色恒常性を利用したカラーCAPTCHAの検討

臼崎 翔太郎[1] 油田 健太郎[1] 山場 久昭[1] 椋木 雅之[1] 朴 美娘[2] 岡崎 直宣[1]

概要：CAPTCHA(Completely Automated Public Turing test to tell Computers and Humans Apart) において，機械耐性のために妨害を加えても人間の正答率が保証されることが重要な要件となる．我々は色恒常性の「機械的に再現が困難である」点と「物体表面色が変化しても本来の色を認識できる」点に着目し，人間に影響を与えない妨害が可能な CAPTCHA を検討した．評価実験では，CAPTCHA として非現実的な色妨害が施されていても色恒常性が働くかを検証した．具体的には，色妨害の加えられた画像を見たグループと，色妨害の加えられていない画像を見たグループ間で κ 係数を用いて判定の一致度を検証した．実験の結果，κ 係数は 0.81 以上となり，高い一致度であることが分かった．この結果から，提案手法は人間の認識能力に影響を与えない妨害が可能な CAPTCHA の形態である可能性が示された．

Investigation of Color-based CAPTCHA Using Color Constancy

SHOTARO USUZAKI[1] KENTARO ABURADA[1] HISAAKI YAMABA[1] MASAYUKI MUKUNOKI[1]
MIRANG PARK[2] NAONOBU OKAZAKI[1]

1. 研究背景

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) は機械によるアカウントの不正取得や，オンライン選挙などでの不正な得票操作，ブログなどでのスパム行為を防ぐ目的のため Web 上で広く利用されている認証技術である [1]．その基本原理は，人間には容易で，機械には困難なタスクを課すことによって，人間と機械を区別するというものである．

CAPTCHA は大きく 3 つのタイプに分類することができる．歪んだ文字列を読み取らせて元の文字列を答えさせるテキスト型 CAPTHCA，人間の知覚特性や心理特性を利用して，提示された画像の中から特定の画像を選ばせる画像型 CAPTCHA，ノイズなどの妨害が加えられた音声から答えとなる文字列を聞き取らせて答えさせる音声型 CAPTCHA である．このうち，現在 CAPTCHA として広く利用されているのが，テキスト型 CAPTCHA と，画像型 CAPTCHA である．しかしながら，これらの CAPTCHA は画像処理や機械学習技術の発展によって高い精度で機械に解決できることが報告されている [2] [3] [4]．この問題を解決するために，さらなる妨害を加えて機械への耐性を向上させる必要があるが，その妨害によって人間の正答率が低下することが知られている [5] [6]．したがって CAPTCHA として成立させるためには，機械耐性のために加えられる妨害が人間の認識能力を損なわない形で提供されている必要がある．

そこで本論文では、人間が持つ色覚特性である色恒常性の，「周囲の照明光の影響を受けても物体本来の色を知覚できる」，「機械的に再現が困難である」という点に着目して，人間の認識能力を損なわない妨害が可能な CAPTCHA の形態を検討した．

実験では色妨害を加えた出題画像について CAPTCHA として成立する程度に色恒常性が働くか調査した．具体的には色妨害を加えた出題画像を見たグループと加えていない出題画像を見たグループ間の色の判定一致度を κ 係数を利用して算出した．実験の結果，我々が検討した妨害フィルタはどちらも κ 係数が 0.81 を超えていた．判定の一致度が極めて高いことから，CAPTCHA として必要な程度色の恒常性が働いていることが分かり，提案手法が人間に影響を与えない妨害が可能な CAPTCHA の形態である可能性が示された．

[1] 宮崎大学
[2] 神奈川工科大学

2. 関連研究

テキスト型 CAPTCHA は，歪んでいたり妨害が加えられたりしている文字列の画像をユーザに提示して，その文字列を入力させる CAPTCHA である．

EZ-Gimpy CAPTCHA は，辞書に載っているものの中から選んだ一つの単語を歪曲させて出題画像として提示し，ユーザに何が書かれているかを入力させるシステム (図 1A) である．この CAPTCHA は，Yahoo 社に利用されていた．しかしながら，OCR 技術の発展によって高い精度で破られることが報告されている．例えば文献 [2] では，EZ-Gimpy CAPTCHA は 97 %以上の精度で機械に破られており，CAPTCHA としての性能が高いとは言えない．

reCAPTCHA[7] は，クラウドソーシング技術としても利用されており，OCR 技術で読み取ることのできなかった文字列を CAPTCHA として出題する (図 1B)．この文字列だけでは機械的に解答を照合することができないため，出題としてもう一つ答えがすでに分かっている文字列を提示する．一方の出題画像によって OCR 技術の発展に役立て，もう一方の出題画像によって CAPTCHA としての要件を満足するようにしている．reCAPTCHA は文献 [3] などのように高い精度で解答が可能であることが指摘されており，Google 社では，現在はこのタイプの CAPTCHA の利用は非推奨となっている [8].

一方画像型 CAPTCHA は，人間の高度な認知能力を利用して人間と機械の区別を図るシステムである．

有名な例が，Google reCAPTCHA v2 (図 1C) である．Google reCAPTCHA v2 は，複数枚の画像か，ある 1 つの画像を複数枚に分割した画像を提示し，その画像群の中から，特定の物体が映っている画像を全て選ばせる CAPTCHA である．この手法は，人間の高い物体認識能力を利用することによって人間と機械の区別を行っている．このシステムはテキスト型 CAPTCHA のように妨害が加えられておらず，またキーボード入力ではなくクリックあるいはタッチによって入力を行うため，人間にとってより容易である．また，一度 CAPTCHA を解くと，今後はチェックボックスをクリックするだけで CAPTCHA 処理をスキップすることができる．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 [4] によれば機械に 83.5%の精度で解かれてしまうことが報告されている．この手法には妨害は含まれていないので，機械耐性を向上させるには妨害が必要となる．

Asirra[9] は，犬と猫の画像から成る 12 枚の出題画像の中から，猫の画像のみを選択させる CAPTCHA である (図 1D)．ペットの里親を探すサービスである Petfinder と提携しており，保護された動物の膨大な量の出題パターンを生成でき，提案時点では 1 日約 10,000 枚の画像が追加されている．CAPTCHA には里親サービスとしての機能もあり，CAPTCHA ユーザは利用された出題画像の中から気に入った犬猫の画像を選択し，里親候補になることができる．人間の正解率は 30 秒以内の解答時間で 99.6%の性能であり，高い人間の正答率が報告されている．しかしながら，機械学習による物体認識技術の進歩により，文献 [10] では 82.7%と高い正答率で機械に回答されている．こちらも，Google reCAPTCHA v2 と同じように妨害が含まれていないため，妨害の検討が必要である．

DeepCAPTCHA[6] は，人間の，物体の相対的なサイズと深さの知覚を利用した CAPTCHA で，6 枚並んだ画像を知覚できるサイズの順番に並べさせるシステムである (図 1E)．人間の成功率は最大 92.32%となっている．人間の感覚を利用することから，CAPTCHA の解答を自動的に生成するのが困難であり，クラウドソーシングを利用して生成している．妨害を複数考案しているが，それによって人間の正答率が 83.7%に低下しているため，CAPTCHA としての性能を落とさない妨害を検討する必要がある．

提案手法に似たアプローチの手法として、Kumar らのカラー型 CAPTCHA 手法 [11] がある (図 1F)．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，あるいは指定されたオブジェクトの色をユーザに答えさせる．評価実験では，職種を問わず 5 歳以上の 1,000 人に CAPTCHA を解かせているが，平均成功率はほとんど 100%と高い数値となっている．Kumar らの手法では，機械が画素値から色の名前を認識できないことを前提に提案されているため，これまでの文字列列 CAPTCHA や画像 CAPTCHA のような妨害が必要なく，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十分に耐性があるとはいいがたい．

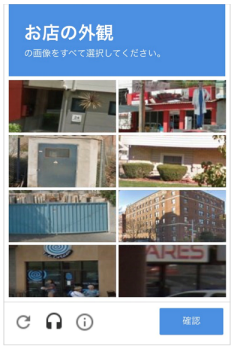
上記の関連研究から，機械耐性のためには妨害を検討しなければならないが，それによって人間の認識能力を低下させない妨害の形態を考える必要がある．

3. 色恒常性

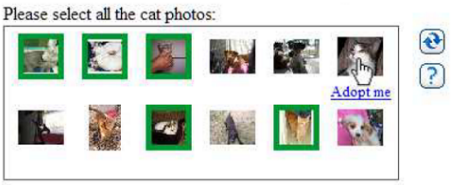
本章では，CAPTCHA に利用する色恒常性を説明する．眼球内に入射波が到達すると，網膜上の 3 錐体が光を 3 種の電気信号に変換し，その信号が脳に到達することによって，人間は今見ている色を認識する [12]．ここで，物体表面が理想拡散面 (lambertian surface) であると仮定し，入射光成分を $\boldsymbol{f}(\boldsymbol{x})$，$\lambda$ を波長，$\boldsymbol{x}$ を空間座標，$e(\lambda)$ を照明光成分，$s(\lambda)$ を物体反射率成分，$\boldsymbol{c}(\lambda)$ を分光反射特性，ω を可視光領域とすると，色の見えを式 (1) で定式化できる [13]．すなわち，物体の色を表現する物体反射率成分，照

A: EZ-Gimpy[2]

B: reCAPTCHA [7]



C: Google reCAPTCHA v2 [8]



D: Asirra [9]

E: Deep CAPTCHA [6]

F: Color CAPTCHA [11]

図 1 既存手法の CAPTCHA の例 (A: EZ-Gimpy, B: reCAPTCHA, C: Google reCAPTCHA v2, D: Asirra, E: Deep CAPTCHA, F: Color CAPTCHA)

明光を表現する照明光成分が混合された光を観測して，人間は色を認識している．

$$\boldsymbol{f}(\boldsymbol{x}) = \int_{\omega} e(\lambda)s(\lambda,\boldsymbol{x})\boldsymbol{c}(\lambda)d\lambda \qquad (1)$$

色恒常性は，周りの環境の照明光が変化しても物体表面の色が変わらないように見せる色覚特性のことである [12]．式 (1) を踏まえると，色恒常性は，入射光 $\boldsymbol{f}(\boldsymbol{x})$ から物体反射率成分 $s(\lambda)$ を抽出して変換することによって，物体表面の色を不変に見せるメカニズムと説明できる．しかしながら，既知な値は入射光 $\boldsymbol{f}(\boldsymbol{x})$ と生理学的に決定されている $\boldsymbol{c}(\lambda)$ のみであり，$e(\lambda)$，$s(\lambda)$ を一意に分離することは数学的には不可能である．したがって，機械的に色恒常性を再現するためには，照明光あるいは物体表面光について仮説を置いて，拘束式を 1 つ増やす必要がある．しかしながら，一般に知られている仮説は全ての画像に対して有効ではないため，色恒常性を再現する手法が今も研究されている [14] [15]．

色恒常性に関わる仮説の 1 つである Gray-World 仮説は，物体表面の反射光を平均すると灰色になるというものである．Gray-World 仮説を説明したものを式 (2) に示す [13]．画像処理の観点から式 (2) を解釈すると，この仮説に基づく画像であれば，全画素の RGB 値の算術平均を求めることによって物体反射率成分を無視して照明光成分のみを抽出することができる（式 (3) にプロセスを示す）．海や山の風景といった元々色の偏りが大きく画素値の平均値が灰色にならない画像を除けば，Gray-World 仮説はほとんど成立することが知られている [12]．

$$\frac{\int s(\lambda,\boldsymbol{x})d\boldsymbol{x}}{\int d\boldsymbol{x}} = k \qquad (2)$$

$$\begin{aligned}\frac{\int \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x}}{\int d\boldsymbol{x}} &= \frac{1}{\int d\boldsymbol{x}} \int\int_{\omega} e(\lambda)s(\lambda,\boldsymbol{x})\boldsymbol{c}(\lambda)d\lambda d\boldsymbol{x} \\ &= k\int_{\omega} e(\lambda)\boldsymbol{c}(\lambda)d\lambda \qquad (3)\end{aligned}$$

4. 提案手法

我々は色恒常性の「周囲の照明光の影響を受けても本来の色を知覚できる」,「数学的に再現が困難である」という特性に注目し，CAPTCHA に利用することにした．

提案手法では，出題画像上の解答エリアと呼ばれる領域の色と近い色を，カラースライダーを用いてユーザに解答させる．解答エリアはランダムな位置に表示されるが，ユーザは解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．提案手法の出題例を図 2 に示す．

色妨害が施されていても，色恒常性の効果によって人間の認識能力を大きく損なうことなく元来の色を認識することが期待できるが，単純に単色のフィルタを重ねるだけでは色恒常性アルゴリズムに対する耐性が十分確保されないことを予備実験で確認した．したがって妨害にも工夫が必要となる．そこで我々は，CAPTCHA に必要な色妨害として，一般的な色恒常性アルゴリズムの 1 つである Gray-World 手法に耐性を持たせることを設計目標とした．妨害としてグラデーションのようなフィルタを重ねる方式（グラデーション型）と，透明な色フィルタを多数配置する方式 (シェイプ型) を取った．また，人間の視認性を考慮して動画型で妨害を構成している．いずれも複数色を出題画像に加え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．この妨害は，他の色恒常性アルゴリズムに対しても有効であると考える．なぜなら本来色恒常性アルゴリズムの対象はある照明光の下で撮影された現実世界の画像であり，このような非現実的な妨害に対して各色恒常性アルゴリズムの仮説は成り立たないと考えられるためである．一方で，非現実的な画像に対しても色恒常性が働くかどうかは明らかになっていないため，非現実なシチュエーションによる色恒常性への影響を検証する必要がある．

4.1 提案手法のパラメータ

共通として持つパラメータは出題画像のサイズ S_{task}，解答エリアのサイズ S_{area} がある (図 3 左)．妨害のパラメータはシェイプ型とグラデーション型でそれぞれ固有なものを設定する．

グラデーション型では，グラデーションの向き Ang，グラデーションの層数 L_{num}，i 番目の層の中心色 C_{G_i}，フレームレート F_{g} をパラメータとして持つ（図 3 中央）．

シェイプ型では，図形の個数 N，図形の辺の数 P_{num}，図形の最大サイズ S_{max}，i 番目の図形の色 $C_{S_{\mathrm{i}}}$ フレームレート F_{s} をパラメータとして持つ（図 3 右）．このフィルタは同じく色妨害を行っている文献 [16] を参考にしている．

C_{G_i}，$C_{S_{\mathrm{i}}}$ は調整のしやすさを考慮して HSV（Hue, Saturation, Value）空間で決定することにした．HSV 空間は色の種類そのものを表現する色相と，色の明るさを表現する明度，色の鮮やかさを表現する彩度で構成される色空間である．色相の範囲は 0 ～ 360°，明度と彩度の範囲は 0 ～ 100% である．

図 2　提案 CAPTCHA の出題例 (左側がグラデーションフィルタ，右側がシェイプフィルタ)

4.2 解答の照合方法

解答の照合には，オリジナル画像と，ユーザの解答した色の画素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 $\boldsymbol{c_{\mathrm{a}}} = (R_{\mathrm{a}}, G_{\mathrm{a}}, B_{\mathrm{a}})$，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)，$(x + S_{\mathrm{area}}, y)$，$(x, y + S_{\mathrm{area}})$，$(x + S_{\mathrm{area}}, y + S_{\mathrm{area}})$ 内の領域を走査しオリジナル画像の画素値 $\boldsymbol{c_{\mathrm{o}}} = (R_{\mathrm{o}}, G_{\mathrm{o}}, B_{\mathrm{o}})$ を求める．画素値の算出には中央値を利用し，物体のハイライト成分や影の成分の影響をなるべく小さくするように工夫した．

オリジナル画像の画素値 $\boldsymbol{c_{\mathrm{o}}}$ を取得すると，ユーザの解答した画素値 $\boldsymbol{c_{\mathrm{a}}}$ との乖離度を計算する．ユーザの解答した画素値と，出題画像の画素値を比較することによって解答照合の自動化を図っている．本論文では，画素値の乖離度を HSV 空間，CIELAB 空間上におけるユークリッド距離で定義した．HSV 空間は人間の認識に近い色の見えを表現するために考案された色空間であり，画素値の乖離度を表現するのに適していると考えたが，色相情報である H が角度であるため，単純にユークリッド距離を求めることはできない．したがって我々は，HSV 空間の座標を文献 [17] で利用された urV 空間に投影してからユークリッド距離を求めることとした．urV 空間は，HSV 空間の値を用いて，式 (4) ～ (6) のように投影される．色相情報を角度として考えた円柱状の空間に投影することで H の値を考慮したユークリッド距離を求めることができる．

$$u = S \times \cos H \tag{4}$$

$$r = S \times \sin H \tag{5}$$

$$V = V \tag{6}$$

一方，CIELAB 空間は，空間上での色の差が，人間の感じる色の差と近くなるように設計された色空間で，L^*，a^*，b^* の 3 つの値で表現される．CIELAB 空間へは，RGB 空

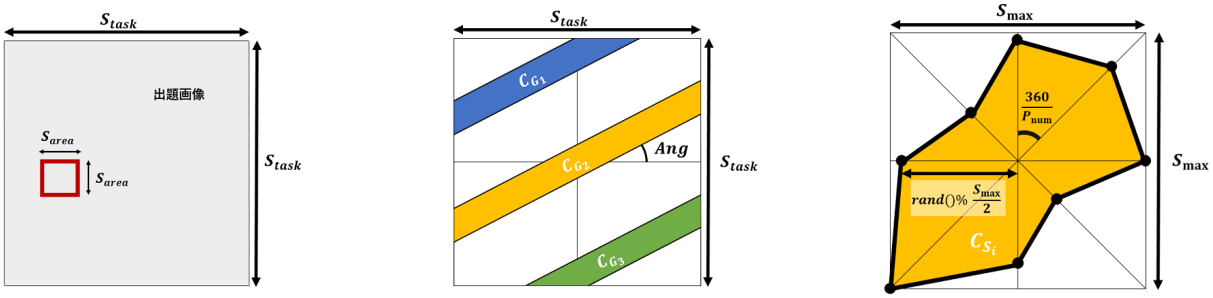


図 3 提案 CAPTCHA のパラメータ (左：共通, 中央：グラデーション型，右：シェイプ型)

間から CIEXYZ 空間という色空間に変換したのち，式 (7) ~ (9) で求めることができる．X_n，Y_n，Z_n は CIEXYZ 空間における白色点を示す座標である．

$$L^* = 116f(Y/Y_n) - 16 \tag{7}$$

$$a^* = 500\{f(X/X_n) - f(Y/Y_n)\} \tag{8}$$

$$b^* = 200\{f(Y/Y_n) - f(Z/Z_n)\} \tag{9}$$

ただし，ここで $f(t)$ は式 (10) で定義される．

$$f(t) = \begin{cases} t^{\frac{1}{3}} & t > (\frac{6}{29})^3 \\ \frac{1}{3}(\frac{29}{6})^2 t + \frac{4}{29} & otherwise \end{cases} \tag{10}$$

5. 評価実験

5.1 色妨害による色恒常性への影響

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かどうか検証する必要がある．よって，本論文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，色妨害の加わった画像のにおいても色妨害のない状況と同じように色を判断できると考えられる．したがって本実験では，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定がどのくらい一致しているかを確認する．色判定の一致度を測る指標として，本論文では κ 係数を利用することにした．本来，色恒常性が成立するかどうかを調査する際には，それ以外の色覚特性が働かない状況を特殊な装置で実現してから行うべきである．すなわち，本実験手法で色恒常性の成立度合いを厳密に評価することは難しい．しかしながら，CAPTCHA として本手法を提供する際には，ユーザが色恒常性以外の色覚特性が働かない特殊な状況で解答することは考えられないため，このような評価方法を取った．

また，実験を行う際，出題画像の候補として，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一般に既知である現実の画像
(2) 物体と色の対応関係が一般に既知である幾何学的な画像
(3) 物体と色の対応関係が一般に未知である現実の画像
(4) 物体と色の対応関係が一般に未知である幾何学的な画像

ここで，対応関係が一般に既知である画像は，イチゴや有名なキャラクターなど，色が一般的に広く知られている物体が写っているものを指す．そして対応関係が一般に未知である画像は，広く浸透していなかったり，服や車などカラーバリエーションが複数存在したりする物体が写っているものを指す．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．画像型 CAPTCHA において，出題として利用する画像は主にオンライン上で公開されたフリー利用可能な画像となることが多いが，その多様さから人間も誤解しやすい画像が含まれることがあり，そのような画像は出題画像として利用しないようにする方策が一般的である [6] [9]．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．すなわち，出題画像に適した画像が出題として選ばれていることを前提として実験を行った．

5.2 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，我々はそれにあてはまるカテゴリとして，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．出題画像は次の条件で収集した；(1) 3 色以上利用されている，(2) 画像内の物体の色の範囲が解答エリアを包含する程度に大きい．実際に利用した画像の例を図 4 に示す（キャラクターは著作権の観点から掲載していない）．

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 グループにそれぞれ 5 人ずつに分けた．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，計 140 箇所を答えさせる．その後，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

図 4　実験で利用した出題画像の例（左：動物カテゴリ、右：食べ物カテゴリ、下：国旗カテゴリ）

表 1　実験環境

蛍光灯の明るさ	ディスプレイの色温度
2900lm	6500K∼9300K

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．

5.3　実験環境

実験時の環境を表 1 に示す．ユーザが CAPTCHA を利用する際にディスプレイの色は個々人によって異なるため，ディスプレイの色温度は特に固定せずに実験を行った．CAPTCHA のパラメータを表 2 に示す．これらのパラメータは経験的に設定されたものである．フィルタの明度と彩度は，低いと妨害としての効果が薄くなってしまうため，少なくとも 50%以上であることが望ましいと考え，50% ∼ 100% との間のうち，経験的に 60%と設定した．

5.4　Cohen の κ 係数

Cohen の κ 係数は，二者間の判定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．計算式を式 (11) に示す．

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{11}$$

ここで，p_o は実際の一致率，p_e は偶然の一致率である．文献 [18] では κ 係数を算出するのに十分な大きさのサンプルサイズを 100 としており，今回の実験では全部で 40 枚の画像であることから，1 枚の画像につき平均 2.5 か所以上指定する必要があった．

κ 係数の目安を Landis らは表 3 のように定義している [19]．κ 係数が 0.81 以上の場合は一致率は十分に高いと判定できる．

5.5　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．基本色名の定義を以下に示す [12].

(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

本論文では，「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．κ 係数は名義尺度と順序尺度の問題以外には適用できないため，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． また，被験者に答えてもらう際に例示として色そのものを表示すると回答に影響があると考えたため，色の名前のみを被験者に提示した．

5.6　実験結果

実験結果を表 4 に示す．κ 係数はグラデーション型が 0.875，シェイプ型が 0.865 と，どちらも 0.81 以上となった．κ 係数が 0.81 以上なら一致率は十分に高いと判定できるため，少なくとも色と物体の色の見え方は妨害の有無にかかわらない可能性が高いことが分かった．このことから，提案手法は人間の認識率を損ねない妨害が可能な CAPTCHA の形態である可能性が示せた．また妨害が加えられても人間の認識できる色が変わりにくいことから，妨害を加えていない従来のカラー型 CAPTCHA [11] 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，従来のカラー型 CAPTCHA よりもセキュリティを高められる可能性が示せた．今後は物体と色の対応関係が一般に未知である画像についても κ 係数の調査を行っていく必要がある．

6.　まとめ

本論文では，色恒常性の，「周囲の照明光の影響を受けても物体本来の色を知覚できる」，「機械的に再現が困難である」という点に着目して，人間の認識能力を損なわない妨害が可能な CAPTCHA を検討した．提案した CAPTCHA では，提示した出題画像に妨害フィルタを加えた上で，ユーザに解答エリア内の色を答えさせる．妨害フィルタによって出題画像の色が妨害されても，色恒常性により元来の色を知覚できる可能性がある．また，妨害フィルタには色恒常性のアルゴリズムとして一般的に知られた Gray-World アルゴリズムに耐性を持たせ，機械耐性を考慮した．

非現実的な色妨害を行っても人間の色恒常性が働くかど

表 2　CAPTCHA のパラメータ

グラデーション型		シェイプ型	
CAPTCHA のサイズ S_{task}[px]	300 × 300	CAPTCHA のサイズ S_{task}[px]	300 × 300
出題エリアのサイズ S_{area}[px]	20 × 20	出題エリアのサイズ S_{area}[px]	20 × 20
フィルタの彩度 [%]	60	フィルタの彩度 [%]	60
フィルタの明度 [%]	60	フィルタの明度 [%]	60
透明度	0.5	透明度	0.5
動作間隔 F_{g}[ms]	5	動作間隔 F_{s}[ms]	33.33
動作間隔あたりの色相増分	0.5	図形の個数 N	100
グラデーションの総数 L_{num}	3	図形の最大サイズ S_{max}[px]	40 × 40 px
		図形の辺の数 P_{num}	10

表 3　Landis and Koch (1977) による κ 係数の目安 [19]

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

表 4　色妨害なしグループと色妨害グループとの平均 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

うかは不明であるため，実験では，κ 係数という二者間の判定の一致度を測る尺度を利用し，非現実的な画像に対しても色恒常性が働くか調査した．具体的には，色妨害の加えられていない画像を見たグループと，加えられている画像を見たグループとの間でどの程度判定が一致するかを確認した．実験の結果，我々が検討した妨害フィルタでは，いずれも妨害の加えられていない画像を見たグループとの間の κ 係数は 0.81 以上を超えており，色の判定が極めて高く一致していることから，色恒常性は少なくとも CAPTCHA が成立する程度に働いていることが分かった．この結果から，色の判定は妨害の有無によって大きく左右されないことが分かったため，従来のカラー型 CAPTCHA に比べて，認識率を下げることなく機械耐性を高めることができる可能性が示唆された．

本論文で確認したのは検討した色妨害フィルタが CAPTCHA の妨害として適切かどうかのみであり，CAPTCHA としての性能は調査できていない．実際に CAPTCHA として被験者に回答させ，人間と機械に回答に差があるか確認して人間の認識率と機械の耐性を調査する必要がある．また，今回の実験において，3 次元の画像の色は具体的に色を答え辛いという意見もあったことや，今回は調査しなかったカテゴリの画像に対して色恒常性がどの程度働くかは不明であることから，人間にとってより適切な出題形式も今後検討していく必要がある．

また，色恒常性アルゴリズムに耐性のある妨害を設計目標として検討を行ったが，今後は画像処理技術への耐性も考慮に入れる必要がある．例えばフレーム画像を収集して時間方向に平均を取ると，色恒常性アルゴリズムを利用せずとも元の色を推定される恐れがある．こうした攻撃に耐性を持たせるよう，画像の一部を切り取って出題画像として提示したり，出題画像そのものを動かしたりするなどの検討が必要である．

謝辞　本研究は JSPS 科研費 JP17H01736, JP17K00139, JP18K11268 の助成を受けたものです．

参考文献

[1] von Ahn, L., Blum, M., and Langford, J.: “Telling Humans and Computers Apart Automatically,” In *Communications of the ACM*, Vol.47, No.2 pp.57–60 (2004).

[2] Moy, G., Jones, N., Harkless, C., and Potter, R.: “Distortion Estimation Techniques in Solving Visual CAPTCHAs,” In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1123–1128 (2004).

[3] George, D., Lehrach, W., Kansky, K., Lázaro-Gredilla, M., Laan, C., Marthi, B., Lou, X., Meng, Z., Liu, Y., Wang, H., Lavin, A., and Phoenix, D.S.: “A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs,” *Science*, Vol. 358, No. 6368 (2017).

[4] Sivakorn, S., Polakis, I., Keromytis, D.A. “I Am Robot: (Deep) Learning to Break Semantic Image CAPTCHAs,” In *2016 IEEE European Symposium on Security and Privacy*, pp. 388–403 (2016).

[5] Yan, J., and Ahmad, A. S. E.: “Usability of CAPTCHAs or Usability Issues in CAPTCHA Design,” In *Proceedings of the 4th Symposium on Usable Privacy and Security*, pp. 44–52 (2008).

[6] Nejati, H., Cheung, N., Sosa, R., and Koh, D. C. I.: “DeepCAPTCHA: An Image CAPTCHA Based on Depth Perception,” In *Proceeding of the 5th ACM Multimedia Systems Conference*, pp. 81–90 (2014).

[7] von Ahn, L., Maurer, B., Mcmillen, C., Abraham, D., and Blum, M.: “reCAPTCHA : Human-based Character Recognition via Web Security Measures” *Science*, Vol. 321, No. 5895, pp.1465–1468 (2008).

[8] recaptcha plugin ⟨`https://developers.google.com/recaptcha/`⟩ (accessed 2018-03-20).

[9] Jeremy, E., Douceur, J., and Howell, J.: “Asirra: a CAPTCHA that Exploits Interest-Aligned Manual Image Categorization,” In *Proceedings of the International Conference on Computer and Communications Security*, pp. 366–374 (2007).

[10] Golle, P.: "Machine Learning Attacks Against the Asirra CAPTCHA," In *Proceedings of the 15th ACM Conference on Computer and Communications Security*, pp. 535–542 (2008).
[11] Kumar, M., and Dhir, R.: "Design and Comparison of Advanced Color based Image CAPTCHAs," *International Journal of Computer Applications*, Vol. 61, No.15, pp. 24–29 (2013).
[12] 内川惠二, 『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).
[13] van de Weijer, J., Gevers, T., and Gijsenij, A.: "Edge-Based Color Constancy," In *IEEE Transactions on Image Processing*, Vol. 16, No. 9, pp. 2207–2214 (2007).
[14] Celik, T., and Tjahjadim T.: "Adaptive Colour Constancy Algorithm Using Discrete Wavelet Transform," *Computer Vision and Understanding*, No. 116, pp. 561–571 (2012).
[15] Oh, S. W., and Kim, S. J.: "Approaching the Computational Color Constancy as a Classification Problem Through Deep Learning," *Pattern Recognition*, No. 61, pp. 405–416 (2017).
[16] Goswami, G., Powell, B.M., Vasta, M., Singh, R., and Noore, A.: "FaceD: Face Detection Based Color Image CAPTCHA", *Future Generation Computer Systems*, Vol. 31, pp. 59–68 (2014).
[17] 後藤 雄飛, 山内 悠嗣, 藤吉 弘亘『CS-HOG：色の類似性に基づいた形状特徴量』, 電子情報通信学会論文誌 D, Vol. 96, No. 7, pp. 1618–1626 (2013).
[18] Cohen, J., "A Coefficient of Agreement for Nominal Scales," *Educational and Psycological Measurement*, Vol .20, pp. 37–46 (1960).
[19] Landins, J.R., and Koch, G.G.: "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, Vol. 33, No. 1, pp. 159–174 (1977).