アモーダル補完を利用した 動画 CAPTCHA の提案

森　拓真† 　宇田　隆哉† 　菊池　眞之†

近年，ボットプログラムよる WEB サービスのアカウントが自動で大量取得され悪用されることが大きな問題となっている．これを防ぐ手法として CAPTCHA を導入するのが一般的となっている．CAPTCHA とは，人間と機械を判別するための認証テストである．現在で最も多く利用されている方式は，文字列 CAPTCHA であるが，OCR 技術の進歩により高い確率で解析されてしまうことがわかっている．また，文字に歪みを加えて読みにくくした CAPTCHA も提案されているが，ユーザより機械の正解率のほうが高いというジレンマを生み出している．これに対し，本稿では，人間の視覚補完能力であるアモーダル補完を動画に応用することで，人間のみ正解できる実用的な CAPTCHA を提案する．

Proposal of MOVIE CAPTHCA Method using amodal completion

TAKUMA MORI† 　RYUYA UDA†
MASAYUKI 　KIKUCHI†

Recently, a part of Bot-program can sign-up and get an account automatically on web service for attacking. These behaviors are problem for many web services. CAPTCHA is a one of solution for this problem. CAPTCHA tests registering user using Character recognition methods whether he is a human or a Bot-program. However, Bot-programs easily recognize the characters in high probability by using OCR technologies. Against for this problem, CAPTCHA changes the shape of the characters to make Bot-programs difficult to analyze it. But as the performance of OCR technology rises, the accuracy rate of the characters are analyzed has risen and is higher than human. Therefore, in this paper we propose a practical method for CAPTCHA only human can answer correctly by applying amodal completion.

†東京工科大学コンピュータサイエンス学部
Tokyo University of Technology School of Computer Science

1. はじめに

近年，WEB サービスの普及により，誰もが自由に様々なサービスを利用することが可能となっている．一方で，ボットと呼ばれる自動プログラムを用いて，サービスのアカウントを大量取得し，スパムメールなどに利用されるなどの事例が多く発生している．こうした攻撃は，サービス側に大きな被害をもたらすと考えられるため，ボットと人間を判別するチューリングテストの導入が必要不可欠である．

現在，WEB サービスで利用されているチューリングテストには，CAPTCHA（キャプチャ，″Completely Automated Public Turing test to tell Computers and Humans Apart″）が利用されている．CAPTCHA とは，相手が人間かコンピュータかを判別するために用いられるシステムである．人間には，容易に解くことが可能であるが，コンピュータには，解くことが難しいものを出題し，正しい解答をした者を人間と判断する．

CAPTCHA は，WEB サービスのアカウント取得や，掲示板やブログでの書き込みなどのサービスに広く利用されている．一般的に利用されている手法としては，文字列 CAPTCHA がある．文字列 CAPTCHA は，画像に描画された文字列をユーザに読み取らせ，テキストボックスへ入力させる方式である．アルファベットや数字をランダムに組み合わせた文字列を生成し，歪みやノイズなどの変形を加えることで，画像処理による解読を困難にしている．しかし，OCR 技術の進歩や，解読アルゴリズムの向上により，文字列 CAPTCHA は容易に突破されるようになってきている．これに対抗するために，歪みやノイズの強化による難読化が行われたが，ユーザの読み取り負荷が増大してしまう結果となった．複雑な背景画像を合成する手法や文字を連結して境界を曖昧にするなどの手法もあるが，いずれにしても読み取り負荷が増えてしまいユーザビリティに対して問題がある．これらの手法では，ユーザが読み取れないどころか，難読化した CAPTCHA でも正答率は，低下するものの解読されてしまっている．つまり，歪みやノイズの強化による手法は，効果を持たないと考えられる．

この問題に対して本提案では，人間の視覚補完を利用することで，ユーザビリティを確保しつつ，コンピュータの突破率を低下させる手法を提案する．提案手法は，物体が遮蔽された状態であっても内容を認知することができるアモーダル補完と呼ばれる視覚補完能力を利用する．アモーダル補完が起こると，遮蔽された文字であったとしても人間は瞬時にその文字が何であるかを知覚することが可能である．しかし，コンピュータは，認識率が大幅に低下する．また，知覚心理学の側面から文字の見やすさに着目し，これを動画に応用することで一意に解答が出せないよう曖昧さを持たせ解析コストを高める．

2. 既存手法

文字列 CAPTCHA の安全性に問題があることが指摘され,画像データを用いる手法や,人間の高次認識系を利用した CAPTCHA などの提案されている．しかしながら，どのような手法でも，完全に安全とは言えず，欠点がある．本章では，既存の CAPTCHA の手法とその問題点について述べる．

2.1 文字列 CAPTCHA

現在利用されている CAPTCHA の大部分は，文字列 CAPTCHA である．図 1(a)は，yahoo メールで利用されている文字列 CAPTCHA である．しかし，このような文字列 CAPTCHA は，OCR(Optical Character Recognition：光学文字認識)によって解読されている．図 1(b)は,文字列 CAPTCHA の一手法として有名な gimpy の簡易版の ez-gimpy である．Mori ら[1]は，これに対して，高い確率で打ち破ることができるとした論文を発表している.その手法では,92%の確率で ez-gimpy を解読することができるとしている．解読率を下げる目的として図 1(c)のようなより歪みやノイズを加えた文字列によって難読化を行ったが,人間による認識が極端に低下しユーザビリティの点で問題がある．図 1(d)は，reCAPTCHA[2]と呼ばれるもので，従来と異なったアプローチでの対策を行っている．その手法は，デジタル化した書籍データの中から，OCR で正しく識別されなかった単語を切り取り，CAPTCHA 用の画像データとして提供するというものである．ただし，CAPTCHA は機械と人間を区別することが主目的であり，正しく入力されたか判定するための「正解」が必要となる．そこで，OCR で正しく識別されなかった単語に加え，正しく識別された単語も用いる．問題として提示される画像データには，2 つの単語が含まれており，一方は正しく識別されており正解が存在する．もう一方は正しく識別されておらず，人間に読み取ってもらう必要のあるものである．しかしこの reCAPTCHA も OCR 技術の読み取り精度の向上によって破られる可能性がある．またそれほど高い精度を持たなくとも，ボットプログラムは，1 日に何万回ものチャレンジを行うので，十分に効果を発揮すると考えられる．既存の CAPTCHA や reCAPTCHA であっても結局は，文字認識のアルゴリズムによって解かれてしまうのが現状である．

2.2 PIX

文字列 CAPTCHA の問題を解決する手段として，画像データを用いた CAPTCHA が提案されている．画像データを用いた CAPTCHA の一手法に PIX がある．PIX は，共通した色や行動，形を認識できる画像を複数枚表示し，認証者に共通する分類を一つ選択させる方式をとっている．その選択が正答であった場合には認証者を人間とみなす．この手法は，画像が表わす情報の共通点を人間の高度な認知に基づきに選択させることで認証者を判断する．画像の内容を理解することは，高度な認知メカニズムであるため，機械的に解読するのは不可能であると考えられていた．そのため画像データを用いた CAPTCHA は，一見有効な手段であるかのように思えるが，実際には，大きな欠点を含んでいる．それは，データベースによる攻撃に弱いという点である．データベース攻撃とは，問題画像とその解を記録したデータベースを構築し，このデータベースを用いて問題を解く方法である．PIX では，画像に対してその特徴別に分類がされている．攻撃者は，何度も認証を繰り返すことで，画像データを取得する．取得した画像データに対して，特徴別に分類を行いデータベースの構築を行う．出題者は，画像データに対して，人間が認識できる特徴から分類を行わなくてはならないため，無作為に選択したデータを用いて出題することができない．従って，問題として構築されるデータベースの情報量は，人間が手入力で行える範囲である．このことから，攻撃者が，データベースを構築するのは，比較的容易であることが推測され，PIX がデータベース攻撃に弱いと考えられる．

2.3 NuCAPTCHA

NuCAPTCHA[3]は従来の文字列 CAPTCHA を動画へ応用したもので，カナダのソフトウェア企業 Leap Marketing Thechnologies が発表した CAPTCHA 手法である．

動画には，複数のフォントを用いたランダムな文字列が動画で表示される．ユーザは，動画上部に表示される色指定などを読み取り，動画中に流れる文字列中から該当文字列を読み取り入力を行う．この CAPTCHA は，文字列を読んで理解するという人間の視覚の高次認識プロセスを利用したものである．人間の知覚情報処理の分野において高次認識を説明するのは，難しい課題とされている．条件として採用される規則が多ければ，複数の組合せを考えることで有効な手段と成り得るが，色指定などの一定の規則であれば，パターンを予め登録しておくことで対処が可能である．動画中に流れる文字列は，一定時間内に数回の画像キャプチャを行えば，全ての文字を評価することが可能であるとともに，描画される文字列は，OCR で解読される CAPTCHA のものと同程度ものであると考えられる．従って，OCR での解析に対して耐性を持つとは考えにくい．

(a) yahoo captcha　(b) ez-gimpy

(c) 難読化したCAPTCHA　(d) reCAPTCHA

図 1　文字列 CAPTCHA の例

2.4 ニューラルネットワークの文字列 CAPTCHA 破りへの応用

Kumar ら[4]は CAPTCHA の解読手法にニューラルネットワークの機械学習を利用している．実験として，Google や Yahoo の CAPTCHA に対して適用した所，突破することが可能であることを示している．成功率は，5%から 50%程度と振れ幅が大きいが，いずれにせよ非常に高い確率で解読することが可能である．また文字の歪みに関して，人間とコンピュータの認識率を比較したものがある．この結果，歪みの度合いを高めた場合にもコンピュータは高い認識率を保っているのに対して，人間の認識率は，大きく低下するといった事態となっている．Kumar らは，これらのことから新たな CAPTCHA の手法として，文字への歪みやノイズに着目すべきでないと述べている．また文字列 CAPTCHA へ新たなアプローチをするのであれば，コンピュータで難しいとされる文字の分割へ目を向けるべきとした．

文字列 CAPTCHA を破る手法の大部分は，OCR 技術や文字認識アルゴリズムによるものが多い．それらは，サービスに提供される CAPTCHA のパターンから生成の仕組みを把握することで，特定のサービスに最適化したものである．そのため，CAPTCHA の生成ロジックを変更すれば即座にボット対策として効果を得られると考えられる．しかしながら，ニューラルネットワークを用いた CAPTCHA 破りのように複数のサービスに対して，適応できる手段は，CAPTCHA を生成する仕組みの変更に対しても効果を発揮することが可能である．攻撃者が現行の手段にニューラルネットワークを用いた場合に文字列 CAPTCHA は，現状よりも脆弱なシステムとなる可能性が高い．

人間の視覚特性における画像の認知メカニズムは，あらゆる方向から研究されており，十分な性能ではないものの補完処理をモデリングできるようになっている．福島邦彦らの研究手法[5]では，ニューラルネットワークの機械学習を利用した文字認識の研究を行っている．研究では，著しく変形したものや，文字の遮蔽度合が高くなければ，元の文字を認識することができることを示している．

3. 提案手法

既存手法で述べたように，文字列 CAPTCHA はコンピュータに対する難読化が行われる一方で，それを利用するユーザにとって，解読しにくくなっている問題がある．その大きな要因は，OCR 技術の進歩に対応するために，過度な歪みやノイズを文字列に加えることで人間による認識の範囲を超えてしまったためである．また，PIX のように人間の高次認識系の知覚特性を利用した画像データを用いた CAPTCHA に対しては，データベース攻撃に対する耐性が低い．データベース攻撃に関しては，攻撃者が手入力によるデータベースを構築するため，対策が難しいと考えられる．これは，画像データが有限であるという特性を持つためである．従って，CAPTCHA の手法は，OCR 技術に対応する手法として別のアプローチが必要となることに加え，無制限な問題生成が可能であることが望ましい．

そこで，本研究では，人間の視覚補完能力を動画に応用した文字列 CAPTCHA を用いることで，これらの問題点を解決する手法を提案する．本手法は，文字の断片とその遮蔽物を動画にして提示する．人間は，動画中に現れる文字を一瞬で知覚することができるため，難読化した文字列と比較して読み取り負荷を軽減することが可能となる．部分的に遮蔽された文字を OCR で解読するためには，遮蔽物に覆われた部分がどのような形態をしているかを解析しなければならない．すなわち，OCR 技術にて直接解析を行うことができなくなる．現在，遮蔽されている文字を復元する試みには，2.4 節で述べた福島らによるモデルがある．遮蔽物から，エッジ(文字の輪郭)の方向を予測することで，元の文字を復元することができる．本稿では，この遮蔽された文字に対する解析時間を解析コストと呼ぶ．問題として提示される画像が静止画の場合には，解析対象が 1 枚であるため，解析コストは少ないが，多くの画像を解析するには高いコストがかかる．本手法では，この解析コストに着目する．CAPTCHA の問題に対して，動画の解析コストよりも短い制限時間を設定することで，CAPTCHA の安全性を高める．

3.1 概要

人間が欠損した文字のような不完全な図形を見る場合，視覚特性上，ある条件下では，瞬間的にその内容が知覚できる．それは，欠損部分が遮蔽物で覆われた状態のことである．この現象を知覚心理学の分野では，アモーダル補完と呼んでいる．しかし，この条件に合わない場合には，よく考えれば内容を把握できる画像か，もしくは内容を知覚することができない画像となる．いずれの画像の場合においても，コンピュータを用いて処理を行えば，人間と同様の結論を導き出せることが過去の文献から明らかになっている．しかしながら，人間が視覚特性上瞬間的に内容を知覚できる画像のみは，人間とコンピュータの間に処理時間に関する決定的な差が生まれる．本研究では，画像解析の際に顕著となる人間とコンピュータにおける処理時間の差を，動画を用いることでさらに拡大して実用レベルまで引き上げる手法を提案する．

提案手法は，動画中にアモーダル補完を促す文字列画像を埋め込み，これを解答させるものである．アモーダル補完についての詳細な説明は，3.2 節で述べる．動画には，文字の断片と文字の特徴点から生成された遮蔽物が表示される．動画に表示される文字は，数字とアルファベットからランダムに 4 つ選択されたものが使用される．動画が再生されると，ランダムに配置された文字の断片と遮蔽物が動画中で移動する．動画中に視覚補完が生起するフレームは 4 箇所あり，ユーザは，それぞれのフレームに対して 1 文字ずつ順々に認識していく．ユーザが読み取るフレームには，知覚できる文字が 1 つあり，他は，人間が知覚しづらい文字となっている．4 つの文字が知覚できた時点で，文字を入力させ予め用意した文字と一緒であれば，解答者を人間として判断する．

2.3 節で述べた NuCAPTCHA は，動画の全部の文字を評価するのに数フレームのキャ

プチャで可能である．これに対して本提案手法では，コンピュータが，人間がどのフレームの文字を知覚したかを評価しなければならない．これには，動画中の全フレームに対して知覚しやすさを計算する必要がある．表示される文字が少なければ，総当たり攻撃も可能である．しかし今回の提案では，動画中に表示される文字の総数は，表示時間に比例して大きくなる．そのため 1 枚に 4 つの文字が表示される今回の枠組みで，フレームレートが 5fps，表示時間が 10 秒である場合には，200 文字に対して評価を行う必要がある．人間は，一瞬で文字を知覚し，すぐに解答を入力することができるが，コンピュータは全フレームを評価する時間を要するので，そこに時間の差が生じる．また，コンピュータの表示される文字の断片から元の文字を推測できたとしても，それが人間の知覚したものかどうかを判断しなければ，突破するのが難しいと言える．3.2 節より，アモーダル補完と不完全図形の「知覚しやすさ」についてその要因について整理する．

3.2 アモーダル補完

本手法では，アモーダル補完と呼ばれる人間の視覚補間能力を利用して，ユーザに対し文字を知覚させる．本節では，人間の補完能力について説明し，アモーダル補完の本手法への応用について述べる．

視覚認知おける補完能力とは，外界から目を通じて脳に入ってきた不完全な情報に対して，ある程度の推測から意味を理解することができるというものである．人間は，不完全な物体を認識する時，例えば，物体の輪郭の一部分が他の物体に遮蔽されていた場合にでも，実際にそこにあるかのように補完して知覚することが可能である．実社会の例を挙げると，電柱に遮られた家，草木に隠れた動物，窓枠の向こうに見える自動車などがある．それらは，物体の一部分が見えないにも関わらず，我々は，そこに物体があるということを信じて疑わない．そこには，網膜から得た 2 次元の情報をもとに 3 次元の情報へ補完する高度な視覚系のメカニズムが作用していると考えられている．図 2 の(a)で，ヒトが知覚しうる図形の可能性は，2 つある．1 つは，図 2 (b) のように四角と円がそれぞれ重なりあって見える場合と，図 2 (c) のように分離して見える場合である．しかしヒトが，図 2 (c)のように分離して知覚することは，稀であり，ほとんどがそこに奥行が存在しているものとして補完する．図 2 (a)の例で言えば，四角形の後ろにもうひとつ四角形が重なっているように存在していることである．これは次節，3. 3 で述べるゲシュタルト心理学の中心的概念を担うプレグナンツの法則によれば，遮蔽された物体をできるだけ滑らかに補完するといういい連続の要因が作用するからだと言われている．

図 2(d)は，遮蔽物を背景色の白で塗りつぶしたものである．人間は，図 2 (d)の画像を見てもすぐに文字を認識することは難しいが，図 2(e)の画像を見ると瞬時に知覚することができる．アモーダル補完は，物体と遮蔽物との間に出現する T 字の交点(T-junction)から，遮蔽されたエッジをできるだけ滑らかに曲線として復元するこ

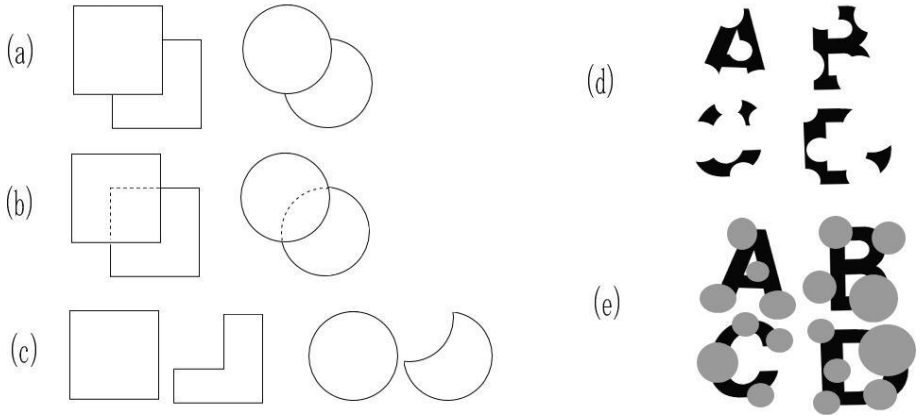

図 2　アモーダル補完の例

で，作用すると考えられる．逆に，図 2 (d)のように，遮蔽物が背景色と同色の場合には，文字が何かに遮蔽されている状況として扱われにくいため補完が起こりにくい．これは，欠損したエッジを補完する場合に補完に必要な情報がどこにあるのかという手掛かりをつかむことができないためである．

Kellman と Shipley[6]は，境界の連続性についてエッジの空間的位置関係によって異なった知覚を与えると述べている．図 3(a)を見ると，エッジの E1 と E2 が楕円形に遮蔽され，その下で連続しているように知覚される．しかし図 3(b)では，連続しているようには，知覚されない．Kellman，Shipley らは，この現象に対して，エッジと遮蔽物の交点が構成する方向に着目し次のようなモデルを提案している．エッジ E1 及び，エッジ E2 の端点から交点 C までの距離をそれぞれ R，r とする．境界補完が成立する条件は，$0 \leqq R\cos\phi < r$ であるという．これは，E1 から E2 の垂線に対して，引いた直線が線分 r の外に行かないことと ϕ が 90° を超えないことを条件にしているということである．この条件下であれば，エッジは遮蔽下で，連続的滑らかな線として補完されるということになる．Kellman，Shipley らは，この制約条件のことを「関係づけの可能性」と呼んでいる．

本提案では，動画中にこのようなアモーダル補完を促すフレームが，タイムライン上に 4 つ用意される．各フレームにおけるユーザに知覚させたい文字は，1 つである．ユーザは，正解を含むフレームを見たときにエッジを滑らかに復元できる形態を瞬間的に知覚し，正解の文字を認識する．

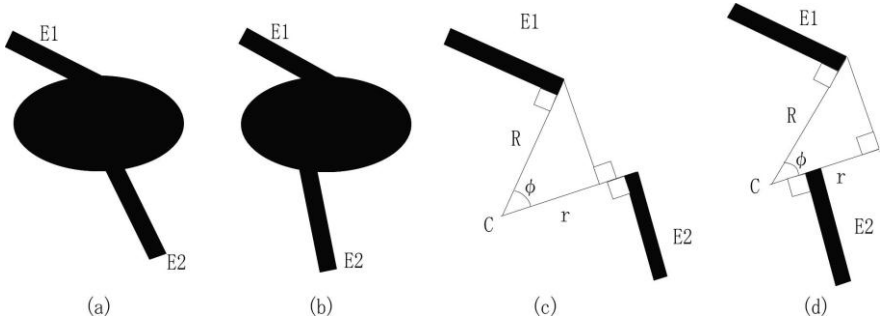

図 3　関係づけの可能性

3. 3　プレグナンツの法則

人が外界の物体を見た場合に，知覚しやすい要因は，3. 2 節で述べた関連付けの可能性以外にも知見がある．それには，ゲシュタルト心理学の中核な概念であるプレグナンツの法則がある[7]．これは，与えられた図形に対して，できるだけ規則的で秩序ある，安定した状態に向かおうとする過程のことである．すなわち，人間は物体を見る時にまとまりの良い，いい形として知覚しようとする．また，複数のまとまりの知覚が可能な場合には,その中で最も簡潔な形を知覚する傾向があると述べられている．

具体的には，次のような要因が働く図形に対して知覚がしやすいとされている．またこれらの要因が働かないものに対しては，知覚がされにくいと考えられる．

(1)　近接の要因

近くにある成分どうしは，まとまりやすいとした要因である．図 4(a)は，黒い円が縦方向に 3 列で並んでいるように見える．

(2)　類同の要因

似ている成分は，まとまりやすいとした要因である．図 4(b)は，円が等間隔で並んでいるが，2 段目の円の色が違うことから，円が横方向に並んでいるように知覚することができる．

(3)　閉合の要因

閉じた図形は，開いた図形よりも知覚しやすいという特性がある．図 4(c)は，向かい合ったかっこの中に四角形があるように知覚できるが，かっこの背面どうしが連続しているようには，知覚しづらい．

(4)　連続の要因

できるだけなめらかな繋がりを持つものは，知覚的にまとまりやすいと言われる．図 4(d)は，円が 2 つ重なっているように見ることができるが，欠けた円が 2 つとラグビーボールのような形が 1 つあるようには，知覚されにくい．

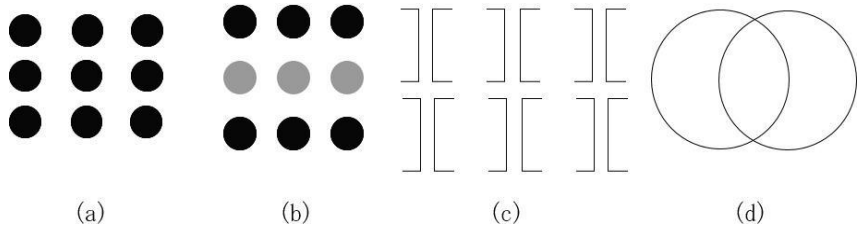

図 4　プレグナンツの法則

3. 2 節で述べた関連付けの可能性や，プレグナンツの法則で述べられる要因が，ユーザに対して知覚のしやすさに影響を及ぼすと考えられる．

4　実装方法

動画 CAPTCHA のフレームは，「背景画像」，「文字が描画されるテキストレイヤー」，「補完を促す遮蔽物が描画されるレイヤー」の 3 つから構成される．ユーザには，これらのレイヤーが埋め込まれた Flash が提示される．Flash に対して埋め込まれるテキストレイヤーは，予め生成されたものからランダムに選択されたものである．表示されるフレームのサイズは，縦横 300px として作成される．また，1 つのフレームは，縦横 150px から成る文字が描画された画像データ 4 枚から構成される．動画の提示時間は，10 秒，フレームレートは，5fps であり，ユーザに対して，提示されるフレームの総数は，50 枚となる．

次節より，実装方法の詳細について言及する．4. 1 節で，提示する文字データの仕様を決定する要因について述べる．次に 4. 2 節で，遮蔽物生成のための特徴点抽出について説明し，4. 3 節では，抽出された特徴点座標より，遮蔽物の描画と文字のエッジに対する動画方法について説明を行う．4. 4 節では，視覚処理における反応速度に対する知見から，フレームレートの設定について述べるとともに，人間の記憶の観点から，動画の提示時間について説明する．

4. 1　文字データの生成とコントラスト

文字データが描画される画像は，PHP の GD ライブラリを利用して作成される．作成された画像データは縦横 150px のサイズである．使用される文字は，大文字アルファベットと数字の 36 文字からランダムに 1 つ選択され描画される．また，使用される文字のフォントサイズは，120pt である．

次に描画する文字と遮蔽物のコントラストについて述べる．補完が行われる場合には，2 つの図形の奥行関係が明瞭である必要がある．図 5(a)は，四角形と三角形が重なっている例である．これらの図形の色相は互いに等質である．この図形を見た場合に，四角形と三角形のどちらかが，手前に置かれているように知覚される．しかしこ

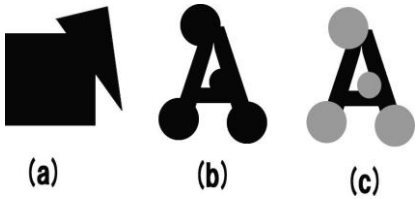

図 5　明るさ，コントラストが奥行に与える影響

の，手前と背景の関係は不確定であり，ある人には正方形が手前に，他の人には，背後に存在するかのように知覚される．また，この関係は，時により意識しなくても注意の仕方によって変化しうるものである．3.2 節において，視覚補完は，図形どうしの奥行関係によって，遮蔽された部分を補完すると述べた．従って，奥行関係を明瞭にさせることは，視覚補完に対して安定した認知を促すと考えられる．図 5(b)を見ると，人間は，アルファベットとそれを覆う円との位置関係が一見して認知しにくい．しかし図 5(c)であれば，遮蔽物が手前に，アルファベットが背景にあるという位置関係が認知しやすい．従って，コントラストの差を設けることが，奥行関係を明瞭にし，安定した認知を促す要因になると考えられる．

本手法で用いられる描画色は，8 ビット(256 諧調)のグレースケールで表現される．提示される各レイヤーのコントラストは，背景画像を白(255)，文字を黒(0)，遮蔽物をグレー(127)として設定される．

4. 2　特徴点抽出

アモーダル補完が起きる要因は，エッジの局所的な働きが大きく作用する．そのため，文字の遮蔽度合が著しく高いものは，補完が起こりにくいとされている．従って，作成される遮蔽物は，文字の輪郭を残す必要がある．本手法での遮蔽物の座標は，文字の特徴点から抽出される．特徴点抽出には，OpenCV ライブラリを用いた．GD ライブラリで作成した図 6(a)の画像に対して canny アルゴリズムを用いて線分化を行い．次に Harris エッジ検出機を適応させた．抽出した特徴点について点を描画した画像が図 6(b)である．次に抽出した特徴点に対して，ランダムに半径を決定し円を描画した例が(c)である．抽出した特徴点と半径の値は，「X 座標，Y 座標，半径」の形式で CSV ファイルとして書き出される．図 6(d)の画像は，図 6(c)の画像と，GD ライブラリで作成した図 6 (a)の画像を合成して，円で囲まれた部分を透過したものである．

生成した図 6 (d)の画像と特徴点座標を書き出した CSV ファイルの組みは，Flash に読み込ませるデータとして用いられる．

4. 3　Flash ファイルの生成

本節では，生成された画像ファイルと CSV ファイルから Flash を生成する方法と動画の手法について述べる．Flash は，予め生成された画像データと組みになった CSV データを 4 つ選び出し，ランダムに配置を行う．配置の仕方は，図 8(a)のようになる．配置が決定したあとは，知覚を起こす順番とタイミングが決定される．タイミングについては，タイムラインからランダムで 4 つ選択される．遮蔽物の描画は，4.2 節で出力された CSV ファイルの情報を用いて行われる．各遮蔽物の X，Y 座標と半径が読み取られ，円が描画される．この円は，フレーム毎にランダムで移動する．円は，動画中で補完画像が表示されるフレームになった時点で，抽出した特徴点の位置へと移動し補完を促す．

次に文字の断片とその動かし方ついての説明を行う．各文字の断片を本稿では，パネルと呼ぶ．各パネルは，「パネルの置き換え」と「ランダムな並行移動」の 2 つで動作をする．フレームは，図 7(a)のように各文字で縦横 50px の分割が行われる．各文字は，150×150 で構成されるので，3×3 のパネルが生成される．文字の断片は，描画される文字のエッジから，元の文字をマッチングさせないための対策として，パネルの移動と入れ替えが行われる．入れ替えられるパネルは，Flash の生成前に問題として選択された文字とは別に，新たに文字を選択し生成されたものが用いられる．Flash は，問題とする文字から予め生成されたパネルを 20 個読み込み，その中からランダムでパネルを入れ替えられる．入れ替えのタイミングは，フレーム毎とし，各文字の領域より 1 または 2 個のパネルを選択して行われる．

図 7(b)は，予め用意された文字のパネルである．フレーム毎にこれらから，全体で 4〜8 枚が選択される．図 7(b)の枠で囲われた部分は，選択されたパネルとして例を挙げたものである．ここでは，例として 5 枚のパネルが選択された場合を示す．図 7 の(b)，(c)，(d)に記述された数字（1〜5）は，それぞれのパネルの位置対応関係を示している．置き換え対象のパネルは，図 7(c)のようにランダムにて選択され，図 7(b)から選択したパネルへ置換される．図 7 (d)は，置換した結果である．知覚文字が表示されるフレームでは，置き換えられたパネルが正規のパネルへ置き換えられるように設定されている．

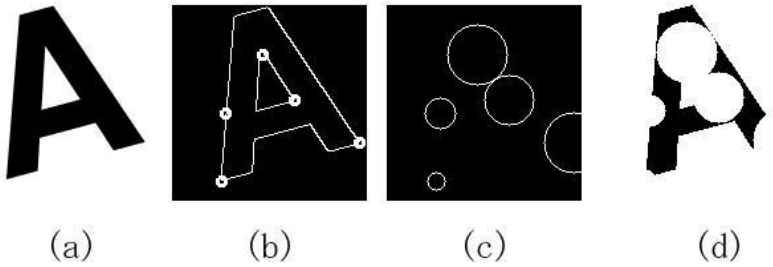

図 6　特徴点の抽出画像

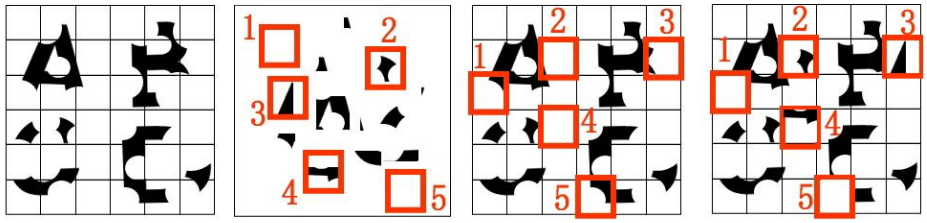

図 7　パネルの置き換え方法

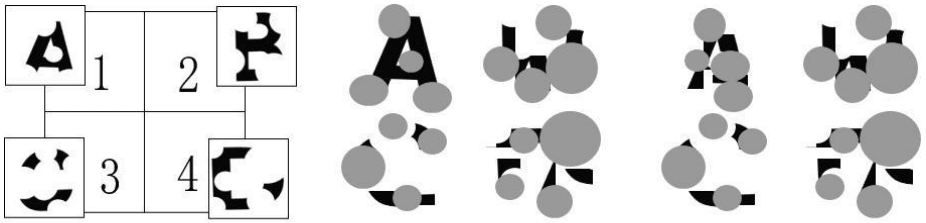

図 8　文字の配置方法と補完画像

4 枚の画像の配置と表示順位について述べる．図 8(a)は，ABCD の文字に対して，表示順位がそれぞれ 1234 として選択された例である．動画では，この順番通りに知覚が促される．図 8(b)は，最初の知覚文字である A が表示された場合の画像である．パネルと遮蔽物の位置が設定値に戻った状態となり，ユーザは一番まとまりのある形として A を知覚することができる．図 8(c)は，知覚文字が表示されないフレームの例である．どの図形もまとまりがなく知覚しにくい状態であるので，正解の文字が表示されていないものとして認識される．

4．4　フレームレートの設定・視覚的補完の処理時間

動画におけるフレームレートは，人間が図形を見て視覚補完が起こり得る値でなければならない．従って，動画を作成するにあたって視覚処理における反応速度や，認知機能の記憶に対して考慮する必要がある．人間の知覚処理における反応速度は，Card らの研究により知見が得られている[8]．Card らによると人間の反応時間が知覚処理，認知処理，運動処理，出力という 4 つの段階に分けて，知覚処理は 100ms，認知処理は 70ms，運動処理は 70ms，出力は 1000ms が必要であると述べられている．このことから，人間は視覚の刺激信号を見てから，その信号を認知するまでの時間は，およそ 170ms である．170ms という反応速度は，どのような対象にも適応されるものではなく，環境や状況によって異なる．今回の場合には，文字の一部分が遮蔽されている場合を問題としている．これらを考慮して，本手法で作成する動画のフレームレートを 5fps(1 フレームあたり 200ms)とした．

フレームレートは，人間の反応時間に対しての上記の知見から設定した．しかし各画像が提示される時間はごくわずかであるため，人間がこれらを想起して解答できるかという問題がある．従って，人間の記憶に対して考慮した動画の長さ(時間)を設定しなければならない．人間の記憶とは，記銘，貯蔵，検索の 3 つの過程からなる情報処理機能のことである．記銘は，経験したことが記憶として取り込まれることを意味し，貯蔵は，取り込まれた情報が保持されることを意味する．想起は，保存した情報を再生することをいう．記憶は，一時的に小さな容量の情報を保持する短期記憶と長期的に大きな容量を保持する長期記憶に分類することができる．人間の認知機能における記憶の保存に対しては，Atkinson や Shiffrin ら[9]が二重貯蔵モデルを説明している．Atkinson や Shiffrin らは，感覚器から入力された情報がまず感覚登録器に一時的に保存され，そこで注意などにより選択された情報が短期貯蔵庫(Short-term Store)に入力され，一定時間保持されるとした．また，これは反復学習などを通じて長期貯蔵庫(Long-Term Store)に入力されるものとした．Sperling[10]の実験では，感覚登録器に保存される感覚記憶(アイコニックメモリー)の持続時間が，500ms だと述べられている．また，感覚登録器に自動的に入力された情報の中で注意を向けられたものに対しては，短期貯蔵庫に格納され，持続時間が 15～30 秒に伸びるとした．本手法ではフレーム内で生起する知覚補完文字がこの注意にあたると考えられる．以上の知見より，本手法の動画の有効解答時間を，30 秒として設定する．

5　評価

福島らの手法[5]により，解析コストの計算を行う．実験環境を，以下の表 1 に示す．実験したところ，1 フレームに表示される図形 4 つの評価に，0.98 秒の解析コストが必要であった．本手法では，提示時間が 10 秒，フレームレートが 5fps である．これらから，動画に対しての解析コストは，提示時間×フレームレート×1 フレームに対しての解析コストで約 50 秒必要となる．動画の有効解答時間は，30 秒に設定してあるので，コンピュータは時間内に解答することができない．このことから，本手法は解析コストの高い動画であると言え，有用性のある CAPTCHA であると考えられる．

表 1　実験環境

OS	Ubuntu10.10
CPU	Intel Core 2 Duo E6850 3.00Ghz
メモリ	4GB

6　考察

本手法では，知覚を想起させるフレームに対して，知覚しやすい図形と知覚しにくい図形の両方を描画することで，曖昧さを生み出し解析コストを高めた．評価では，フレームに表示される各図形の解析にかかる時間のみを解析コストとして測定した．この評価には，コンピュータが，人間が知覚したであろう正解の文字を選択するというコストを含めていない．この問題をコンピュータが解析するためには，曖昧さを解決する手段が必要である．例えば，この曖昧さに対して，図形の見えやすさに着目するのであれば，関連付けの可能性やプレグナンツの法則に対して定量的に評価する必要性がある．関連付けの可能性に対しては，3.2 節のように Kellman, Shipley らが，定量化する手法を提案している．これに加え近年では，この知覚的体制化の法則を定量的に明らかにする試みが登場しており，知覚的体制化のゲシュタルト要因について量的操作と測定が可能であることが示されている．今後は，人間の視覚特性における曖昧性についての解析コストも基準に加えて評価を行っていかなければならない．

7　今後の課題

動画中のフレームにおいて，4 つの文字を知覚する上でどのような情報処理が行われるかを把握することは重要である．ゲシュタルト心理学では，初期視知覚系の初期過程として並列型の処理について述べている[11].並列処理による知覚する文字列は，人間がその中から，読みやすいものを選択するよりも，読みにくいものを判断する方が 2 倍の時間がかかると言われている．今回の手法では，動画中で表示される 50 フレーム中の 46 フレームは,読みにくい文字が描画されていると判断しなければならないため，ユーザに対して大きな負担になっている可能性がある．これに対しては，今後検討の余地がある．

文字認識の速度に対する問題がある．3.2 節にて，人間が文字認識を行う時間として 170ms を要するとしたが，これは，個人差による数値を甘味していない．文字の認識速度に対しては，年齢や環境によって異なることが予測される．また今回は，実験的に文字サイズを 120pt としたが，より認識しやすい文字サイズを検証していかなければならない．

8　まとめ

現在の CAPTCHA が，OCR 技術や文字認識アルゴリズムの向上の他にも，ニューラルネットワークを用いた手法に対して脆弱であることを述べた.これに対して本稿では，画像解析の際に顕著となる人間とコンピュータにおける処理時間の差を，動画を用いることでさらに拡大して実用レベルまで引き上げる手法を提案した．静止画に対しての評価では，人間とコンピュータにおける処理時間には，差が生じた．この差を本手法で述べた動画への応用により，実用レベルまで差を広げることができた．このことから，本手法は CAPTCHA 手法としてその有効性を確認することができたと言える．

参考文献

[1]　G.Mori and J. Malik, “Recognizing objects in adversarial clutter: Breaking avisual CAPTCHA”, Proc.CVPR'03 , pp.134-144 , 2003.

[2] reCAPTCHA http://www.google.com/recaptcha

[3] NuCAPTCHA http://www.nucaptcha.com/

[4] Kumar Chellapilla, Kevin Larson, Patrice Simard, Mary Czerwinski, “Computers beat Humans as Single Character Recognition in Reading based Human Interaction Proofs”, Proceedings of the Second Conference on E-mail and Anti-Spam(CEAS), 2005.

[5] 福島邦彦, “遮蔽されたパターンの認識：神経回路モデル”, 電子情報通信学会技術研究報告. NC, ニューロコンピューティング 99(686), 151-158, 2000-03-15

[6] Kellman, P. J., Shipley, T.F. , “A theory of visual interpolation in object perception. CognitivePsychology”, 23, 141-221 , 1991.

[7] カニッツァ（著), 野口　薫（訳), 視覚の文法 , サイエンス社 , 1987.

[8] S.K. Card, T.P. Moran, and A. Newell: “The Psychology of Human-Computer Interaction”, Lawrence Erlbaum Associates, 1983.

[9] Atkinson, R.C., Shiffrin, R.M. , “Chapter: Human memory: A proposed system and its control processes”, In Spence, K.W.; Spence, J.T.. *The psychology of learning and motivation (Volume 2)*, New York: Academic Press , pp.89-195, 1968

[10]　Sperling, G. “The information available in brief visual presentations. Psychological Monographs“: General and Applied , 74(11, Whole No. 498). pp.1-29 , 1960.

[11] 下條 信輔　(著), 視覚の冒険－イリュージョンから認知科学へ, 産業図書, 1995.