色の恒常性を利用した colorCAPTCHA の提案

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

Proposal of ColorCAPTCHA Using Color Constancy

Ryusei FUJI, Shohei KAWAKAMI, Hisaaki YAMABA, Kentaro ABURADA, Naonobu OKAZAKI

Abstract

Activities by bots such as illegally acquiring accounts, sending spam mail and participating in online voting are problem. In order to counter these problems, a technology called CAPTCHA to discriminate humans and bots was developed. CAPTCHA is widely used on the web sites to prevent the spam behavior. CAPTCHA, which is currently the mainstream, has a function that allows a user to solve a character displayed on the screen or to distinguish what appears in a photograph. These existing CAPTCHAs are getting easily broken by the development of OCR technology and machine learning technology. In order to counter this problem, we propose a new CAPTCHA, which is called color homeostasis, utilizing advanced cognitive ability naturally provided to humans and difficult to imitate by bots. The proposed CAPTCHA uses an image obtained by adding a color interference filter to the image. We use 11 basic colors (black, white, red, green, yellow, blue, brown, purple, pink , orange, gray) for the answer. The answer area is displayed at a random position, but the user can move the answer area to an arbitrary place by dragging or clicking. A case in which the color selected by the user from the color palette selected is the closest color to that of the original image, the color selection is regarded as correct.

Keywords: CAPTCHA, color, color constancy, security

1. はじめに

ボットによるアカウントを大量に不正取得しそれらを用いてスパムメールを送信することやボットがオンライン投票に参加するなどの不正行為が問題視されている。このような問題に対抗するためには CAPTCHA と呼ばれる人間とボットを判別する技術が開発された [1)]。CAPTCHA はスパム行為を防ぐために Web 上で広く利用されている。CAPTCHA の基本原理は人間には容易であるが、機械には困難である問題をユーザに出題し、解答者が人間であるかを判別する技術である。現在主流である CAPTCHA はユーザに画面上に表示されている文字を解答させるものや写真に写っているものを判別させるものなどがある [2)3)]。

これらの既存の CAPTCHA は時代に伴う OCR 技術 (文字認識技術) や機械学習技術の発達により、容易に突破されるようになってきている。このような問題に対抗するために、色の恒常性と呼ばれる、人間に自然に備わっていてボットによる模倣が難しい高度な認知能力を利用した新たな CAPTCHA を検討する。

[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

2. 関連研究

2.1 3Dtext ベース CAPTCHA

3Dtext ベース CAPTCHA とは OCR(Optical Character Recognition) 技術の向上により、従来の 2Dtext ベースの CAPTCHA の限界を克服するために検討された CAPTCHA(図 1) である。画像から 3D オブジェクトを知覚する人間の視覚システムに自然な能力を利用するように設計されており、コンピュータプログラムが 3D コンテンツを識別することが困難であることが基本的なセキュリティの前提となっている。代表的な 3Dtext しては、人気のあるブログ管理ツール (ワードプレス) などで使われている Super CAPTCHA や、OCR Research Team により設計された、従来の 2Dtext ベース CAPTCHA に見られる設計の欠陥と弱点を分析することによって設計された Teabag3D などがある。しかし文献 [2)] によると、これらは短時間で高い精度で解かれてしまうことが報告されている。このように画像認識技術や OCR 技術の向上により、textCAPTCHA ではセキュリティを保つことが難しくなってきている。

2.2 Google reCAPTCHA

textCAPTCHA ではボットに読み取られてしまうため、様々な CAPTCHA が開発されている。その中の一つが image-CAPTCHA である。最も代表的な例でいえば、Google 社が開発した Google reCAPTCHA というシステム (図 2) である。reCAPTCHA は、ユーザに画像を選択させて人間とボットを識別する。具体的には、CAPTCHA の画面上部には見本の画像あるいは画像に関する説明文が表示され、画面下部に

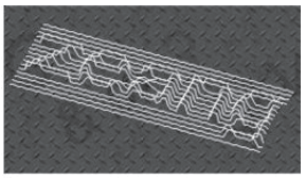

図 1. 3DtextCAPTCHA

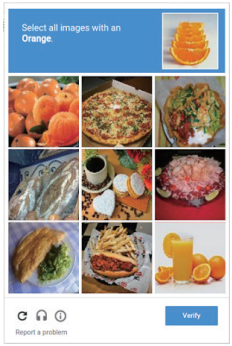

図 2. imageCAPTCHA

出題された 9 つの画像から説明文に沿った画像をすべて選択させる CAPTCHA である。このシステムは textCAPTCHA のように今までの歪んだ文字列を読み取って入力する作業はないため、textCAPTCHA よりも人間にとって非常に容易であるが、ボットにとっては多様な映り方をしている物体を網羅的に認識することが困難なのでその性質を利用して人間とボットを区別しようとしている。しかし、文献 [4] によれば、Facebook の imageCAPTCHA はボットに 83.5 %の精度で解かれてしまうことが報告されている。

2.3 colorCAPTCHA

これまで紹介した CAPTCHA はいずれもセキュリティ面が脆弱であった。これらの手法のセキュリティを高めるためには、文字や画像にさらなる妨害を加えることが考えられるが、その分人間の CAPTCHA 成功率が下がってしまうおそれがある。Kumar らは、コンピュータは色の名前を認識することが困難で、人間は容易に色の名前を認識できることを利用した colorCAPTCHA を提案した [5]。

Kumar らの手法では、CAPTCHA の出題としてランダムで選択されたカラー画像から、指定された部分の色、画像にあるオブジェクトの色をユーザに答えさせる。

評価実験では、職種問わずに 5 歳以上の 1,000 人に colorCAPTCHA を解かせているが、その結果、colorCAPTCHA は従来の textCAPTCHA や imageCAPTCHA よりも、「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解率の高い 100 %であることが分かった。Kumar らの手法では、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、これまでの textCAPTCHA や imageCAPTCHA のように妨害が必要なく、これによって正解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の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非常に高いため、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。そこで本研究では、色の恒常性という人間の高度な知覚能力を利用することで、従来の colorCAPTCHA と同程度の人間の正解率を保証しつつセキュリティを向上させる、新たな colorCAPTCHA を検討する。

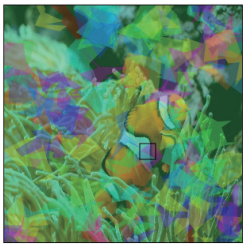

図 3. 出題例

図 4. カラーパレット

3. 提案手法

3.1 提案 CAPTCHA

提案する CAPTCHA は、画像に色妨害フィルターを加えた画像から、回答エリアと呼ばれる領域の色と近い色を、基本色 11 色 (黒、白、赤、緑、黄、青、茶、紫、ピンク、オレンジ、灰色) のカラーパレットを用いてユーザに回答させる。回答エリアはランダムな位置に表示されるが、ユーザは回答エリアをドラッグ、あるいはクリックすることで任意の場所に移動させることができる。ユーザがカラーパレットから選択した色が、ユーザの解答領域においてオリジナル画像（色妨害フィルタを加える前の画像）の色に一番色差の近い色を選択した場合を正解とした。提案手法の出題例を図 3 に、カラーパレットの画像を図 4 にそれぞれ示す。この手法は色の恒常性の「周囲の照明光の影響を受けても本来の色を知覚できる。」という人間に自然に備わっている高度な認知能力を利用することにより、人間にとっては正解の色を選択させることが可能だが、ボットは色の恒常性の原理をアルゴリズムとして表現することが困難であるということが基本的なセキュリティの前提である。

3.2 色の恒常性

色の恒常性とは、照明光の条件が変わってもその照明光の色に引きずられることなく、同じ物体は安定して同じ色として知覚させられる色覚特性である。通常の環境内では照明光の強度や分光特性は様々に変化する。したがって、表面からの反射光の輝度や色度もこれに合わせてさまざまに変化する。しかし、人間は照明環境が変わっても同じ表面を同じ色であると認識してしまう。例えば赤いリンゴは青い照明の下でも赤く感じられる。このような人間に自然に備わっていてボットによる模倣が難しい高度な認知能力を利用することにより人間とボットを区別することが提案 CAPTCHA の目的である。

3.3 偶然突破認証確率

画像の解答領域において 11 色のカラーパレットから 1 回選択しただけで人間とボットを判断することは、偶然突破認証確率に問題が生じると考えた。カラーパレットからランダムに色を選択するとなると 1/11 である色が選ばれる。11 色のカラーパレットのどの色を選択しても正解の色が当たる確率が均等であると仮定すると (実際には使用される画像の RGB 値の配分率によって変わる。) ボットがランダムに色を選択した場合、正解する確率は 1/11 となる. 文献 [6] によると、CAPTCHA の偶然突破認証確率として、1/4096 を確保できれば、Token Buckets Scheme を用いて誤答が多いユーザからのアクセスを

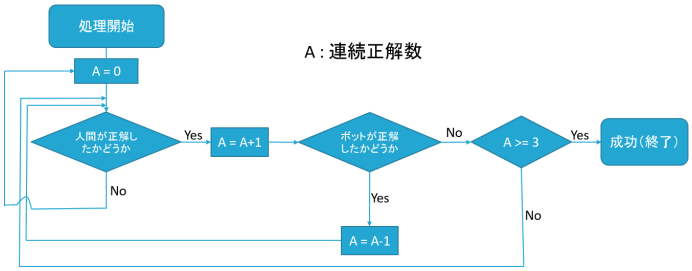

図 5. フローチャート

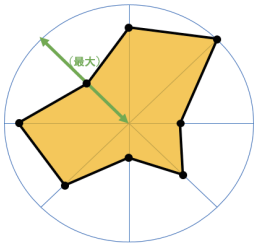

図 6. シェイプ型 (単体)

図 7. シェイプ型 (全体)

遮断することで、実質的な総当たり数を 560 万通りまで高めることが可能であることが示されている。そこで、本論文の提案 CAPTCHA においても、1/4,096 が CAPTCHA が有するべき総当たり数であると想定すると、提案 CAPTCHA において 1 回のみ正解の色を選択したことで人間と判別することにより生じる偶然突破認証確率は 1/4,096 と比べると高い値となる。よって、提案 CAPTCHA においては、平均偶然突破認証確率が 1/4,096 より低い確率になるように検討した。ボットの偶然突破認証確率を下げるために、CAPTCHA 成功には連続正解することを必要とした。CAPTCHA 成功までのアルゴリズムを図 5 に示す。提案 CAPTCHA には内部ボットプログラムを搭載した。これは、色妨害が施されている画像の解答領域の色を抽出して 11 色のカラーパレットから一番色差の値が小さい色を選択するプログラムである。この内部ボットプログラムが正解してしまうと色妨害の効果が期待できないとみなすことができる。さらに、色の恒常性の効果が働くことが期待できないのでボット耐性が著しく低下すると考えられる。

ユーザに表示される画像は正解不正解に関わらず、1 回解答すれば別の画像をランダムで表示するようにした。人間が 3 回連続正解し、内部ボットプログラムが 3 回連続不正解となることが提案 CAPTCHA では理想であるが、この場合、1 回の正解で CAPTCHA を成功する際の偶然認証突破確率を 1/11 と仮定すると、3 回連続で正解することを CAPTCHA 成功の条件にした際の偶然認証突破確率は 1/1,331 となる。しかし、予備実験において、内部ボットプログラムの正答率は 2 割程度あることが確認でき、ユーザが 4 回以上の連続正解を約 50% の確率で強いられることが計算上、予想できる。よって、ユーザが 4 回以上連続正解を強いられたときの偶然突破認証確率は 1/14,641 以下の数値となる。また、提案 CAPTCAH は最低でも 1/1,331 の偶然突破認証確率を有することが分かっているので、平均偶然突破確率を考えると 1/4,096 を下回る数値を期待できることを考慮し、CAPTCHA 成功に必要な最低連続正解回数を 3 回と定めた。Google reCAPTCHA の偶然認証突破確率は 1/511 であることを考えると、提案 CAPTCHA の最低偶然認証突破確率の 1/1,331 という数値は悪くないと考えた。

3.4 妨害色フィルタ

提案 CAPTCHA で使用している色妨害フィルタは 2 種類あり、単色の色妨害フィルタと図形を画像に多数表示させるフィルタである。本論文では後者の色妨害フィルタをシェイプ型色妨害フィルタと呼ぶことにする。単純に単色のフィルタを重ねるだけではボット耐性が十分でないことが確認できた。

1 つ目の単色色妨害フィルタは画像全体に色妨害を施している単色のフィルターであり、色に関しては HSL 色空間を用いている。HSL 色空間は色相、彩度、輝度からなり、色相はランダム、彩度は 100%、輝度は 50%に設定した。色相の範囲は 0〜360%、彩度の範囲は 0〜100%であり、100%で純色である。輝度の範囲は 0〜100%であり。50%で純色である。本論文の実験においては透明度は写真画像では 70%, イラスト画像では 50%に設定した。透明度に関しては予備実験を行い、人間に色の恒常性が働き、高い確率でオリジナル画像の色が認識できる限界を決めた。ただし、この値は 1 人のみの予備実験で決めたため、最適な値とは限らない。

2 つ目の色妨害フィルターは図 6 に示すように 8 つの頂点を持ち、各辺の長さは図形の中心から最大サイズ（円の半径 (40px)）までランダムに設定し、各頂点を繋ぎ合わせた図形を用いている。この図形を画像の様々な箇所にランダムで 50ms の間隔で 100 個の図形の発生と消滅を繰り返すようにした。図形を 100 個表示している画像を図 7 に示す。HSL 色空間における値は透明度以外は単色色妨害フィルターの設定と同じである。これら 2 つの妨害色フィルターを用いることによって、平均画素値をかく乱することを目的としている。

3.5 解答の照合方法

解答の照合には、オリジナル画像、出題画像、11 色の中からユーザの解答した色の RGB 値、および解答エリアの座標を利用する。

ユーザの解答した画素値 $Ca = (Ra、Ga、Ba)$、解答エリアの左上の座標（x,y）が与えられると、システムはオリジナル画像の (x,y)、$(x + Sarea、y)$、$(x,y + Sarea)$、$(x + Sarea、y + Sarea)$ 内のユーザの解答領域を走査しオリジナル画像の画素値 $Co = (Ro、Go、Bo)$ を求める。画素値の算出には解答領域内の中央値を利用した。

オリジナル画像のユーザ解答領域の画素値を取得すると、カラーパレットの 11 色それぞれとの色差を計算し、色差の値が最小値となるカラーパレットの色が正解となり、この色とユーザが選んだ色が一致すると人間と判別するシステムとなっている。本論文では、画素値の色差を CIEDE2000[8] を用いて計算した。

4. 実験・評価

4.1 実験の目的

本論文の実験では提案 CAPTCHA に上限解答回数を設け、上限を 10 回とした。解答回数が 10 回を超えてしまうと CAPTCHA としてかなりの時間がかかってしまうため、本

論文の実験においては、この 10 回の内に提案 CAPTCHA のクリア条件を満たした被験者を CAPTCHA 成功者としてみなした。

実験の目的は本論文で検討した CAPTCHA の正答率、被験者の解答領域における内部ボットプログラムの正答率、提案 CAPTCHA を成功した人の平均回答回数、CAPTCHA 成功率を調べることにより、提案 CAPTCHA の有用性を調査することである。また、SUS(System Usability Scale) 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 CAPTCHA におけるユーザビリティの調査を行った。

4.2 実験方法

実験には宮崎大学工学部生 15 人が参加した。実験に用いる画像は 300 × 300px の動物、食べ物、国旗、キャラクターの画像各 10 枚計 40 枚を用意した。このうち動物、食べ物の画像は現実世界で撮影されたものを用いている。国旗、キャラクターの画像はコンピュータで作成されたイラストを用いている。このように写真とイラストの 2 種類の画像に分けて実験した。写真とイラストとでは照明条件が大きく違う。現実世界で撮影された写真には撮影時の照明条件が画像の画素値の色成分に大きく影響を与えている。例えば、影で撮影された写真と太陽の日が差す真下で撮影された写真の画素値の色成分は違う。また、写真とイラストでは 1 枚の画像に含まれる色の種類 (各画素の RGB 値) の数が多くなる。これらの理由から画像を写真とイラストの 2 種類に分けた。被験者には、写真とイラストにおいてそれぞれ 20 枚の画像からランダムで 1 枚ずつ表示するようにした。また、出題画像の解答領域は被験者が正解できる自信のある個所を選んでもらった。

4.3 結果と考察

被験者 15 人に提案 CAPTCHA を解いたときに得られた実験データである被験者の平均正答率、ボットの平均正答率、平均解答回数 (成功した人のみ)、CAPTCHA 成功率の値を、写真画像を用いた実験結果は表 1 に、イラスト画像を用いた実験結果は表 2 にそれぞれ示す。表 1、表 2 より、透明度が 70%の写真画像に比べ、イラスト画像の透明度は写真画像に比べて 20%低い 50%という数値に設定したのにも関わらず、被験者の平均正答率や CAPTCHA 成功率は写真画像よりイラスト画像の方が高い値となった。このような結果になった原因を考察すると、オリジナル写真画像は写真が撮られた照明環境により、本来の色とは違っている可能性がある。したがって、被験者はオリジナル写真画像においても色の恒常性が働いていた可能性がある。しかし、提案 CAPTCHA の正解の色はオリジナル画像の色であるため、正解の色を選ぶのに妨げになっていた可能性がある。

写真画像、イラスト画像ともに、上限を 10 回としたときに CAPTCHA 成功率は、textCAPTCHA の平均正答率 92%には及ばなかったので今後の課題であるといえる。

CAPTCHA を成功した時の連続正解数は、写真画像、イラスト画像ともに、9 割近くが 3 回連続正解の時であった。この結果より平均偶然認証突破率は 1/4092 より低い数値となってしまった。この結果の原因は、提案 CAPTCHA の一問ごとの正答率が写真画像、イラスト画像ともに 80%を下回っているのでそもそも 4 回以上正解することが困難になっているためである。

表 1. 実験結果 (写真)

被験者の平均正答率 [%]	71.58(68/95)
ボットの平均正答率 [%]	21.05(20/95)
平均回答回数 (成功した人のみ) [回]	5.000
CAPTCHA 成功率 [%]	73.33(11/15)

表 2. 実験結果 (イラスト)

被験者の平均正答率 [%]	79.49(62/78)
ボットの平均正答率 [%]	14.10(11/78)
平均回答回数 (成功した人のみ) [回]	4.462
CAPTCHA 成功率 [%]	86.67(13/15)

4.4 ユーザビリティ評価

本論文の提案 CAPTCHA の平均 SUS スコアは 83.17 であった。[9] によると SUS の平均スコアは 68 とされている。さらには、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。

5. まとめと今後の課題

本論文では人間の「色の恒常性」と呼ばれる高度な認識能力を利用した colorCAPTCHA のコンセプト提案・実装・評価を行った。人間のより高度な認知能力を利用している提案 CAPTCHA は人間の平均正答率には課題を残したが、偶然認証突破確率などのボット耐性における信頼性は高いと考えられる。また、ユーザビリティの評価の指標となる SUS スコアも高い数値が得られ、提案方式の有用性が示された。今後の課題は一問ごとの正答率を高めることである。そのためには、色妨害フィルターの透明度の最適化、画像の選別、画像の色成分によって色妨害の色を画像ごとに変えることなどが必要になってくる。

参考文献

1) von Ahn, L., Blum, M., and Langford, J: Telling Humans and Computers Apart Automatically, Communications of the ACM., Vol.47, No.2, pp.57-60, 2004.

2) Vu Duc Nguyen, Yang-Wai Chow, Willy Susilo: On the security of text-based 3D CAPTCHAs, Computers & Security, Vol.45, pp.84-99, 2014.

3) https://support.google.com/recaptcha, (accessed 2018/01/28).

4) S. Sivakorn, I. Polakis and A. D. Keromytis: I am Robot: (Deep) Learning to Break Semantic Image CAPTCHAs, 2016 IEEE European Symposium on Security and Privacy (EuroS&P), Saarbrucken, pp.388-403, 2016.

5) Mandeep Kumar, Pathankot Renu Dhir: Design and Comparison of Advanced Color based ImageCAPTCHAs, International Journal of Computer Applications(0975 - 8887), Vol.61-No.15, 2013.

6) Jeremy Elson, John (JD) Douceur, Jon Howell, Jared Saul: Asirra: A CAPTCHA that Exploits Interest-Aligned Man-

ual Image Categorization, Association for Computing Machinery, 2007.

7) 内川　恵二: 色の恒常性と認識, 映像情報メディア学会誌, Vol.58, No.5, pp.662-668, 2004.

8) Cui, G.and Rigg, B: The development of the CIE 2000 colour–difference formula: CIEDE2000, Color Research & Application, Vol.26, pp.340-350, 2001.

9) Jeff Sauro: MEASURING USABILITY WITH THE SYSTEM USABILITY SCALE (SUS), Measuring U, https://measuringu.com/sus/, (accessed 2018/01/28).