Color CAPTCHA のための色妨害画像に対する色恒常性の成立度合いの検討

臼崎　翔太郎 a)・中原　裕介 b)・山場　久昭 c)・油田　健太郎 d)・岡崎　直宣 e)

Investigation of the Color Constancy of Color Obstruction Images for CAPTCHA

Shotaro USUZAKI, Yusuke NAKAHARA, Hisaaki YAMABA,
Kentaro ABURADA, Naonobu OKAZAKI

Abstract

Currently, the text-based CAPTCHA which makes the user answer the original characters from the distorted those is standardly used. In addition, reCAPTCHA has recently used as image-based CAPTCHA which is easy for humans to understand. However, these standard CAPTCHAs are broken with high success rate because of the progressing of machine learning and image processing technology. Although obstacles that make it impossible to read the machine are necessary, it causes the owing to the success rate of the human decrease. Therefore, the requirement for CAPTCHA is not only more to be difficult for the machine but also to be easy for humans. On the other hand, Color-based CAPTCHA has a high human success rate, but it is likely to be solved mechanically. This CAPTCHA does not add disturbance unlike the conventional ones can be easily solved by humans, however, it has a vulnerable in terms of security. In this paper, we propose a color CAPTCHA that uses color obstruction images has higher tolerance to machines. Due to the influence of color constancy, it is considered that high human success rate can be guaranteed even if there is color obstruction. We evaluated how much color constancy is functioned for color obstruction images, and confirmed that color recognition almost matches as in normal state.

Keywords: CAPTCHA, color constancy, gray world assumption

1. はじめに

今ではコンピュータは教育や販売、仕事、コミュニケーションなどあらゆる場面で利用され便利な時代になっている。ところが、便利に使っている反面、ボットウイルスに感染した機器による不正行為などが多発している。そこで、導入されたシステムが CATPCHA である。現在、導入されている CAPTCHA の中で最もスタンダードなものが text CAPTCHA であり、ゆがんだ文字列を読み取って入力する CAPTCHA である。開発された当時のコンピュータは文字を歪ませると読み取ることができなかったので、解答は困難だったが、最近のコンピュータではこれらも読み取られてしまう事例も発生しており、安全性と信頼性が疑問視されている。そこで、機械にとって text CAPTCHA より困難な CATPCHA を開発することが求められるが、それだけでなく人間にとってより容易に解答できることも重要視される。color CAPTCHA は従来の CAPTCHA とは異なり、妨害を加える必要がないので、人間にとって容易に解くことが可能になる。その反面、セキュリティ面において非常に弱いところが欠点である。

そこで、本論文では、よりセキュリティ面の高い color CAPTCHA を検討する。色恒常性を利用することで、色妨害を加えた画像を使った color CAPTCHA においても人間の認識率を保証することができる。実験においては色の妨害を加えた画像についてどれだけその色恒常性が成立するかを確認した。具体的にはグラデーション型、シェイプ型の色妨害を加えた画像を 1 枚ずつ表示し、実験者が指した部分の色を基本色の 11 色から選択させ、色妨害を加えていない画像を見た人と色妨害を加えた画像を見た人の選択色を比較して、κ 係数を使って色恒常性の成立度合いを調べた。

2. 関連研究

2.1 EZ-Gimpy CAPTCHA

EZ-Gimpy CAPTCHA とは text CAPTCHA の一種で、辞書に載っているものの中から 1 つ選んだ単語について、妨害を加え歪ませた文字を表示画像として提示し、ユーザに何が書かれているかを入力させるシステム (図 1) のことである。妨害が加えられた文字について一字ごとに分割して認識することは機械には困難であったため、人間と機械を区別することが可能であった。しかし、文献 [1] によれば、EZ-Gimpy CAPTCHA

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

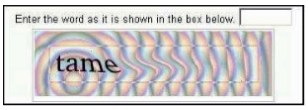

図 1. EZ-gimpy CAPTCHA

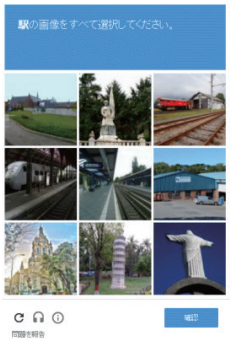

図 2. 画像型 CAPTCHA

は 97 %以上の精度で読み取られてしまうため、CAPTCHA のセキュリティが高いとは言えない。

2.2 Google reCAPTCHA

text CAPTCHA では機械に読み取られてしまうため、様々な CAPTCHA が開発されている。その中の一つが画像型 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、画像を選択させるだけで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、画面下部に出題された 9 つの画像から上の画像か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 text CAPTCHA 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、text CAPTCHA よりも人間にとって非常に容易であるが、機械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機械を区別しようとしている。しかし、文献 [2] によれば、Facebook の image CAPTCHA は機械に 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 color CAPTCHA

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 color CAPTCHA を提案した [3]。

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 color CAPTCHA を解かせているが、その結果、color CAPTCHA は従来の text CAPTCHA や image CAPTCHA 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、機械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 text CAPTCHA や image CAPTCHA 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、機械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、将来的に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 color CAPTCHA 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 color CAPTCHA を検討する。

3. 提案手法

3.1 提案 CAPTCHA

提案する CAPTCHA は、色妨害を使った画像から、解答エリアに示された領域の色を答えさせる CAPTCHA である。図 4, 図 5 のように、画面上部には色妨害がされてある画像と解答エリアである四角の枠、画面下部には色の選択バーが表示されている。解答方法は、まずその画像の範囲内でランダムに四角の枠が一つ表示される。次に、その指示されてある四角の枠内の色の色について、ユーザに色の選択バーで自分が一番近いと思う色を選んでもらい、送信することで CAPTCHA を解くことができる。

この提案 CAPTCHA の利点として、色妨害が含まれているため、従来の color CAPTCHA よりもセキュリティが強くなることがあげられる。また、画像に色妨害フィルターをかぶせても、人間は色の恒常性が働くことにより、元の画像の色を保持しようとするため、人間における正解率が従来と同程度を保つ可能性が高いことが考えられる。

3.2 色の恒常性

色の恒常性とは周りの環境の照明光の色彩や明度がどんなに変化しても表面の色が変わらないように見せる色覚特性のことを指す。

人間が色を認識するメカニズムを説明すると、光源からの照明光が物体の表面に反射して反射光となり、観察者の眼に入射する。その反射光は眼球内を通り、網膜の 3 錐体により 3 種の応答に変換され脳へと送られる (図 3)。脳がこの応答を認識することで、人間は今見ている色を感じることができる。人間の眼に到達する反射光には、物体の色を表現する表面反射率成分と、照明光成分が混じり合っている (図 3)。 普通の環境内では、照明光の強度や分光特性、表面反射光の輝度がさまざまに変化していくので、反射光に含まれる表面反射率成分と照明光成分は一般的に未知である。よって数学的にこれらを分離することは不可能であるが、人間は視覚系が表面反射率成分と照明光を分離し、物体の色を不変に見せている。

3.3 灰色仮説

灰色仮説とは、色の恒常性を説明するための 1 つの仮説で、物体表面の反射光を平均すると灰色になることを指す。これによって物体反射率成分を無視することができ、照明光成分のみを抽出することができる。色の恒常性を、こうして抽出された照明光成分を除去することによって色を不変に見せている現象と説明するのが、この灰色仮説である。この仮説は、照明光に偏りのある画像を補正するホワイトバランスの手法としてビデオカメラなどの光源色の調整に用いられていることが多い [4]。例えば文献 [5] では、反対色の関係になるように各色相カテゴリから色を選択する手法を提案し、灰色仮説成

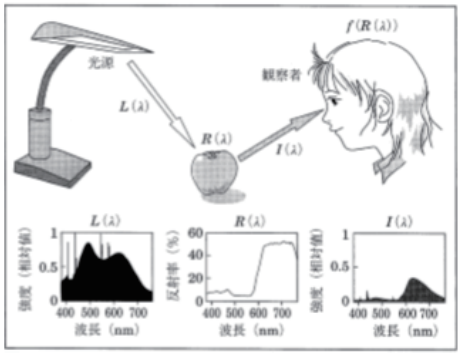

図 3. 色の恒常性 (文献 [4] 参照)

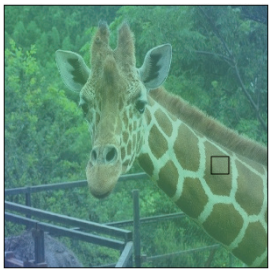

図 4. グラデーション型

図 5. シェイプ型

立可否を判定することで照明光推定を行い、画像の色を補正している。

青い空や緑の草原などのようなほぼ単色のみで構成された風景などでは、灰色仮説が成立しないが、通常の風景においてこの仮説はほとんど成立していることが知られている。つまり、ほとんどの画像は灰色仮説を利用した画像補正手法で照明光の影響をキャンセルすることができるため、機械的に CAPTCHA が解答されないためには、灰色仮説にそぐわない色妨害を行う必要がある。

3.4 色妨害の検討

灰色仮説の説明文にもあったように、ほとんどの画像はこの灰色仮説を利用して、色恒常性を再現することができる。そこで、その色補正手法に耐性を持たせるために妨害を加える必要がある。本研究では、灰色仮説に耐性を持たせるために、画像中に複数色のフィルタを加える。これにより、灰色仮説においての色補正に必要な画素値の平均値が歪められるので、色補正がより困難になり、照明光成分を除去することが難しいと考えられる。

複数色のフィルタの作成にあたって、次の 2 つのタイプを検討した。1 つ目のグラデーション型 (図 4) は複数色の色が混ざり合って構成されており、複数色が混ざることで画素値が単色フィルタのときよりも乱れやすいため、色補正が困難になる。2 つ目のシェイプ型 (図 5) は、複数の色付きの図形で構成されており、それらがランダムに配置されることで画素値に乱れが生じ、色の補正がより困難になる。

4. 評価実験

画像に色の妨害を加えているが、CAPTCHA として成立するためには、色妨害画像に対しても色の恒常性が働く必要がある。よって、本研究で検討した非現実的な妨害フィルタを加えても色の恒常性が働くかを確かめるのが本実験の目的である。もし、色の恒常性が働くのであれば、どのような色妨害に対しても色妨害のない状況と同じ判断ができると考えられる。したがって評価実験では、被験者を色妨害のある画像を見るグループと、色妨害のない画像を見るグループに分け、グループ間で色の判断がどのくらい一致しているかを確認する。判断の一致度を測る指標として、本研究では κ 係数を利用することにした。

また、実験を行う際、出題画像[f]の候補として、画像を以下の 4 種類に大別した。

1. 物体と色の対応関係が既知である現実の画像
2. 物体と色の対応関係が既知である幾何学的な画像
3. 物体と色の対応関係が未知である現実の画像
4. 物体と色の対応関係が未知である幾何学的な画像

物体と色の対応関係が未知であるものは解答にばらつきが生じてしまう可能性があるため、CAPTCHA に向かないと考えた。よって、CAPTCHA として有利となると思われる、1. と 2. にあてはまる画像で実験を行った。

4.1 実験方法

物体と色の対応関係が既知である現実の画像のうち、動物、食べ物、国旗、キャラクターの画像を各 10 枚計 40 枚用意した。被験者は工学部の 20 代男女計 15 人で、オリジナル画像、グラデーション型の画像、シェイプ型の画像の 3 グループに分けた。被験者には 1 枚ずつ画像を見せ、実験者が指定したエリアについて、基本色 11 色から一番近いと思う色を被験者に答えさせた。指定したエリアについては、1 枚の画像につき 2 か所から 5 か所で決定した。これを 40 枚すべての画像に対して行い、解答エリアは合計で 140 か所となった。オリジナル画像を見たグループとグラデーション型の画像を見たグループ、オリジナル画像を見たグループとシェイプ型の画像を見たグループの間で、選択した色を比較し κ 係数を求める。これを求めることによって、グループ間の判定一致率を測ることができるからである。

4.2 Cohen の κ 係数

2 者間の符号化、評定などがどれだけ一致しているかを示す指標である。これは、名義尺度や順序尺度の問題に利用できる。計算式を (5) 式に示す。

$$P_0 = \frac{\text{解答一致数}}{\text{全評定数}} \tag{1}$$

$$= \frac{\text{分割表の対角成分の和 (トレース)}}{\text{分割表の全成分の和}} \tag{2}$$

$$P_e = \text{各解答カテゴリの期待一致率} \tag{3}$$

$$= \text{分割表の対角成分毎の周辺確率の積の和} \tag{4}$$

とすると、κ 係数は

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \tag{5}$$

[f] ただし、画像はそれぞれ 3 色以上使っていてかつ、色のエリアがある程度大きいことを条件とする

表 1. : Landis and Koch (1977) による κ 係数の目安 [6)]

κ 係数の範囲	一致率
0.0〜0.2	わずかに一致
0.21〜0.40	まずまずの一致
0.41〜0.60	中等度の一致
0.61〜0.80	かなりの一致
0.81〜1.0	ほぼ完全 or 完全一致

となる。

表 1 のように κ 係数が 0.2 以下の場合、評者間の一致率は非常に低く、0.81 以上の場合は一致率は十分に高いと判定できる。ただし、κ 係数を求めるには 1 者につき 100 個以上の値が必要であり、今回の実験では全部で 40 枚の画像しかないので、1 枚の画像につき平均 2.5 か所以上指し示す必要があった。

4.3 カテゴリカル基本色

色の基本的なカテゴリーとして、黒、白、赤、緑、黄、青、茶、紫、ピンク、オレンジ、灰の 11 色が提唱されている。これらを基本色名 (Basic Color Terms) と呼ぶ。基本色名の定義 4 つを以下に示す。

1. 全ての人の語彙に含まれること
2. 人によらず、使用時によらず、安定して用いられるようにすること
3. その語彙が他の単語に含まれないこと
4. 特定の対象物にしか使われない色でないこと

本研究では、「すべての人の語彙に含まれる」という性質から、この基本色を色の判定実験で用いることにした。基本色のうち、どの色に見えるかを答えさせることで、疑似的に色の判定問題を名義尺度の問題としている。

5. 結果・考察

実験の結果、表 2 に示す通り、色妨害を加えていない画像を見たグループとグラデーション型の色妨害を加えた画像を見たグループとの平均 κ 係数は 0.875、色妨害を加えていない画像を見たグループとシェイプ型の色妨害を加えた画像を見たグループとの平均 κ 係数は 0.865 となった。それぞれ「両者の判定がほとんど一致している」という目安である 0.81 を超えていることが確認できたため、色妨害を加えても選択した色の一致率が高く、妨害を加えていない color CAPTCHA と同程度に人間の認識率が高いことが考えられる。出題画像に対して色妨害を加えていることを考えると、人間の認識率を損ねることなく、従来の color CAPTCHA よりもセキュリティを高められる可能性が示せた。

今回の実験において被験者に解答させた部分の大半はそれらの色が基本色 11 色の中に含まれていたが、中には青と緑が混ざった色や、ピンクとオレンジが混ざった色、肌色など感覚的に基本色 11 色にはない色も存在し、そのような場合では妨害の加えられていないグループの被験者同士でも解答が分かれた。このとから、解答エリアを決定する際、曖昧な色の領域を選ばないようにすることが必要となる。

表 2. 色妨害の無いグループとの間の平均 κ 係数

画像タイプ	グラデーション型	シェイプ型
平均 κ 係数	0.875	0.865

また、妨害フィルタのグループにおいて、例えば本来はナスの色である紫か黒を選ぶと予想される箇所において、別の野菜と間違えて緑と答えるといったことも確認された。以上のことから、映っている物体を認識することも、色の恒常性に影響があると考えられる。

6. まとめ

本研究では、人間の持っている色恒常性に注目し、色妨害を加えた出題画像の色をユーザに答えさせることで, 従来と同程度の人間正解率を保ちつつ、従来手法よりもセキュリティ耐性の高い color CAPTCHA を検討した。評価実験では、非現実的な色妨害画像にも CAPTCHA に必要な程度の色恒常性が働くかどうか調査した。具体的には、色妨害がない画像を見た人と 2 種類の色妨害を加えた画像を見た人をそれぞれ 5 人ずつの 3 グループに分け、1 ジャンル 10 枚で合計 40 枚の画像を 1 枚ずつ見せ、実験者が指した部分の色を基本色 11 色の中から選択させ、色妨害なしの画像を見た人と色妨害ありの画像を見た人の間で選択した色を比較してから κ 係数を求めた。

その結果、κ 係数が 0.81 を超えていたことから、色妨害がない状態とある状態での色の判定がほとんど一致していることが分かった。よって、人間における色の恒常性が働いたと考えられるため、このシステムにおいて人間の認識率が従来と同程度を保ち、従来にはなかった色妨害が加えられていることにより、セキュリティ面での耐性も高いことが考えられる。今後の課題として、紛らわしい色や妨害ありの画像を見た人の物体の勘違いなどを考慮する必要がある。

参考文献

1) A. Bansal, D. Garg, and A. Gupta: Breaking a Visual CAPTCHA : A Novel Approach using HMM ⟨https://pdfs.semanticscholar.org/3c2c/9af1e9a3b7095edaf8f205dfbadc30f917fb.pdf⟩, 2008 (accessed 2018-03-04).
2) S. Sivakorn, I. Polakis, and A. D. Keromytis: I Am Robot: (Deep) Learning to Break Semantic Image Captchas, in *IEEE European Symposium on Security and Privacy (EuroS & P)*, pp. 388–403, 2016.
3) M. Kumar, and R. Dhir: Design and Comparison of Advanced Color based Image CAPTCHAs, *International Journal of Computer Applications(0975 - 8887)*, Vol. 61, No.15, pp. 24–29, 2013.
4) 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.
5) 川村春美、米村俊一、大谷淳、松浦宣彦: 色相に着目した灰色仮説に基づく照明光推定法の一提案, 研究報告オーディオビジュアル複合情報処理（AVM）, No. 6, pp. 1–6, 2010.
6) J.R. Landis, and G.G. Koch: The Measurement of Observer Agreement for Categorical Data, *Biometrics*, Vol. 33, No. 1, pp. 159–174, 1977.