色の恒常性を利用したCAPTCHAのユーザビリティと機械攻撃耐性向上のための検討

臼崎　翔太郎[a)]・砂本　佑紀[b)]・岡崎　直宣[c)]・山場　久昭[d)]・油田　健太郎[e)]

A Study on Improving the Usability and Machine Resistance of Color Constancy CAPTCHA

Shotaro USUZAKI, Yuki SUNAMOTO, Naonobu OKAZAKI, Hisaaki YAMABA, Kentaro ABURADA

Abstract

CAPTCHA is now widely used to prevent the bot from sending unauthorized service requests. CAPTCHA is a system that imposes tasks that are easy for humans and difficult for machines and thereby detect access from the bot. Unfortunately, many researchers have already reported that bots can break current CAPTCHAs by OCR and machine learning technology. Therefore, it is necessary to add interference for machine resistance. However, such interference also has a negative effect on human perception. Therefore, we adopt a color constancy to CAPTCHA. The color constancy is a human visual characteristic and is known to be difficult to reproduce mathematically. However, this CAPTCHA has a problem that human success is low and machine success rate is high. To solve this issue, we proposed an improved CAPTCHA by processing the base image and creating a new color interference filter. The first contrivance is to leave only a specific object color in the base image for defending the attack by the Gray-World algorithm, which is vulnerable to color bias. The second contrivance solves the weakness that the color filter does not always work effectively. We evaluated the effect of these devices on the accuracy rate of humans and machines and usability.

Keywords: CAPTCHA, color constancy, gray-world, gray-scale, system usability score

1. はじめに

ボットによるアカウントの不正な大量取得や、不正なサービス要求を大量に行うDoS攻撃が発生している。これらの問題を解決する対策として、一般的なWebサービスではCAPTCHAが用いられている。CAPTCHAは機械と人間を区別するためのチューリングテストであり、ユーザに対して問題を出し、人間か機械かを区別するシステムである。CAPTCHAの代表例として、歪んだ文字列をユーザに読み取らせる文字列型CAPTCHAや、表示された複数の画像の中から条件に合う画像をユーザに選択させる画像型CAPTCHAがある。これらの手法はOCR技術や機械学習技術の発達により突破されつつある。しかし、既存の手法の難易度を上げると、機械による解読を阻止することができる一方で、人間による解読も困難になってしまう。

そこで、人間の視覚特性である色の恒常性を利用し、ベース画像に色妨害フィルターを加えた画像中の色の区別をさせるCAPTCHAが提案された[1)]。しかしこのCAPTCHAには、人間の正解率が低く、機械の正解率が高いという課題がある。そこで本研究では、ベース画像の加工手法と新しい色妨害フィルターの作成手法を提案し、人間の正解率向上と機械の正解率低下を目指す。またこれらの工夫によって、人間の正解率、機械の正解率に与える影響を検証する。

以下、2章では先行研究を紹介し、その問題点を指摘する。3章では提案手法、ベース画像、色妨害フィルター、解答の照合方法について解説する。4章では、提案手法の性能評価実験について、その目的と内容を説明し、その後、実験を行った結果について考察を行う。5章ではまとめと今後の課題について述べる。

2. 先行研究

2.1 color CAPTCHA

colorCAPTCHAは、コンピュータには色の名前を認識することが困難であるが、人間には容易に色の名前を認識できることを利用したCAPTCHAである[2)] これは、表示されるカラー画像の中から、指定された部分の色の名前を答えさせるというものである。評価実験では、職種を問わず5歳以上の1,000人にcolorCAPTCHAを解かせた。その結果、「色の名前を知らない」か「入力した色の名前のスペルにミスがある」という2つの事柄を除いて、正答率が100%であり、colorCAPTCHAは従来の文字列型CAPTCHAや画像型CAPTCHAよりも優れていることが分かった。この手法ではボットが色から名前を認識できないことを前提に提案されているため、これまで

[a)]物質・情報工学専攻生産工学教育コース大学院生
[b)]情報システム工学科学部生
[c)]情報システム工学科教授
[d)]情報システム工学科助教
[e)]情報システム工学科准教授

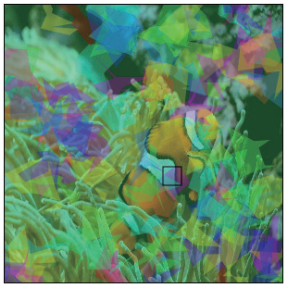

図 1. 色の恒常性を利用した colorCAPTCHA の出題例 [1)]

の文字型 CAPTCHA や画像型 CAPTCHA のように妨害が必要なく、これによって正答率が高くなったとしている。ただし、ボットへの耐性は実験の評価対象とされておらず、現在の機械学習の技術の進歩を考えると、色から名前を認識できる可能性が高いため、セキュリティ面においては十分に耐性があるとはいいがたい。

2.2 色の恒常性を利用した color CAPTCHA

colorCAPTCHA のように、画像に何の妨害も加えずに、単に見えている色の名前を答えさせるだけではボットに突破されてしまう恐れがあるため、人間の正解率を保証しつつセキュリティを向上させる、色の恒常性を利用した colorCAPTCHA が提案された [1)]。この手法は、色の恒常性という人間に自然に備わっている高度な認知能力を利用している。色の恒常性とは「周囲の照明光の影響を受けても本来の色を知覚できる」というものである。色の恒常性の原理は完全には解明されておらず、アルゴリズムとして表現することが困難であることが知られている。色の恒常性により人間には容易に色を認識できるが、機械にとっては再現が困難であり、これがそれぞれ本 CAPTCHA のユーザビリティとセキュリティの前提となっている。具体的には、色妨害フィルターを加えた画像中の解答領域の色と最も近いと思う色を、11 色のカラーパレット（黒、白、灰、赤、青、緑、黄、紫、茶、オレンジ、ピンク）を用いてユーザに解答させる。ユーザはドラッグやクリックにより、画像中の任意の位置に解答領域を移動させることができる。最終的にカラーパレットの色と、解答領域におけるオリジナル画像の色との色差が最も近いものをユーザが選択した場合を正解としている。

色妨害フィルターには、単色フィルターとシェイプ型フィルターの 2 種類を用いているのだが、シェイプ型フィルターは動的なもので、図形の発生と消滅を繰り返しており、このCAPTCHA を数秒ほど録画した動画のフレーム平均をとることで、シェイプ型フィルターを取り除くことができる。これによって、単色フィルターのみの妨害となってしまうため、機械耐性が弱くなってしまう。

またこの CAPTCHA は、人間の正解率が低いことが課題としてあげられている [1)]。

3. 提案手法

3.1 概要

CAPTCHA に使用する画像は、ベース画像と色妨害フィルターで構成されている。本研究で提案するのは、人間の正解率の向上と機械の正解率の低下を目的とした、ベース画像の

図 2. カラーパレット [1)]

加工手法と色妨害フィルターの作成手法である。CAPTCHA の出題形式そのものは先行研究である「色の恒常性を利用した colorCAPTCHA」と変わっておらず、色妨害フィルターを加えた画像中の解答領域の色と最も近いと思う色をカラーパレットを用いてユーザに選択させるものである。カラーパレットは、後述するベース画像をグレースケール化する都合で、黒、白、灰を選択肢から除外している。

3.2 ベース画像

先行研究では色妨害フィルターを重ねる以外に画像に加工は加えていない。しかし本研究では、色妨害フィルターの作成、色恒常性アルゴリズムに対する耐性を持たせるために、画像を加工している。

事前実験において、Gray-World という色恒常性アルゴリズムに強力な色妨害の除去効果があることがわかった。文献 [3)] によると、Gray-World は灰色仮説という、画像の全ピクセルの RGB 値の平均を取ると灰色になるという仮説に基づいたアルゴリズムである。しかし、色に偏りがある画像の場合にはこの仮説が成り立たないため、図 3 と図 4 のような画像を用意し、画像の一部分を除いてグレースケール化を行っている (図 5)。この工夫によりベース画像の色を灰色に偏らせ、Gray-World に耐性を持たせることにする。

また、ベース画像の色を偏らせることで、色妨害フィルターの色とベース画像の色が重なりにくくするという意図がある。色が重なっている状態というのは、ベース画像の座標 (x,y) と色妨害フィルターの座標 (x,y) の画素が同じ色（画素値が完全に一致する）か、近い色（画素値の色差が小さいもの）である状態である。このような状態だと、機械が正解の色を答えられる可能性が高くなるという問題がある。

3.3 色妨害フィルター

本研究の色妨害フィルターは、図 6 のような色抜き画像を元に決定した 3 色を用いたグラデーション画像である。2.2 で説明したように、動的なフィルターでは除去耐性が低いため、静的なフィルターとしている。事前実験において、色恒常性アルゴリズムが単色のフィルターに強いことがわかっていたので、複数色を用いたグラデーションの静的な画像を色妨害フィルターとした。作成手順は以下のようになっている。

1. 画像を 3 等分する。
2. 3 等分した画像のうち、最も多く色が残っている画像の平均色 (RGB) を取得し、HSV に変換する。
3. 平均色 (HSV) から以下のような 3 色を作成する

$$(H,S,V) = \begin{cases} ((H+180) \bmod 360, 80, 80) & (1) \\ ((H+90) \bmod 360, 80, 80) & (2) \\ ((H-90) \bmod 360, 80, 80) & (3) \end{cases}$$

図 3. ベース画像 [4)]

図 4. マスク画像 [4)]

図 5. グレー化画像 [4)]

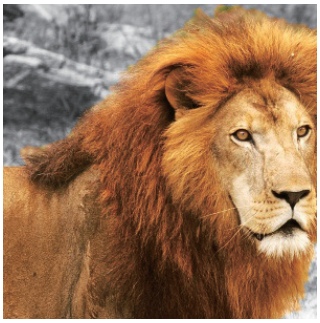

図 6. 図 5 を 300x300 にリサイズした画像

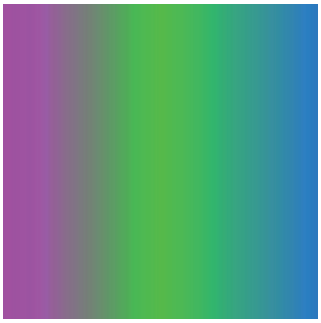

図 7. 色妨害フィルター

4. 作成した 3 色を使って色妨害フィルターを作成する (図 7)。

色妨害フィルターに使われている色がベースとなる画像の色とできるだけ重ならないようにするため、手順 3 のように色を作成している。ベースとなる画像に残っている色の種類が多いと、色の重なりを避けて妨害フィルターを作成することが難しくなるため、今回はベースとなる画像に残る色が 1、2 色になるものを選んでいる。また、色 (1) は手順 2 で取得した平均色の補色となっており、補色が重なっている領域が人間にとって最も色を識別しやすくなるのではないかと考えたため、色妨害フィルターを作成する時、3 等分した画像のうち最も多く色が残っている画像に色 (1) が重なるようにしている。

3.4 解答の照合方法

ユーザが CAPTHCA を成功したとするのは、ユーザが出題画像における解答領域内の色と最も近いと思ったカラーパレットの色と、ベース画像における解答領域内の色と最も近いカラーパレットの色が一致した時である。解答の照合は、ユーザが選択したカラーパレットの RGB 値、解答領域の座標、ベース画像を用いて行う。具体的にはまず、ベース画像における解答領域内の RGB 値を求め、その値とカラーパレットの各色との色差を計算し、色差が最も小さかった色を正解の色とする。そしてその色がユーザの選んだ色と一致していれば成功とする。本研究では人間の色の見えを考慮して解答の照合ができるようにするために、色差の計算に CIEDE2000[5)] を使用した。CIEDE2000 は CIE(国際照明委員会) が 1976 年に $L^*a^*b^*$ 色空間上の 2 点間のユークリッド距離を規定した計算式を人間の色の違いによる感度、即ち人間の特性を色差を求める計算式に組み込む修正を行った色差を求める計算式である.

4. 実験・評価

4.1 実験の目的

本論文で提案した手法によって、人間の正解率と色妨害フィルターの機能の有無が、先行研究と比較してどのように変化するのか、また、色恒常性アルゴリズムを適用した場合の機械正解率がどの程度あるのか調査することが目的である。加えて、SUS(System Usabiliity Scale)[6)] と呼ばれるユーザビリティの数値的な評価が可能である指標を用いてアンケート調査を行い、提案手法のユーザビリティの調査も行った。

表 1. 人間の正解率 [%]

本研究	先行研究
76.67（138/180）	71.58（68/95）

表 2. 色妨害フィルターが機能していなかった画像の割合 [%]

本研究	先行研究
8.33（15/180）	21.05（20/95）

4.2 実験内容

被験者は、宮崎大学工学部の学生 18 人である。実験画像には、現実世界で撮影された人物、乗り物、動物、風景等の画像を 49 枚用意し、画像サイズは 300px×300px とした。今回の画像は Open Image Datast V5[4] のものを使用している。画像をグレースケール化した時、グレーではない部分が 3 色以上で構成されるような画像は、色妨害フィルターとの色の重なりによって妨害効果が弱まる可能性があるため、実験画像に含めていない。被験者には、ランダムに選ばれた 10 枚を出題した。また、出題画像中に色がある箇所とない箇所があることを説明し、解答領域は被験者が正解できる自信がある箇所を選んでもらった。

4.3 結果と考察

4.3.1 人間正解率

人間の正解率について、本研究で行った被験者 18 人に対する実験結果から得られた正解率と先行研究で行われた被験者 15 人に対する実験結果から得られた正解率を表 1 に示す。

表 1 より、人間の正解率は約 5 ポイント上昇したことがわかる。このことから、色妨害フィルターを作成する時に色の恒常性が働きやすい色をうまく設定できたのではないかと考えられる。

しかし、人間の正解率は上昇したものの 8 割には到達しておらず、標準的な CAPTHCA である reCAPTCHA[7] の正解率の 84.1%を下回ってしまった。これをさらに上昇させるには、より良い色妨害フィルターの作成、ベース画像の選択、カラーパレットに使う色の選択が必要であると考える。

4.3.2 妨害フィルター機能成功率

本研究と先行研究の色妨害フィルターの機能の有無について、フィルターが機能していない画像の割合を表 2 に示す。

表 2 より、フィルターが機能していない画像の割合は約 10 ポイント減少したことがわかる。このことから、色妨害フィルターを作成する時に、ほとんどの画像で色の重なりのない色の選択ができたのではないかと考えられる。今後フィルターが機能していない画像を 0% にするためには、色妨害フィルターを構成する色の決定方法や、ベース画像の選別について見直すことが必要である。

表 1、表 2 より、提案手法の工夫によって人間の正解率が向上し、色妨害フィルターの機能していない画像が減少したことが分かる。このことから、画素値が大きく変化するような色妨害フィルターの彩度と明度の設定、ベース画像との合成比率を設定できたのではないかと考えられる。

4.3.3 機械耐性

色恒常性アルゴリズムである Gray-World を画像に適用した場合の機械の正解率を表 3 に示す。

表 3. Gray-World を画像に適用した時の機械の正解率

色恒常性アルゴリズム	正解率 [%]
Gray-World	25.00（45/180）

表 3 より、Gray-World を画像に適用した場合、4 分の 1 が機械に正解されてしまっていることがわかる。Gray-World に耐性を付けるのなら、ベース画像をグレースケール化する以外の画像の加工や、本研究で提案した色妨害フィルターの改良、もしくは全く新しいフィルターを考える必要がある。

4.3.4 ユーザビリティ

本研究における提案 CAPTCHA の平均 SUS スコアは 83.05 であった。[8] によると SUS の平均スコアは 68 とされている。さらに、ユーザビリティに優れた上位 10%に入るには、SUS スコアが 80.3 を超えるスコアが必要とされている。したがって提案 CAPTCHA の SUS スコアはかなり高い値と言える。また、先行研究における提案 CAPTCHA の平均 SUS スコアは 83.17 であり、もともと高かったスコアを維持しつつ CAPTCHA の改良ができたと考えられる。

5. まとめと今後の課題

本研究では、先行研究である色の恒常性を利用した colorCAPTCHA の課題解決のため、ベース画像の加工、色妨害フィルターの作成と、それらの効果を確認する実験、評価を行った。

先行研究では人間の正解率の低さに加え、ベース画像との色の重なりにより色妨害フィルターが意味を成していない画像が 20%を超えていた。また、動的な色妨害フィルターが除去できてしまうという問題があった。そこで、人間の正解率の向上, 色妨害フィルターのベース画像との色の重なりの回避および除去耐性の向上を考慮した、ベース画像の加工法と色妨害フィルターの作成法を検討した。

実験の結果、人間の正解率については約 5 ポイント程度ではあるが、先行研究よりも向上した。加えて、色妨害フィルターが機能していない画像は、先行研究から約 10 ポイント削減でき、新たな色妨害フィルターの作成方法は有用性があるのではないかと考えられる。また、ユーザビリティの評価の指標となる SUS は先行研究とほとんど変わらなかったことから、ユーザビリティを損なうことなく CAPTCHA の性能を向上させることができたといえる。

色恒常性アルゴリズムである Gray-World を使った場合の機械による正解率は、25.00%と非常に高い割合になってしまった。

今後の課題として、人間の正解率を向上させること、Gray-World への耐性を獲得することが挙げられる。そのためには、より細かい色妨害フィルターの色の選択や彩度・明度・透明度の設定、解答の照合に用いるカラーパレットと色差の計算方法、画像の選別などが必要である。

参考文献

1) 藤竜成, 川上翔平, 山場久明, 油田健太郎, 岡崎直宜: 色の恒常性を利用した colorCAPTCHA の提案 , 宮崎大学工学部紀要, Vol.48, pp.223-227, 2019.

2) M. Kumar and P. R. Dhir: Design and Comparison of Advanced Color based Image CAPTCHAs, International Journal of Computer Applications, Vol.61,No.15,pp.24-29, 2013.

3) J. V. D. Weijer,T. Gevers and A. Gijsenij: Edge-Based Color Constancy,IEEE Transactions on Image Pro-cessing, Vol.16, No.9, pp.2207–2214, 2007.

4) Google: Open Image Dataset v5, https://storage.googleapis.com/openimages/web/factsfigures.html,(accessed 2020/01/28).

5) K. Minolta: 新しい色差式（CIE DE2000）について,https://www.konicaminolta.jp/instruments/knowledge/color/section2/06.html, (accessed 2020/01/28).

6) J. Brooke: SUS—A Quick and Dirty Usability Scale, Usability Evaluation in Industry, Taylor and Francis, 1996.

7) N.Jiang, H.Dogan and F.Tian: Designing mobile friendly CAPTCHAs: An exploratory study, Proceedings of the 31st British Computer Society Human Computer Interaction Conference, Vol.92, pp.1-7, 2017.

8) J. Sauro: MEASURING USABILITY WITH THE SYSTEM USABILITY SCALE (SUS),Measuring U,https://measuringu.com/sus/, (accessed 2020/01/28).