# キメラCAPTCHA：3DCGを利用した違和感画像CAPTCHA

# キメラ CAPTCHA：3DCG を利用した違和感画像 CAPTCHA

藤田 真浩，池谷 勇樹，可児 潤也，西垣 正勝*

**概要**．本稿では，「人間のより高度な認知能力を利用する」，「問題の自動生成が容易である」という二つの要件を満たす CAPTCHA「キメラ CAPTCHA」の提案，実装，攻撃耐性の考察を行う．提案方式では，ランダムに選んだ二つの 3 次元オブジェクトをマージすることでキメラオブジェクトを生成する．そして，複数の通常の 3 次元オブジェクトの中に，一体のキメラオブジェクトを配置した一枚の画像を問題画像として出題する．キメラオブジェクトはユーザの常識から逸脱した形をしており，人間であれば常識から外れたものに対して「違和感を覚える」ため，そのオブジェクトを発見することは容易である．キメラ CAPTCHA のプロトタイプシステムを実装し，基礎実験とアンケート調査を行った結果，高い正答率と利便性が確認され，提案方式の実用性が示唆された．また，現状の画像認識技術を考慮したうえで，提案方式が高い攻撃耐性を持つことを定性的に考察した．

## 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 CAPTCHA や動物画像の判別を用いた Asirra [1]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 [2][3]，人間のより高度な認知能力を利用して画像 CAPTCHA の攻撃耐性を高めることが求められている[4]．一方で画像 CAPTCHA には，「攻撃耐性の向上」という要求と同時に「問題の自動生成」という要求も存在する．これら二つの要求の両立は困難であり，実際，現在までに提案された CAPTCHA の多くは，どちらかの要求を満たせていない．

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，という二つの要求を両立する，違和感を利用した画像 CAPTCHA「キメラ CAPTCHA」の提案，実装，考察を行う．

## 2 キメラ CAPTCHA

### 2.1 コンセプト

「違和感」は，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．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．加工方法には種々のアプローチが存在するが，本稿では，任意に選んだ二つの 3 次元オブジェクトをマージすることで新しいオブジェクト（以下，「キメラオブジェクト」と呼ぶ）を生成する．

キメラ CAPTCHA では，複数の通常の 3 次元オブジェクトの中に，一体のキメラ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キメラ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キメラオブジェクトを発見することは容易である．

### 2.2 キメラオブジェクトの自動生成

キメラオブジェクトは「三次元平面の任意の一座標に，2 体の 3 次元オブジェクトを配置する」ことで容易に生成可能である．同座標に配置された 2 体の三次元オブジェクトは，共通部分が隠されることで互いにマージされた状態となり，人間には 1 体のキメラオブジェクトとして認識される．本手法であれば，大量の 3 次元モデルの中から，任意に抽出した 2 体の 3 次元オブジェクトをマージすることが可能であり，事実上無数のキメラオブジェクトを自動生成することが可能である．


*静岡大学，nisigaki@inf.shizuoka.ac.jp

### 2.3 システム開発

提案方式は，既存の 3DCG プラットフォームを用いて問題画像の自動生成が可能である．下記手順を実装し，キメラ CAPTCHA システムを開発した．開発したシステムで生成された問題画像例（画像中のオブジェクト数 N=8）を図 1 に示す．なお，実運用では，システムには通常の 3 次元モデルが大量に登録されていることを前提とする．

① 3 次元モデル N-1 体をランダムに選ぶ．
② 各オブジェクトに対して，アフィン変換を施すことでスケール変更と回転を行う．
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 1 体をランダムに選ぶ．
⑤ ④のオブジェクトに対して，②と同様アフィン変換を施すことでスケール変更と回転を行う．
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⑥のオブジェクトと同座標に（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し，出題画像を生成する．

## 3 基礎実験

ユーザ（人間）が 提案方式に正答可能であることを基礎実験によって確認した．被験者はセキュリティ系研究室所属の学生 7 名である．CAPTCHA 画像中のオブジェクト数（セキュリティパラメータ）N は 25（体）に設定した．各被験者には 3 問の問題を解くことを求めた．

基礎実験の結果，正答率は全ユーザ平均で 90.5%（7 名×3 回＝21 回うち，正答 19 回，失敗 2 回）であった．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約 5.7 秒／問であった．文字判読型 CAPTCHA の平均所要時間は約 12 秒／問であるため，文字判読 CAPTCHA より短時間で解ける CAPTCHA であることが示された．

## 4 攻撃耐性に関する考察

提案方式への攻撃として，画像中から「オブジェクト同士がめり込んでいる」部分を検出す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（例：二つのオブジェクトをマージした際の境目）を利用するものと推察される．本攻撃に対する理論的な評価は今後の課題であるが，以下に示す二つの理由から，提案方式が本攻撃に一定の耐性を有することを期待している．

図 1. キメラ CAPTCHA 問題画像例（N=8）

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 1 中の左上に配置されている「椅子」と「猫」）．空間認識能力が低い機械にとっては，めり込み関係／遮蔽関係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，元来めり込んでいる形で，複数の物体によって構成されるオブジェクトがある（たとえば，図 1 左下：「草」と「鉢」から成るオブジェクト）．今後，画像解析技術が進み，機械が複数の物体によって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めり込みであるか，常識から外れためり込み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

人間の違和感を覚える能力を利用した「キメラ CAPTCHA」の提案，実装，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，問題画像の自動生成が容易な点が特長である．

今後は，実験条件を変えながら評価実験を繰り返し，提案方式の可用性をより深く調査する．攻撃耐性についての理論的な評価も進めていく予定である．

### 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．作者の方に厚く御礼申し上げます．

## 参考文献

[1] J.Elson, et al, Asirra: a CAPTCHA that exploit interest-aligned manual image categorization, 2007 ACM CSS, pp.366-374, 2007.
[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] P.Golle: Machine Learning Attacks, Against the ASIRRA CAPTCHA, 2008 ACM CSS, pp.535-542, 2008.
[4] 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.