たオブジェクトの選定基準を踏まえて改良し，利便性 (正解率，回答時間) の再検証を行った．結果として，物体数 4 体の場合において，正解率は以前の検証 0.65 および既存手法 0.67 に対し提案手法はそれを上回った (0.76) が，目標値 0.9 には達せず，正解率の向上にはさらなるオブジェクト選定のや正解オブジェクトの描画画素数の調整が必要であるなどの課題を見出した．また，物体数の増加に伴い，利便性は，既存研究では大きな変化が見られなかった一方で，提案手法では減少するため，物体数 4 体付近が最適であることがわかった．総当たり攻撃耐性の検証では，十分な総当たり数を確保するためには複数枚の出題が必要であるが，出題数を減らすために正解の描画割合を調整する方法は，利便性を損なわない程度に制限する必要があり，今後の検証で確かめていくとした．今後は，課題の解決とともに，機械攻撃手法を実装し，さらなる安全性の検証をしていく予定である．

参考文献

[1] 西原大貴，新井イスマイル：物体のサイズ感を利用した 3DCG 画像 CAPTCHA の検討，研究報告コンピュータセキュリティ (CSEC)， Vol. 2016-CSEC-75, No. 5, pp. 1–5 (2016).
[2] 藤田真浩，池谷勇樹，可児潤也，西垣正勝：非現実画像 CAPTCHA：常識からの逸脱を利用した 3DCG 画像 CAPTCHA，情報処理学会論文誌，Vol. 56, No. 12, pp. 2324–2336 (2015).
[3] 小林　司，藤堂洋介，森井昌克：画像認識の困難性を利用した CAPTCHA 方式の提案，電子情報通信学会技術研究報告, LOIS, ライフインテリジェンスとオフィス情報システム， Vol. 110, No. 207, pp. 37–42 (2010).
[4] Elson, J., Douceur, J. R., Howell, J. and Saul, J.: Asirra: A CAPTCHA that Exploits Interest-Aligned Manual Image Categorization, *Proc. of ACM CCS2007*, pp. 366–374 (2007).
[5] Golle, P.: Machine Learning Attacks Against the Asirra CAPTCHA, *Proc. of ACM CCS2008*, pp. 535–542 (2008).
[6] 可児潤也，鈴木徳一郎，上原章敬，山本　匠，西垣正勝：4 コマ漫画 CAPTCHA, 情報処理学会論文誌, Vol. 54, No. 9, pp. 2232–2243 (2013).