人間の視覚補完能力を用いた文字列 CAPTCHA の提案 —側面形状からの立体文字の推定—

藤　竜成[a)]・星野 理彦[b)]・山場　久昭[c)]・油田　健太郎[d)]・岡崎　直宣[e)]

A Proposal of a Reading Text CAPTCHA Using Human Visual Complementation Ability

Ryusei FUJI, Masahiko HOSHINO, Hisaaki YAMABA, Kentaro ABURADA, Naonobu OKAZAKI

Abstract

Recently, a program called CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) has been introduced in many web sites because it prevents misconduct by automatic programs called bots. CAPTCHA distinguishes bots from legitimate human users by requiring some questions that are easy for human users to solve but difficult for bots. However, with the development of various technologies such as OCR, object recognition, machine learning, bots come to be able to solve CAPTCHAs. In order to outcome such troublesome bots, more sophisticated CAPTCHA is desired. In this paper, we propose a new CAPTCHA scheme that uses amodal completion. Amodal completion is a visual supplement function that complements the missing part in the brain and recognizes the whole image when a part of the object is hidden or can not be seen. In the proposed method, characters are stereoscopically rendered. We set the color of the front part and the back part of the three-dimensional characters to the same color (black) as the color used in the background of CAPTCHA. By Amodal completion, we can correctly recognize characters from only the shape of the side part of the three-dimensional characters. But it is hard for bots to recognize the characters. We carried out several experiments to confirm that the proposed CAPTCHA scheme has enough resistance against bots attacks using OCR and object recognition. Other experiments were also carried out to evaluate the usability of the proposed CAPTCHA scheme. To evaluate the usability, we used system usability scheme (SUS). The results of the experiments showed that the proposed CAPTCHA scheme is usable and has enough resistance against bots attacks.

Keywords: CAPTCHA, bot, amodal compretion, OCR, object recognition

1. はじめに

今日、CAPTCHA(Completely Automated Public Turing test to tell Computers and Humans Apart)[1)] と呼ばれるプログラムが、多くのウェブサービス提供サイトに導入されている。これは、ボットと呼ばれる自動プログラムがブログのコメント欄を利用して自動で書き込みを行ったり、無料メールサービスなどを利用して大量のメールアカウントを取得したりする不正行為を防ぐためである。CAPTCHA は、サービスを利用しているユーザーの操作がコンピュータのプログラムによるものではないことを確認するための認証方式であり、人間には容易に回答できるがコンピュータには判読が困難である問題をユーザーに出題し、正解できたユーザーを人間と判断する技術である。

近年では、ボットにとって模倣が困難な人間の高度な認知能力を必要とした CAPTCHA が提案されている [2)]。というのも、OCR(Optical Character Recognition:光学文字認識)や物体認識、機械学習などの様々な技術の発達によって、ボットによる CAPTCHA への攻撃手段が増加しており、従来の CAPTCHA の安全性の低下が問題となっている [3)] からである。

本論文では、アモーダル補完と呼ばれる人間の視覚補完能力を利用することで、ユーザビリティを確保しつつ、ボットの突破率を低下できる手法を提案する。

以下、本概要の構成を述べる。第 2 節では既存の文字列 CAPTCHA について述べ、問題点を指摘する。第 3 節では提案方式について述べる。第 4 節では利便性に関する評価実験を行った結果を、第 5 節ではボット耐性の検証実験を行った結果を示す。第 6 節ではまとめと今後の課題について述べる。

2. 文字列 CAPTCHA

文字列 CAPTCHA は、現在最も広く利用されている CAPTCHA である。ボットによる文字の解読を阻止するために、様々な歪みを加えたランダムな文字列をユーザーに表示し、回答をテキストボックスに入力させ、正解であればそのユーザーが人間だと判断する仕組みとなっている。代表的な文字列 CAPTCHA に、Google ReCAPTCHA や Microsoft CAPTCHA(図 1) などが存在する。

[a)] 工学専攻機械・情報系コース大学院生
[b)] 情報システム工学科学部生
[c)] 情報システム工学科助教
[d)] 情報システム工学科准教授
[e)] 情報システム工学科教授

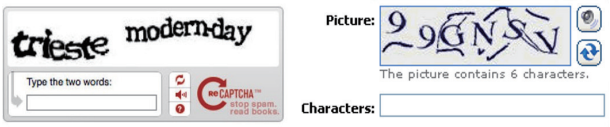

図 1. Google ReCAPTCHA(左) と Microsoft CAPTCHA(右)

図 2. アモーダル補完の例 (左) と提案 CAPTCHA で使用する立体文字の例 (右)

文字列 CAPTCHA を用いるメリットとして、Web サイト上で実装することが簡単であることと、総当たり攻撃に対して高い耐性を持つことが挙げられる。しかし、文字列 CAPTCHA は OCR を備えたボットによって突破可能であることが指摘されている [4]、[5]。

3. 提案方式

3.1 アモーダル補完を用いた CAPTCHA

提案手法ではアモーダル補完を利用する。アモーダル補完とは、対象物の一部が隠れて見えないとき、欠けた部分を脳内で補完して全体像を認識する視覚補完機能である。人間をはじめ、脊椎動物全般が有する視覚機能として知られている。例えば図 2 の (左) では、実際には存在しない三角形があたかも存在しているように見える。

提案 CAPTCHA の概要を以下に説明する。本手法では使用する文字を立体化したうえで、立体文字の前面部分と背面部分の色を、CAPTCHA の背景で使用されている色と同じ色 (黒色) にする。人間はアモーダル補完により、立体文字の側面部分の形状のみの情報からでも正しく文字を判読することが出来る。例えば図 2 の (右) ではアモーダル補完が生じ、“A” という文字が浮かび上がっているように見える。しかし、これはボットには判読が困難であると考えられる。

さらに、以下の 3 つの工夫を加えることで、ボット耐性をより強固なものにしている。

1 つ目は、立体文字の側面部分の色をランダムに選択するようにすることである。文字が生成される度に側面部分の色が異なるので、ボットは「文字の色の情報を文字ごとに記憶して、文字を判読する」ことが困難になる。

2 つ目は、文字を回転させることである。文字が回転することで、既存の動画 CAPTCHA と同様、ボットは文字の判読が困難になる。しかし人間にとっては、文字を様々な角度で視認出来るため、むしろその判読が容易になる。

3 つ目は、縦に回転する文字と横に回転する文字を双方用意し、問題文で指定した方向に回転する文字を回答させることである。ボットは問題文が示す内容を理解したうえで、文字が回転している方向が縦であるか横であるかを判別する必要があるため、正答を得ることがより困難になる。

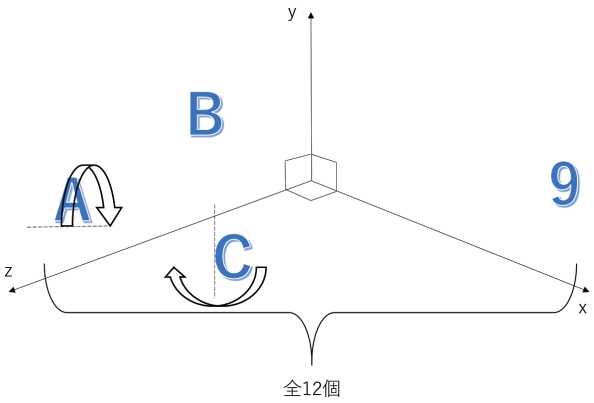

図 3. 3D 空間の構造と各立体文字の動き

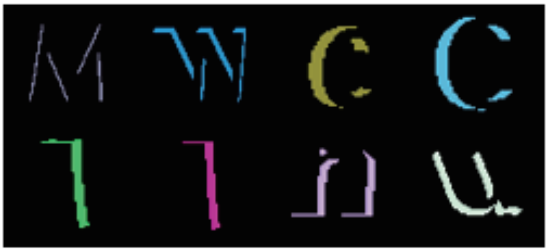

図 4. 読み間違いが多かった文字の一例

また、今回は、ユーザーに回答させる文字数を 6 文字に設定した。一般的に、ユーザーに回答させる文字数は 4 文字から 6 文字のいずれかであることが推奨されている [6] ので、その中でも偶然突破確率が一番低い 6 文字を採用することにした。縦回転する文字と横回転する文字をそれぞれ 6 文字ずつ用意するので、全体として 12 個の立体文字が画面上に表示される。図 3 は、提案 CAPTCHA の 3D 空間の構造と各立体文字の動きを示している。

3.2 提案 CAPTCHA で使用する文字

提案 CAPTCHA では基本的に数字、アルファベットの大文字、小文字を使用している。しかし、事前実験を行った結果を参考にし、以下に示す 17 文字を使用しないことにした。

大文字 I, M, O

小文字 b, c, d, l, n, o, p, q, s, u, v, w, x, z

これらの文字を使用しなかったのは、M は W、c は C、I は l、n は u などのように、それぞれ形状が似ている文字と読み間違えてしまうケースが多く見られたからである (図 4)。

3.3 実装

提案 CAPTCHA の開発言語は JavaScript である。提案手法では、WebGL のライブラリの一つである Three.js を用いて 3D 空間を作成し、その空間上に 12 個の立体文字を配置し表示する。配置された 12 文字のうち、左から 1、2、5、7、11、12 番目の 6 文字を縦回転、それ以外の 6 文字を横回転させる。立体文字の回転速度はランダムである。各立体文字の前面部分と背面部分の色を (R, G, B) = (0, 0, 0) と設定し、側面部分の色は RGB 値の 0 から 255 の範囲からランダムに選択する。各立体文字の大きさを 20px、幅を 2.5vh で統一し、フォントタイプは typeface.js 形式のフォントファイル

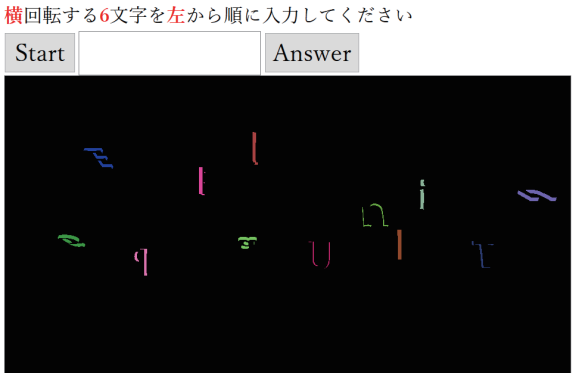

図 5. 提案 CAPTCHA の認証画面

である optimer_bold.typeface.json を使用している。3D 空間の背景の色は (R, G, B) = (0, 0, 0) であり、立体文字の前面部分、背面部分と同じ黒色で統一している。

3.4 認証手順

図 5 に、実装した CAPTCHA の認証画面を示す。CAPTCHA プログラムが動作を始めると、回答するべき文字が縦回転の方なのか横回転の方なのかを指示する問題文が画面の一番上に表示される。ユーザーが Start と書かれたボタンを押すと、3D 空間が生成され、縦もしくは横に回転する立体文字が表示される。ユーザーは問題文に従い、指定された方向に回転している文字を左から順にテキストボックスに入力し、Answer と書かれたボタンを押す。問題文で指定された文字列を正しく入力できていれば人間と判定される。

4. 利便性に関する評価実験

4.1 実験目的

実験目的は以下の二つである。実装した提案 CAPTCHA が人間にとって容易に回答が可能であるかどうかを正答率と回答時間から評価する。また、ユーザビリティ評価を行うことで CAPTCHA としての実用性を調査する。

4.2 実験方法

本実験の被験者は、宮崎大学に所属する学生 20 名である。

実験の手順を以下に示す。

(1) 説明

各被験者には、まず 3.4 で説明した、提案 CAPTCHA の回答手順を説明した。この時、提案 CAPTCHA で使用されている文字と使用されていない文字については、それらを記述したプリントを配布したうえで、使用されていない文字が存在する理由を具体例を挙げて説明した。

(2) 練習

文字列の入力習熟のために各被験者が十分慣れたと判断するまで練習を行ってもらった。

(3) 解答

提案 CAPTCHA を 10 回連続して解答してもらった。出題される文字列は毎回ランダムに選ばれるようにした。この時、各解答の正誤と開始ボタンをクリックしてから送信ボタンをクリックするまでの所要時間を実験者が記録した。また、被験者が CAPTCHA の解答に失敗した際には、間違えて入力した文字と正解である文字の組をそれぞれ記録した。

表 1. 実験・アンケート結果

平均正答率 [%]	91.0
平均所要時間 [秒]	21.4
最大所要時間 [秒]	114.0
最小所要時間 [秒]	7.0
平均 SUS スコア	73

(4) 評価

各被験者の測定終了後、ユーザビリティ評価を行うために、アンケート調査を行った。アンケート調査では、SUS(System Usability Scale) という、システムの使いやすさを測定するための指標を用いた。

4.3 SUS(System Usability Scale)

SUS(System Usability Scale)[7] は John Brooke が 1986 年に開発した、当時の文字ベースの PC の評価指標である。その後は携帯電話やハードウェア、IVR などの評価軸としても利用されている。

この指標は 10 項目のアンケートから構成されており、全ての項目を 0 から 4 で評価する。奇数項目がポジティブな質問、偶数項目がネガティブな質問となっている。奇数項目は回答番号から 1 を、偶数項目は 5 から回答番号を引き、それらを足しあわせた数値を 2.5 倍して 0 から 100 のスケールへ変換する。

スケール後の数値が高いほど、システムとして良い評価が与えられる。SUS スコアは、Jeff Sauro らによる調査結果 [7] から平均スコアが 68 とされており、ユーザビリティに優れた上位 10%に入るには、80.3 を超えるスコアが必要とされている。

4.4 実験結果と考察

被験者 20 名が 10 回ずつ提案 CAPTCHA を解いたときの 200 個のデータから得られた、平均正答率、平均所要時間、最大所要時間、最小所要時間および、SUS によるユーザビリティ評価の平均スコアを表 1 に示す。

代表的な文字列 CAPTCHA である ReCAPTCHA の平均正答率は 75.0%[8] であるので、本手法は正答率 (91.0%) においては十分実用的であると考えられる。被験者が解答に失敗した原因として、形状が似ている文字や側面部分が暗色である文字を判読することが難しいといった理由があげられた。この問題は、文字のフォントをより明確に判読できるものに変更することや、側面部分の RGB 値の取る範囲を見やすいものに制限することで改善が可能であると考えられる。

一方、平均所要時間 (21.4 秒) は、ReCAPTCHA の平均所要時間 11.9 秒 [8] と比べて長く、改善の必要があると考えられる。この問題は、入力文字数を 6 文字から 4 文字に減らすなどの変更を行うことで改善が可能であると考えられるが、同時に偶然突破確率が高くなり、総当たり攻撃耐性が低下してしまうといった問題が懸念される。しかし、一般的に推奨されている偶然突破確率は 1/4,096 とされており [4]、提案 CAPTCHA の文字数を 4 文字に減らした場合の偶然突破確率は $1/44^4 = 1/3{,}748{,}096$ であるので、文字数を 4 文字に減らしても総当たり攻撃耐性は十分なレベルに維持できると考えられる。

今回の SUS に基づいたアンケート調査のスコアは 73 であり、平均スコアである 68 を上回る結果となったので、本手法は最低限のユーザビリティを確保できると考えられる。しかし、多くの被験者が提案 CAPTCHA を利用し始める前に知っておくべきことが多くあると回答した。これは提案 CAPTCHA で使われていない文字を被験者が事前に把握する必要があったことが原因であると考えられる。この問題は、例えば使用する文字をアルファベットの小文字を除く、数字と大文字の 2 種類に縮減することで、被験者は容易に CAPTCHA で使われていない文字を把握することが容易になると考えられる。しかし、使用する文字の種類を縮減することで総当たり攻撃耐性が低下する危険性がある。ユーザビリティを確保しつつ、十分な総当たり攻撃耐性を持つような文字の使用範囲を今後検討する必要がある。

5. ボット耐性の検証実験

5.1 実験目的

提案手法の CAPTCHA に対して、OCR による文字認識とテンプレートマッチング法による物体認識を行い、その結果から、ボットにとって提案手法の CAPTCHA が困難なものになり得るかを検証する。

5.2 文字認識技術への耐性の検証実験

OCR とは画像データ上にある文字と思われる部分を解析し、コンピューター上で扱える文字 (テキスト) データに変換する技術のことである。

提案 CAPTCHA が文字認識技術に耐性を持つことを確認する実験を行った。文字認識は、Google Cloud Platform が提供するサービスの一つである Google Cloud Vision API と呼ばれる画像分析 API を利用した。実行中の提案 CAPTCHA をスクリーンショット機能を使用して 0.1 ミリ秒間隔で撮影した 300 枚の画像を入力画像として保存した。保存した入力画像を 1 枚ずつ、Google Cloud Vision API の OCR 機能を使用して文字認識を行い、テキストデータに変換した。

実験の結果、正しく認識出来た文字は存在しなかった。

5.3 物体認識技術への耐性の検証実験

画像の中から特定の文字や模様などのパターンを検出する手法の一つに、テンプレートマッチング法がある。テンプレートマッチング法とは、入力画像の中からテンプレート画像に似た画像パターンを探し出す手法である。テンプレート画像を被検出画像上でスライドさせ、テンプレート画像と被検出画像の領域を比較し、類似度の高い領域を検出することで物体認識を行う。

提案 CAPTCHA がテンプレートマッチングに耐性を持つことを確認する実験を行った。テンプレートマッチングは、画像処理ライブラリである OpenCV
(Open Source Computer Vision Library) を利用した。提案 CAPTCHA で使用する 44 個の立体文字を、人間にとって一番判読が容易であると思われる角度で配置し、それらを 1 文字ずつ撮影したものをテンプレート画像として使用した (図 6)。文字認識の耐性の検証実験と同様の手法で撮影した 300 枚の画像を 1 枚ずつグレースケールで読み込み、44 枚のテンプレート画像のそれぞれとテンプレートマッチングを実行した。マッチングが成功すれば、検出結果から検出領域の位置を取得し、検出領域が四角で囲んで保存される。

図 6. テンプレート画像の例

実験の結果、正しく検出出来た文字は存在しなかった。

5.4 考察

実験結果から、提案 CAPTCHA は OCR による文字認識とテンプレートマッチング法による物体認識への耐性を持つことが確認できた。

しかし、テンプレートマッチング法で使用した、各文字に対して 1 枚ずつ撮影したテンプレート画像だけでは不十分だった (違った角度で撮影したテンプレート画像を多数用意しておけば、マッチするものが存在していた) 可能性がある。今後はテンプレート画像の枚数をさらに増やしてテンプレートマッチングを行い、物体認識の耐性の検証実験を行う必要がある。

6. まとめ

本論文では、人間の高度な視覚補完機能であるアモーダル補完に着目し、自動プログラムによる攻撃への耐性を持たせた CAPTCHA を提案した。提案方式の実装、利便性に関する評価実験、ボット耐性の検証実験を行い、本方式の有用性を示した。

今後の課題として、提案 CAPTCHA で使用する立体文字のフォントや側面部分の RGB 値がランダムに選択される範囲の見直し、文字の使用範囲の検討、機械学習などの新たな攻撃耐性に関する理論的評価などがあげられる。また、テンプレートマッチングを行う際に使用するテンプレート画像の枚数を増やすことについても検討予定である。

参考文献

1) L. von Ahn, M. Blum, N. Hopper, and J. Langford: CAPTCHA: Telling humans and computers apart, Advances in Cryptology, Eurocrypt' 03, Vol.2656 of Lect. Notes Comput. Sci., pp.294-311, 2003.

2) 藤田 真浩, 池谷 勇樹, 可児 潤也, 西垣 正勝: Locimetric 型メンタルローテーション CAPTCHA, 情報処理学会論文誌, Vol.57, No.9, pp.1954-1964, 2016.

3) George D, Lehrach W, Kansky K, et al.: A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs, Science, Vol.358, No.6368, 2017.

4) 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.

5) Yan, J. and EI Ahmad, A.S.: Breaking Visual CAPTCHAs with Naive Pattern Recognition Algorithms, Proc. ACSAC2007, pp.279-291, 2007.

6) E. Bursztein, S. Bethard, C. Fabry, J. C. Mitchell, and D. Jurafsky: How Good Are Humans at Solving CAPTCHAs? A Large Scale Evaluation, 2010 IEEE Symposium on Security and Privacy, pp.399-413, 2010.

7) https://measuringu.com/sus/, (accessed 2019/01/23)

8) Marek R. Ogiela, Natalia Krzyworzeka, and Lidia Ogiela: Application of knowledge-based cognitive CAPTCHA in Cloud of Things security, Concurrency and Computation Practice and of Experience, Vol.30, No.21, 2018.