非現実画像CAPTCHA：オブジェクトのめり込みを利用した違和感画像CAPTCHA

非現実画像 CAPTCHA：オブジェクトのめり込みを利用した違和感画像 CAPTCHA

藤田 真浩[†1] 池谷 勇樹[†1] 可児 潤也[†1] 西垣 正勝[†1]

1. はじめに

現在，多くの Web サービス提供サイトでは文字画像判読型の CAPTCHA（図 1 左）[1]や動物画像の判別を用いた Asirra（図 1 右）[2]など，画像を利用した CAPTCHA がマルウェアの DoS 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，これらの CAPTCHA は OCR や機械学習の機能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ため[3][4]，人間のより高度な認知能力を利用した画像 CAPTCHA が求められており[5]，いくつかの研究が進められている．しかし，既存研究には，出題画像の自動生成，出題画像の解読耐性などに関して課題が残っており，著者らの知る限り，全ての課題を解決する理想的な CAPTCHA は未だ提案されていない．

そこで本稿では，(i)人間のより高度な認知能力を利用する，(ii)出題の自動生成が容易である，という二つの要件を実現する，違和感を利用した画像 CAPTCHA「非現実画像 CAPTCHA」の提案を行う．

2. 非現実画像 CAPTCHA

2.1 コンセプト

「違和感」は，人間が，自分のもつ常識から逸脱した場面に遭遇した時に生じるものであると考えられる．また，現在のところ，人間のように常識を備えたコンピュータは実現できていない．したがって，「常識に基づいた事象」と「違和感を覚える（常識から逸脱した）事象」を識別できるか否かをユーザに問うことで，そのユーザが人間であるか否かを判定可能である．

提案方式では 3 次元モデルを「常識を基づいた事象」として利用する．3 次元モデルを利用した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．3 次元モデルは，動物や車のように，現実に存在する有形物をモデル化したものであることが多い．人間はモデル化前のオブジェクトを少なくとも一度は現実世界で見た経験を持っており，それらのモデルを「常識」として保持していると考えられる．

また，常識から逸脱した（違和感を覚える）事象については，現存の 3 次元オブジェクトを適切に加工することで生成する．加工の方法には種々のアプローチが考えられる

図 1. 既存の画像 CAPTCHA（左：文字画像判読型 CAPTCHA，右：Asirra）

が，本稿では，ランダムに選んだ二つの 3 次元オブジェクトどうしをめり込ませることで新しいオブジェクト(以下，「非現実なオブジェクト」と呼ぶ）を生成する．

非現実画像 CAPTHA ではこれら 2 種類のオブジェクトを利用して CAPTCHA 画像を作成する．具体的には，複数の通常の 3 次元オブジェクトの中に，一体の非現実なオブジェクトを配置した一枚の画像を CAPTCHA として出題する．人間であれば，常識から逸脱した不自然なめり込みをしている非現実なオブジェクトに対して「違和感を覚える」ため，通常のオブジェクトの中に紛れる非現実なオブジェクトを発見することは容易である．

2.2 出題画像の自動生成

提案方式の出題画像作成手順を以下に示す．システムには通常の 3 次元オブジェクトのモデルが大量に登録されていることを前提とする．1 枚の出題画像中に含まれるオブジェクトの個数 N はセキュリティパラメータである．

① 3 次元モデルのオブジェクト N-1 体をランダムに選ぶ．
② 各モデルに対して，アフィン変換を施すことでスケールの変更と回転を行う．
③ ②の N-1 体のオブジェクトをそれぞれが重ならないように三次元空間平面 α 上に配置する．
④ 3 次元モデルのオブジェクト 1 体をランダムに選ぶ．
⑤ ④のオブジェクトに対しても，②と同様アフィン変換を施すことによってスケールの変更と回転を行う．
⑥ ③で設置した N-1 体のオブジェクトの中から，ランダムに 1 体のオブジェクト選択する．
⑦ ⑤のオブジェクトを三次元空間平面 α 上の⑥のオブジェクトと同座標に（⑥のオブジェクトと⑤のオブジェクトが互いにめり込むように）に配置する。
⑧ 三次元空間平面 α 上のオブジェクト群を二次元画像へ投影することによって，出題画像を生成する．

†1 静岡大学

本手順で生成した非現実画像 CAPTCHA の出題画像の例（N=8）を図 2 に示す．図 2 では，画面右下に犬と車がめりこんだ非現実なオブジェクトが配置されている．

3. 基礎実験

ユーザ（人間）が 非現実画像 CAPTCHA に正答することが可能であることを基礎実験によって確認した．被験者は情報セキュリティ系の研究室に所属する学生 7 名である．CAPTCHA 画像中のセキュリティパラメータ N は 25（体）に設定した．各被験者は，3 回行う実験本番の前に，各被験者が十分と思えるまでの回数の練習を行った．

基礎実験の結果，非現実画像 CAPTCHA の正答率は全ユーザの平均で 90.5%（全ユーザ 7 名×3 回＝21 回の試行を行ったうち，正答が 19 回，失敗が 2 回）であった．一般的な文字判読型 CAPTCHA の平均正答率は約 93%であるため，ほぼ同程度の正答率が得られた．

回答に要する平均所要時間は一問あたり約 5.7 秒であった．一般的な文字判読型 CAPTCHA の平均所要時間は約 12 であるため，非現実画像 CAPTCHA は文字判読 CAPTCHA より短い時間で解ける CAPTCHA であることが示唆された．

4. 攻撃耐性に関する考察

4.1 総当たり攻撃

マルウェアが画像解析によって出題画像中からすべてのオブジェクトを抽出できれば，マルウェアは，抽出したオブジェクトの総数 N-1（オブジェクトの総数は N 体であるが，その内の 2 体はお互いにめり込んでいるため，機械には 1 体に見える）に対して，1/(N-1)の確率で正答することができる．総当たり数を増やすためには，出題画像中のオブジェクト数 N をある程度大きな値にする，問題数を増やすといったパラメータ変更による方法が考えられる．

4.2 非現実オブジェクトの検出

非現実画像 CAPTCHA に対する攻撃の一つに，画像中から，「オブジェクト同士がめり込んでいる」部分を検出しようと試みる攻撃が考えられる．機械が「めり込み」を検出するためには，二つのオブジェクトをマージした時に現れる特徴を利用するものと推察される．以下に，非現実オブジェクト検出に用いられる可能性がある特徴例を列挙する．

- 二体のオブジェクトの境目：二つのオブジェクトをマージした際の境目が，エッジ抽出によって検出可能である可能性がある
- 色：非現実オブジェクトは，通常のオブジェクトよりも多くの色を利用している可能性がある．また，通常のオブジェクト内で使用されにくい色の組み合わせが使われている可能性もある．
- 空間周波数：非現実オブジェクトは，通常のオブジェクトよりも空間周波数が高くなる可能性がある

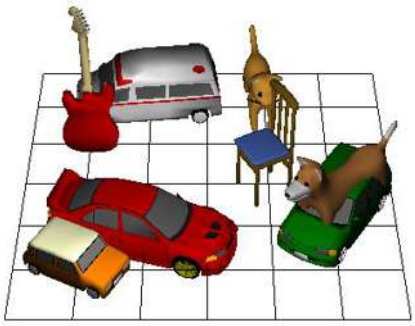

図 2. 非現実画像 CAPTCHA の認証画面例(N=8)

本攻撃に対する理論的な評価は今後の課題である．ただし，筆者らは，以下に示す二つの理由から，提案方式が本攻撃に対して一定の耐性を有することを期待している．

第一に，出題画像中には，2 体のオブジェクトの位置の前後関係によって，めり込んではいないが遮蔽関係にあるオブジェクトも存在し得る（たとえば，図 2 中の右上に配置されている「椅子」と「猫」）．空間認識能力を持たない機械にとっては，めり込んでいるか／遮蔽関係にあるかの判別は困難であると期待される．

第二に，3 次元オブジェクトの中には複数の物体から構成されるオブジェクトがある（たとえば，「花」と「鉢」から成るオブジェクト）．今後，画像解析技術が進み，機械が複数の物体から構成されるオブジェクトを一つ一つの物体に分解できたとしても，機械にとって，それが常識に基づいた構成であるか，常識から外れた構成であるかを区別することは困難であると期待される．

5. まとめ

本稿では，人間の違和感を覚える能力を利用した「非現実画像 CAPTCHA」の提案と考察を行った．人間のより高度な認知能力を利用しているため攻撃耐性が高く，かつ，出題画像の自動生成が容易である点が特長である．

今後は，3 次元オブジェクトの種類，出題数といった条件を変えながら評価実験を繰り返し，提案方式の可用性についてより深く調査していく．また，攻撃耐性についても理論的な評価を進めていきたい．

謝辞 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です．ここで，作者の方に厚く御礼申し上げます．

参考文献

1) Unlocking Google's Gmail CAPTCHA http://www.gmailhelp.com/2009/10/unlocking-googles-gmail-captcha/
2) J.Elson, et al, Asirra: a CAPTCHA that exploit interest-aligned manual image categorization, 2007 ACM CSS, pp.366-374, 2007.
3) J.Yan, A.S.E.Ahmad, Breaking Visual CAPTCHAs with Naïve Pattern Recognition Algorithms, 2007 Computer Security Applications Conference, pp.279-291, 2007.
4) P.Golle, Machine Learning Attacks, Against the ASIRRA, CAPTCHA, 2008 ACM CSS, pp.535-542, 2008.
5) K.Chellapilla, el al., Computers beat humans at single character recognition in reading-based Human Interaction, Proofs(HIPs), 2nd Conference on Email and Anti-Spam (CEAS), 2005.