色の恒常性を利用したCAPTCHAの機械攻撃耐性の評価と分析

永井 麻裕[1]　臼崎 翔太郎[1]　油田 健太郎[1]　山場 久昭[1]　椋木 雅之[1]　岡崎 直宣[1]

概要：人間の高度な認知能力を利用したCAPTCHAの1つとして色の恒常性を利用したcolorCAPTCHAが提案されている．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能力のことである．色恒常性CAPTCHAは，色妨害フィルタを加えた画像をWebページに提示し，ユーザが元の画像の色を推定できるか否かで人間と機械を判別する．人間には色の恒常性が備わっているので容易に認識できるが，機械にとっては再現が困難であり，それがセキュリティとユーザビリティの前提になっている．しかし，このCAPTCHAにはいくつかの機械攻撃が考えられる．そこで本論文では，機械攻撃に耐性のある新しい色妨害フィルタを作成し，ユーザビリティ，機械攻撃耐性の評価を行った．

キーワード：CAPTCHA，認証，色恒常性

Evaluation and analysis of the color constancy CAPTCHA against automated attacks

Mayu Nagai[1]　Shotaro Usuzaki[1]　Kentaro Aburada[1]　Hisaaki Yamaba[1]　Masayuki Mukunoki[1]
Naonobu Okazaki[1]

Abstract: Color constancy CAPTCHA has been proposed as one of the CAPTCHAs that use the advanced cognitive abilities of humans. Color constancy is a human characteristic that enables humans to recognize the object's original color by ignoring the effects of illumination light. This CAPTCHA tells computers and humans apart based on whether or not the user can estimate the original image's color when presented with images with a color filter. Humans have color constancy and can easily recognize colors, but computers have difficulty reproducing them, which is a prerequisite for security and usability. However, this CAPTCHA may be vulnerable to some automated attacks. In this paper, we developed a new color filter that is resistant to automated attacks and evaluated its usability and resistance to automated attacks.

Keywords: CAPTCHA, authentication, color-constancy

1. はじめに

Webサービスの普及により，誰でも様々なWebサービスを利用することが可能となっている． しかし一方で，機械によるアカウントの不正取得や，ブログなどでのスパム行為といった迷惑行為が発生している． その対策として，CAPTCHA(Completely Automated Public Turing test to tell Computers and Humans Apart) と呼ばれる人間と機械を判別する技術が広く利用されている．CAPTCHAの代表例としては，歪曲やノイズが付加された文字列画像をWebページに提示し，ユーザがその文字を判読できるか否かを試す文字列型CAPTCHAや，複数の画像をWebページに提示し，ユーザが条件に合う特定の画像を選択できるか否かを試す画像型CAPTCHAなどがある．しかしながら，これらのCAPTCHAはOCR技術や機械学習技術の発展によって機械が高い精度で突破できることが報告されている [1, 2]． そのため，人間による正答率が高く機械による正答率が低い，新たなCAPTCHAが必要である．

[1] 宮崎大学
University of Miyazaki

図 1: colorCAPTCHA-1 [3]

図 2: colorCAPTCHA-2 [3]

そこで，人間の色覚特性である色の恒常性を利用した colorCAPTCHA が提案された．この CAPTCHA は，ベース画像に色妨害を加えた画像を Web ページに提示し，ユーザがベース画像（色妨害を加えていない状態の画像）の色を推定できるか否かで人間と機械を判別する．しかしこの CAPTCHA には，機械攻撃耐性が低いという課題がある．そこで本研究では，新しい色妨害フィルタの作成手法を提案し，人間の正答率を維持したまま，機械の正答率低下を目指す．また，この工夫によって人間の正答率，機械の正答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 章では提案手法について解説する．4 章ではユーザビリティ，機械攻撃耐性の評価を行う．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，色の名前を認識することがコンピュータには困難であるが，人間には容易であることを利用した CAPTCHA である [3]．具体的には，ランダムに選ばれたカラー画像を Web ページに提示し，ユーザが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の名前を認識できるか否かで人間と機械を判別する．評価実験では，職種を問わず 5 歳以上の 1000 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．この手法では機械が色から名前を認識できないことを前提に設計されているため，これまでの文字列型 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，機械攻撃耐性は実験の評価対象とされておらず，現在の機械学習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難い．

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機械に突破されてしまう恐れがあるため，人間の正答率を保証しつつ，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 [4]．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の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，色妨害フィルタを加えた画像を Web ページに提示し，解答領域の色と最も近いと思う色を 8 色のカラーパレット（赤，青，緑，黄，紫，茶，オレンジ，ピンク）から選択させる．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ベース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．色恒常性 CAPTCHA の出題画像はベース画像と色妨害フィルタで構成されており，ベース画像には有名な色恒常性アルゴリズムである Gray-World に耐性を持たせるために，画像の一部分を除いてグレースケール化を行っている．色妨害フィルタは，ベース画像を基に決定した 3 色を用いたグラデーション画像（以降，グラデーション型フィルタ）である．しかしこのグラデーション型フィルタに対しては，平均画素値を利用した攻撃，画像検索攻撃，局所領域抽出攻撃が考えられる．

平均画素値を利用した攻撃とは，出題画像の平均画素値の補色が，ベース画像の平均画素値と同等であると仮定した攻撃である．先行研究 [4] では，ベース画像と色妨害フィルタの合成比率を 4:6 と，色妨害フィルタの割合を高く設定している．そのため，出題画像の平均画素値を求めた時に色妨害フィルタの色が強く表れる．グラデーション

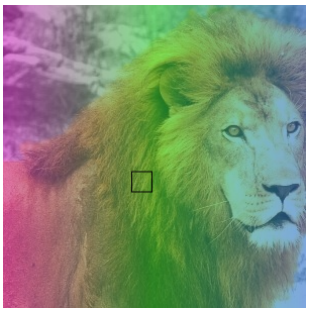

図 3: 色の恒常性を利用した colorCAPTCHA の出題例 [4]

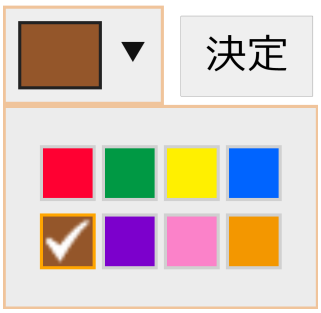

図 4: カラーパレット [4]

型フィルタはベース画像の平均色の反対色で構成されているため，出題画像の平均画素値の補色を求めることでベース画像の平均色に近い色に戻すことができる．

画像検索攻撃とは，CAPTCHA の出題画像を Web 上の検索エンジンで検索し，色妨害が掛かっていない画像を取得しようとする攻撃である．色妨害が掛かっていない画像を取得された場合，容易に突破できることが予想される．

局所領域抽出攻撃とは，画像の一部分を抽出して色恒常性アルゴリズムを適用し，色妨害フィルタを除去しようとする攻撃である．グラデーション型フィルタを使用した画像の場合，画像の一部分を抽出することで単色フィルタのようになってしまう．その場合，色恒常性アルゴリズムによって容易に色妨害フィルタを取り除くことができる．

3. 提案手法

本章では，提案 CAPTCHA の概要，2.2 で説明した機械攻撃に耐性がある色妨害フィルタの作成手法，解答の照合方法について述べる．

3.1 概要

CAPTCHA の出題形式そのものは先行研究 [4] と変わっておらず，色妨害フィルタを加えた画像を Web ページに提示し，その画像上の解答領域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ものである．解答領域はランダムな位置に表示されるが，ユーザは解答領域をドラッグ，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．

色選択の手段として，先行研究 [4] ではカラーパレットを使用していたが，本研究ではカラースライダーを用いる．コンピュータの計算によって正解となる色を求めているため，色の選択肢が少ないカラーパレットでは人間の感覚と異なる色が正解として選ばれることがある．意図しない形での人間の正答率低下を防ぐために，カラースライダーを採用した．

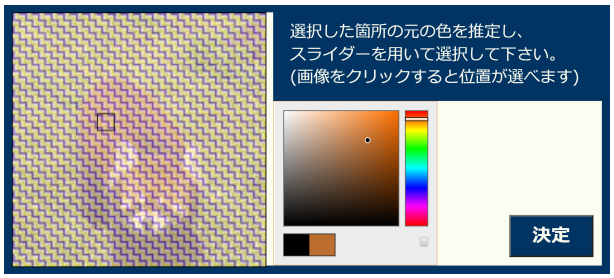

図 5: 提案 CAPTCHA

提案 CAPTCHA の出題例を図 5 に示す．左側に出題画像，右側にカラースライダーを配置している．出題画像は，元となるベース画像，及び色妨害フィルタで構成されている．ベース画像には，先行研究 [4] のようなグレースケール化は行っておらず，トリミング加工のみを行った． 本研究では，機械攻撃耐性向上を目的とした色妨害フィルタの作成手法を提案する．

3.2 色妨害フィルタ

本研究の色妨害フィルタは，図 6 のようなグラデーション画像を縮小し，タイル状に敷き詰めた図 7 のような画像（以降，タイル型フィルタ）である．2.2 で説明した機械攻撃に耐性を持たせるため，いくつかの工夫を施した．具体的には，局所領域抽出攻撃対策としてタイル型に，色恒常性アルゴリズム及び平均画素値を利用した攻撃への対策として構成色を工夫した．

タイル型にしたのは，同じ色の範囲が狭ければどの部分を抽出されても単色フィルタにならないと考えたためである．構成色に関しては，事前実験において Gray-World 及び Max-RGB 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分かっていたので，耐性を高められるような色選択を行った．具体的には，Gray-World 対策として画像の代表色の補色，Max-RGB 対策として白色を使用した．Gray-World は，RGB の平均値が無彩色になるという仮説に基づいたアルゴリズムである [5]． 色には反対色同士の色を平均すると灰色になるという性質があるので，ベース画像の色とその補色とで相殺されて灰色になり，本来の色を復元できなくなるのではないかと考えたため，補色を採用した． また，補色は対の関係にあるため，画像の代表色 1 色の補色にした場合，攻撃者に推測されやすくなると考えたため代表色 2 色の補色とした．Max-RGB は，RGB の最大値から光源色を推定するアルゴリズムである [5]．RGB の最大値である白色を混ぜることで，本来

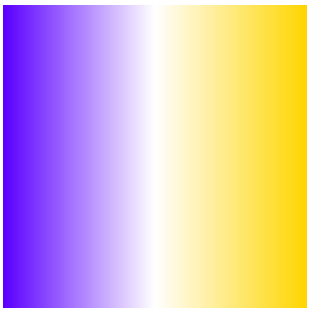
図 6: グラデーション画像

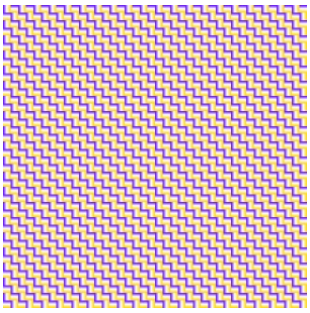
図 7: タイル型フィルタ

の光源色を推定できなくなるのではないかと考えたため白色を採用した．

色空間は，人間の認識に近い色の見え方を表現するために考案された HSV 色空間を使用した．HSV 色空間は，色相 (0˜360°)，明度 (0˜100%)，彩度 (0˜100%) の組み合わせで色を表現する．色相は画像の平均色を基に決定し，明度，彩度は 100%に設定した．また，透明度に関しては人間に色の恒常性が働き，高い確率でベース画像の色を認識できる，70%に設定した．ただし，これは一人のみの予備実験で決定したため，最適な値とは限らない．タイル型フィルタの作成手順を以下に示す．

(1) 画像を 10 × 10 に 100 等分する．

(2) 100 等分した画像それぞれの平均色 (RGB) を取得し，HSV に変換する．

(3) 平均色 (HSV) の色彩の値 H をもとに，3 つのグループに分類する．

$$RED = \{0 \leqq H < 60,\ 300 \leqq H < 360\} \quad (1)$$

$$GREEN = \{60 \leqq H < 180\} \quad (2)$$

$$BLUE = \{180 \leqq H < 300\} \quad (3)$$

(4) 要素数の多い 2 グループの中央値をそれぞれ h1, h2 とし，以下のような 2 色を作成する．

$$(H,S,V) = \begin{cases} ((h1 + 180) \bmod 360, 100, 100) & (4) \\ ((h2 + 180) \bmod 360, 100, 100) & (5) \end{cases}$$

(5) 作成した 2 色と白色を用いてグラデーション画像を作成する (図 6)．

(6) 作成したグラデーション画像を 1/40 に縮小し，画像 A を作成する．

(7) 画像 A を右に 90 度回転させ，画像 B を作成する．

(8) 画像 A と画像 B を交互に縦横 40 枚ずつ敷き詰め，タイル型フィルタを作成する (図 7)．

3.3 解答の照合方法

解答の照合には，ベース画像，ユーザが解答した色の画素値，及び解答領域の座標を利用する．

ユーザが解答した色の画素値 Ca = (Ra，Ga，Ba)，解答領域の左上の座標（x,y）が与えられると，システムはベース画像の (x,y)，($x + S$area，y)，(x,$y + S$area)，($x + S$area，$y + S$area) 内の領域を走査し，解答領域におけるベース画像の画素値 Cb = (Rb，Gb，Bb) を求める．画素値の算出には，物体のハイライト成分や影の成分の影響を小さくするため，解答領域内の中央値を利用した．

そうして得られた，解答領域におけるベース画像の画素値 Cb とユーザが解答した色の画素値 Ca の色差を計算し，色差が閾値以下であれば人間と判定される．色差の計算には，人間の色の見え方を考慮した解答の照合をするために，CIEDE2000 [6] 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. 評価実験

本章では，提案したタイル型フィルタが人間の認識を著しく損ねないものであることを確認するためのユーザビリティ評価，及び 2.2 で説明した機械攻撃にどの程度耐性があるのかを確認するための機械攻撃耐性の評価を行う．また，先行研究のグラデーション型フィルタにも同様の攻撃を実行し，比較対象とした．

4.1 ユーザビリティ評価

本節では，タイル型フィルタが人間の認識を著しく損ねないものであることを確認する．正答率，所要時間に加え，実験終了後に行った System Usability Scale(SUS) を用いたアンケートによって評価する．

System Usability Scale(SUS) は，1986 年に John Brooke が開発した，ユーザビリティについての主観的な評価の指標である [7]． 評価に用いられる 10 項目の質問を以下に記す． 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おり，1(強く反対する) から 5(強く賛成する) の 5 段階で評価される．

(1) この CAPTCHA をしばしば利用したいと思う

(2) この CAPTCHA を利用するには説明が必要となるほど複雑であると感じた

(3) この CAPTCHA は容易に使いこなす事ができると思った

(4) この CAPTCHA を利用するのに専門家のサポートが必要だと感じる

(5) このCAPTCHAにあるコンテンツやナビゲーションは十分に統一感があると感じた

(6) このCAPTCHAでは一貫性のないところが多々あったと感じた

(7) たいていの人は、このCAPTCHAの利用方法をすぐに理解すると思う

(8) このCAPTCHAはとても操作しづらいと感じた

(9) このCAPTCHAを利用できる自信がある

(10) この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思う

評価値の集計方法は以下のようになっている．奇数項目(ポジティブな質問)は回答番号から1を引き，偶数項目(ネガティブな質問)は5から回答番号を引く．全ての項目を0から4で評価し，足し合わせた合計数値を2.5倍して0から100のスケールへ変換する．各項目のスコアをN_1からN_{10}とすると，合計スコアSは式(6)で表すことができる．

$$S = (\sum_{i=1}^{10} N_i) \times 2.5 \tag{6}$$

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の平均スコアは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[8].

宮崎大学の工学部生13名を対象として，実験を行った．被験者には，色恒常性CAPTCHAについての説明を行い，慣れるまで練習してもらった後にランダムに選ばれた10問を出題した．その際，出題画像，ユーザが解答した色の画素値，解答領域の座標，解答領域におけるベース画像の画素値，所要時間を記録した． そして実験終了後，SUSによるアンケート調査を実施した．実験画像は先行研究[4]と同様の，現実世界で撮影された人物，乗り物，動物，風景等の49枚の画像を使用した．画像サイズは300px×300pxで，入手元はOpen Image Dataset V5 [9]である．

被験者13名が提案CAPTCHAを10回ずつ解いて得られた130件のデータから，平均正答率，平均所要時間，最小所要時間，最大所要時間を算出した．これらの値と，SUSの平均スコアを表1に示す．

平均正答率は78.5%であり，代表的な画像型CAPTCHAであるreCAPTCHAv2の平均正答率84.1% [10]には及ばなかった．本研究では解答方式をカラースライダーとしたが，アンケートでは「選択肢が多すぎて選びにくい」という意見があったためUIの改善が必要だと考える．また，実験に使用する画像を人物，乗り物，動物，風景等の画像としていたが，アンケートでは「何の写真か分からないものがあった」,「もっと色が分かりやすい写真にしてほしい」という意見があったため，物体と色の対応関係が一般に既知である画像に変更する必要があると考える．平均所要時間は13.2秒であり，画像型CAPTCHAの平均所要時間6.71秒[10]と比べて長く，これに関しても対策が必要であると考える．平均SUSスコアは80.6で，優れたユーザビリティを示すスコア80.3と同程度であった．

表1: 実験・アンケート結果

平均正答率	78.5%(102/130)
平均所要時間	13.2秒
最小所要時間	2.33秒
最大所要時間	41.2秒
平均SUSスコア	80.6

4.2 機械攻撃耐性の評価

4.2.1 平均画素値を利用した攻撃

本節では，タイル型フィルタが平均画素値を利用した攻撃にも頑強であることを確認する．ベース画像と色妨害フィルタの合成比率を先行研究では4:6，本研究では3:7と，どちらも色妨害フィルタの割合を高く設定している．また，色妨害フィルタの構成色として先行研究ではベース画像の平均色の反対色，本研究ではベース画像の代表色2色の補色を使用している．

そのため攻撃手順としてはまず，出題画像の平均画素値を求めることで，色妨害フィルタの平均色を取得した．次に，出題画像の平均画素値の補色を求めることで，ベース画像の平均色に近い色に戻した．そうして得られた色を解答とし，その色と正解の色との色差が閾値以下であれば攻撃成功とした．ユーザビリティの実験において得られたデータ（出題画像，ユーザが選択した解答領域の座標，解答領域におけるベース画像の画素値）を使用して，平均画素値を利用した攻撃を行った結果を表2に示す．

表2から，平均画素値を利用した攻撃に対してはタイル型フィルタの方が頑強であることが分かる．色妨害フィルタの構成色にベース画像の代表色1色ではなく代表色2色を使用したことが有効だったのではないかと考える．

表2: 平均画素値を利用した攻撃の成功率

色妨害フィルタ	成功率
グラデーション型フィルタ	23.3%(42/180)
タイル型フィルタ	6.92%(9/130)

4.2.2 画像検索攻撃

本節では，タイル型フィルタが画像検索攻撃にも頑強であることを確認する．Google画像検索[12]を用いて提案CAPTCHAの出題画像を検索することで，色を認識するのに十分な画像が検出されるかを調査した．Google画像

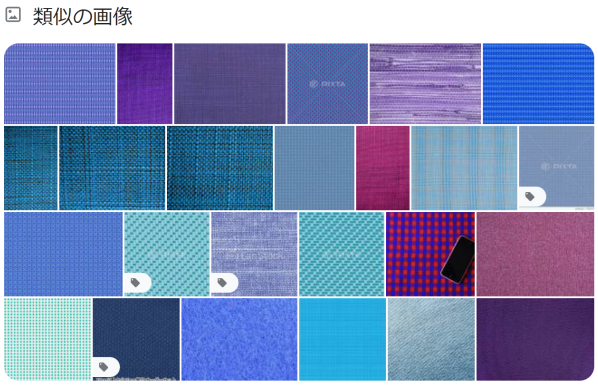

図 8: 画像検索攻撃の失敗例 [12]

検索の検索結果には関連コンテンツの他に「この画像の最良の推測結果」として，入力した検索キーの画像ファイルから推測されるキーワードが返され，「類似する画像」として，入力した検索キーの画像ファイルから類推される画像の一覧が出力される．出題画像を 1 枚ずつ検索にかけ，元の画像が表示された場合，もしくは「類似する画像」の中に，色を認識するのに十分な類似画像が表示された場合を攻撃成功とみなした．ユーザビリティの実験で使用した 49 枚の出題画像を対象として，画像検索攻撃を行った結果を表 3 に示す．表 3 から，タイル型フィルタには画像検索攻撃がほとんど成功していないことが分かる．「この画像の最良の推測結果」には複数の画像で，pattern，vertical，dot というキーワードが表示され，全ての画像において正しいキーワードは表示されなかった．また「類似する画像」には，ほとんどの画像において，布のような画像，又は格子状の画像が表示された．その一例を図 8 に示す．これらのことから，画像検索エンジンが物体の輪郭情報を正しく認識できていないのではないかと考える．局所領域抽出攻撃への対策として行ったタイル型にする工夫が．画像検索攻撃に対しても有効であることが分かった．

表 3: 画像検索攻撃の成功率

色妨害フィルタ	成功率
グラデーション型フィルタ	32.7%(16/49)
タイル型フィルタ	4.08%(2/49)

4.2.3 局所領域抽出攻撃

本節では，タイル型フィルタが局所領域抽出攻撃にも頑強であることを確認する．攻撃手順としてはまず，解答領域の座標・パラメータ $Size$ を基に，図 9 のように解答領域から $Size$px 拡大した領域を抽出した．次に，抽出した画像に色恒常性アルゴリズムを適用し，解答領域中の RGB の中央値を解答とした．その色と正解の色との色差が閾値以下であれば攻撃成功とした．色恒常性アルゴリズムは，Gray-World，Max-RGB，Gray-Edge，2nd-Gray-Edge，Multi Layer Perceptron(MLP) を使用した [11]．ま

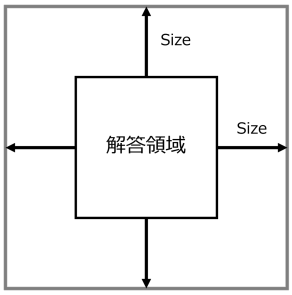

図 9: 抽出する領域

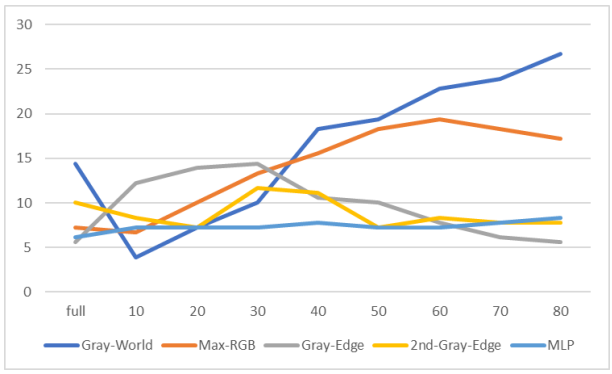

図 10: グラデーション型フィルタに対する局所領域抽出攻撃の成功率

た，ユーザビリティの実験において得られたデータ（出題画像，ユーザが選択した解答領域の座標，解答領域におけるベース画像の画素値）を使用した．グラデーション型フィルタに対する局所領域抽出攻撃の結果を図 10 に，タイル型フィルタに対する局所領域抽出攻撃の結果を図 11 に示す．

図 10，図 11 から，グラデーション型フィルタの場合は，局所領域を抽出する ($Size = 10\text{~}80$) ことで Gray-World，Max-RGB の攻撃精度が上昇しているが，タイル型フィルタの場合は，局所領域を抽出しても攻撃精度がほとんど上昇していないことが分かる．その原因としては，タイル型にしたことによってどの部分を抽出しても同じように色妨害が掛かっていたこと，もしくは Gray-World，Max-RGB への対策として行った構成色の工夫が効果的に働き，本来の光源色を推定できなかったことが考えられる．しかし，画像全体に色恒常性アルゴリズムを適用する攻撃（$Size = full$）の成功率が高くなってしまった．それはタイル型フィルタが画像全体で見れば，単色フィルタのようになってしまうことが原因ではないかと考える．最大の攻撃成功率は，グラデーション型フィルタの場合 26.7%，タイル型フィルタの場合 26.9%であり，同程度の耐性であると言える．

4.3 考察

実験結果から，平均画素値を利用した攻撃，画像検索攻撃に対しては本研究で提案したタイル型フィルタの方が

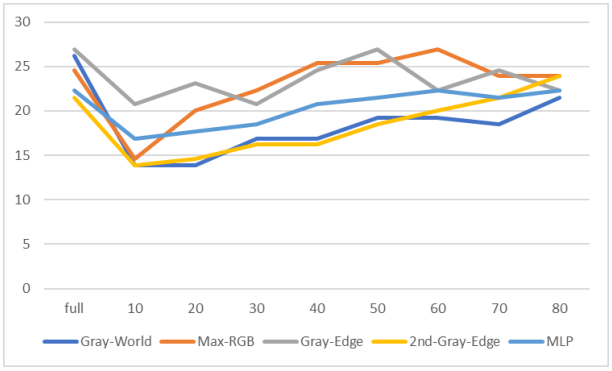

図 11: タイル型フィルタに対する局所領域抽出攻撃の成功率

頑強であり，局所領域抽出攻撃に対しては先行研究のグラデーション型フィルタと同程度の耐性であることが分かった．平均画素値を利用した攻撃，局所領域抽出攻撃に対しては色妨害フィルタの構成色の工夫が，画像検索攻撃，局所領域抽出攻撃に対しては形状の工夫が効果的だったのではないかと考える．また，グラデーション型フィルタは画像の一部分を抽出することで単色フィルタのようになってしまい，局所領域抽出攻撃に耐性が低くなるため，本研究では部分的に抽出されても単色フィルタにならないようにタイル型に変更していた．しかし，タイル型にすることで画像全体で見たときに単色フィルタのようになってしまったため，局所領域抽出攻撃の $Size = full$ のとき，つまり色恒常性アルゴリズムによる攻撃の精度が上がってしまった．このことから，色恒常性アルゴリズムによる攻撃と局所領域抽出攻撃はトレードオフの関係にあるのではないかと考える．

先行研究のグラデーション型フィルタに対する最も効果的な攻撃は画像検索攻撃の 32.7%で，本研究で提案したタイル型フィルタに対する最も効果的な攻撃は局所領域抽出攻撃の 26.9%であった．本研究で提案した色妨害フィルタの作成手法によって機械攻撃の成功率を約 5 ポイント削減することができた．他の CAPTCHA が 60%以上で突破されている [1, 2] ことを考えると提案 CAPTCHA の 26.9%は，その半分以下であり機械攻撃耐性は高いのではないかと考える．

ユーザビリティに関しては，平均正答率は 78.5%，平均所要時間は 13.2 秒であり既存の画像型 CAPTCHA と比較して劣っていたが，これは解答方式のカラースライダーが扱いにくいことや，使用した画像が CAPTCHA に向いていなかったことが原因としてある．ユーザが解答しやすい方法に変更すること，及び使用する画像をカラーバリエーションが少ない，物体と色との対応関係が一般に既知である画像に変更することで，正答率，所要時間は改善されるのではないかと考える．

5. おわりに

本研究では，先行研究である色の恒常性を利用した colorCAPTCHA の機械攻撃耐性向上のため，新しい色妨害フィルタの作成手法を提案し，その効果を確認するためユーザビリティ評価，機械攻撃耐性の評価を行った．

先行研究のグラデーション型フィルタには，平均画素値の補色がベース画像の平均画素値と同等であると仮定した，平均画素値を利用した攻撃，色妨害が掛かっていない画像を取得しようとする画像検索攻撃，画像の一部分を抽出して色恒常性アルゴリズムを適用し，色妨害フィルタを除去しようとする局所領域抽出攻撃に脆弱であるという課題があった．そこで，これらの機械攻撃への耐性を向上させるため，構成色，形状を工夫した色妨害フィルタの作成手法を提案した．

実験の結果，平均画素値を利用した攻撃，画像検索攻撃に対してはタイル型フィルタの方が頑強であり，局所領域抽出攻撃に対しては同程度の耐性であることが分かった．また，タイル型フィルタに対する機械攻撃の成功率は最大 26.9%であり，先行研究よりも機械攻撃の成功率を約 5 ポイント削減することができた．一方，人間の正答率は 78.5%，平均所要時間は 13.2 秒であり，既存の画像型 CAPTCHA と比較して劣っていることが分かった．

今後の課題としては，人間の正答率を向上させること，所要時間を短縮させることが挙げられる． そのためには，UI の改善，画像の選別などが必要だと考える．

謝辞 本研究は JSPS 科研費 JP18K11268，JP21K11849 の助成を受けたものです．

参考文献

[1] Jeff Yan, Ahmad Salah El Ahmad: A Low-cost Attack on a Microsoft CAPTCHA, ACM conference on Computer and communications security, pp.543-554, 2008.

[2] Fatmah H.Alqahtani, Fawaz A.Alsulaiman: Is image-based CAPTCHA secure against attacks based on machine learning? An experimental study, Computers&Security, vol.88, No.101635, 2020.

[3] Mandeep Kumar, Pathankot Renu Dhir: Design and Comparison of Advanced Color based ImageCAPTCHAs, International Journal of Computer Applications(0975-8887), Vol.61, No.15, pp.24-29, 2013.

[4] 臼崎翔太郎, 砂本佑紀, 岡崎直宣, 油田健太郎, 山場久昭: 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, 宮崎大学工学部紀要, Vol.49, pp.251-256, 2020.

[5] J Van De Weijer, T Gevers, A Gijsenij: Edge-Based Color Constancy, IEEE Transactions on Image Processing, Vol.16, No.9, pp.2207–2214, 2007.

[6] M.R.Luo, G.Cui and B.Rigg: The development of the CIE 2000 color–difference formula:CIEDE2000, Color Research& Application, Vol.26, pp.340-350, 2001.

[7] J.Brooke: SUS: A Quick and Dirty Usability Scale, Usability Evaluation in Industry. Taylor & Francis, pp.189-

194, 1996.
[8] Jeff Sauro: MEASURING USABILITY WITH THE SYSTEM USABILITY SCALE (SUS), Measuring U, https://measuringu.com/sus/, (Accessed 2022/01/24).
[9] Google: Open Image Dataset V5, https://storage.googleapis.com/openimages/web/factsfigures_v5.html, (Accessed 2022/01/24)
[10] N.Jiang, H.Dogan, F.Tian: Designing mobile friendly CAPTCHAs:An exploratory study, 31st British Human Computer Interaction Conference 2017, Vol.92, pp.1-7, 2017.
[11] Sami M, Khaled M, https://github.com/MinaSGorgy/Color-Constancy, (Accessed 2022/01/24)
[12] Google: Google 画像検索, https://www.google.co.jp/imghp?hl=ja&tab=wi, (Accessed 2022/01/24).