視線を用いたターゲット選択に基づく 動画内物体の注釈情報提示システム

島里恵多† 河野恭之†

概要：本研究では，動画内の物体を視線で選択し，その注釈情報を提示するシステムを提案する．動画を視聴する際，動画に映る物体やキャラクタについての注釈情報を得ることで動画内容の理解が深まる．しかし，一つのシーンに対して複数の物体やキャラクタが存在し，その物体やキャラクタ毎に注釈情報を付与したい場合，注釈情報同士の重畳による注釈情報の可読性の低下や注釈情報の多さによる動画の見易さの低下が生じる．そこで本研究では，動画内の物体を選択しその注釈情報を提示するシステムを提案する．本システムは，動画の一つのシーンに対して複数の物体やキャラクタが存在し，その物体やキャラクタ毎に注釈情報を付与した場合，または付与した注釈情報が単語や短文ではなく長文や画像の場合でも動画の見やすさを保ちつつ注釈情報を提示できる．本システムにおける動画内の物体の選択は，マウスやタッチ操作よりも素早いカーソルの移動と選択が可能な視線を用いて行う．本稿では，提案システムとその評価について述べる．

キーワード：視線インタフェース，ポインティングタスク，グラフィカルユーザインタフェース

1. はじめに

本研究では，視線を用いて動画内の物体を選択し，その注釈情報を提示するシステムを提案する．動画を視聴する際，動画に映る物体やキャラクタについての注釈情報を得ることで動画内容の理解や魅力の向上が考えられる．例えば，近年 COVID-19[1]の影響で普及されつつあるオンライン動物園[2][3]や水族館[4][5]において生物の動画を視聴する際，その動物の種名や生態を知ることでその動画内容の理解が深まる．しかし，近年の動画配信サービスにはこのような動画内に映る物体やキャラクタ毎に注釈情報が提示される機能は少ない．利用者投稿型動画配信サービスであるニコニコ動画[6]は，視聴者が動画にコメントを付与できる機能を提供している．この機能により，異なる場所と時間で動画を視聴していながらも他の視聴者と一緒に動画を観ているかのような感覚を与え，コンテンツの魅力を向上させている．また，YouTube[7]には，動画作成者が動画内にテキストやリンクを組み込む機能があり，より動画視聴の利便性を向上させている．しかし，ニコニコ動画や YouTube のこれらのような機能には，アノテーションによって動画が見にくいという問題点がある．そのため特にオンライン水族館や動物園のような，動画の一つのシーンに対して複数の物体やキャラクタが存在し，その物体やキャラクタ毎にアノテーションを付与したい場合，この問題が顕著に表れることが予想される．また各物体やキャラクタに付与したいアノテーションが単語や短文ではなく長文や画像などの情報が多いものの場合もこの問題が起こり得る．

そこで本研究では，動画内の物体を選択し，その注釈情報を提示するシステムを提案する．動画内の物体全ての注釈情報を提示するのではなく，ユーザにより選択された物体の注釈情報のみを提示することで動画の見やすさを保ちつつ注釈情報を提示できる．本研究の提案システムの開発を目指すにあたり，ユーザが選択するターゲットは動画内の物体である．そのためユーザには移動するターゲットのポインティングタスクが求められる．近年 GUI における一般的なターゲット選択はマウスやタッチ操作で行われる．しかし，移動を伴うターゲットの選択はユーザにとって困難なタスクである[8][9][10]．またユーザの選択したい物体は短いシーンしか映らない場合もある．そのため本システムではより素早いターゲット選択がユーザに求められる．そこで本研究では，マウスやタッチ操作より素早いターゲット選択が可能であると報告[11]のある視線のみを用いて動画内の物体を選択し，その注釈情報を提示するシステムの開発を目指す．

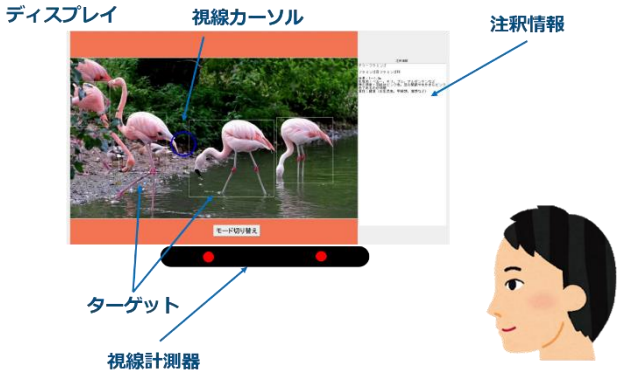

図 1 提案システム

2. 関連研究

2.1 移動するターゲットの選択に関する研究

AttachedShock[12]は，動画内のターゲット幅を拡大し，指でターゲットをストロークすることでターゲットを選択する手法である．Comet[13]は，ターゲットの軌跡を基にター

† 関西学院大学
Kwansei Gakuin University

ゲットを拡大することで移動するターゲットであってもターゲット選択を容易にする手法である．しかし動画内のターゲット数が多い場合，このようなターゲットを拡大させる手法は動画の見やすさを低下させる．また動画内に選択したいターゲットが現れたときそのシーンを一時停止してポインティングタスクを行い，ターゲットを選択する手法[14]が提案されている．この手法は短いシーンしか映らないターゲットも選択できるという利点を持つが，動画の一時停止により視聴する効率性が低下する．本研究で提案するシステムでは動画の見やすさと効率性を保ちつつ移動するターゲットの選択も可能にするシステムの開発を目指す．

2.2 視線を用いたターゲットの選択に関する研究

凝視時間とキーボード入力を組み合わせた選択手法はマウスによる選択より短時間で選択できる[15]．しかし，この手法では視線のみによる操作の利点である素早さを損なってしまう．視線のみを用いたターゲット選択手法として，ターゲットの周りに小さなマーカを円状に移動させてそのマーカに対する追従眼球運動を検出することで選択する手法[16]がある．追従性眼球運動[17]とは，見たいものの動きに合わせて同じ速さで眼球を動かす運動である．また，ユーザ及びターゲットごとに動的に凝視時間を最適化して選択する手法[18]が提案されている．崔ら[19]は，視線カーソルにバブルカーソル[20]を用いてターゲットを選択する手法を提案した．バブルカーソルとは，カーソルが常に一つのターゲットを含むように大きさを動的に変更し続ける円形のカーソルである．しかしこれらの手法は静止したターゲットの選択を想定した手法である．本研究においてユーザは動画内の物体の選択が求められる．本研究では視線を用いて移動するターゲットの選択手法を用いる．

Pursuits[21]は，ピアソンの積率相関係数を用いてディスプレイ上における視線移動の軌跡と移動するターゲットの軌跡との間に相関関係があるときターゲットを選択する手法である．またカメラ画像からユーザの視線の移動を推定し追従性眼球運動を検出する手法[22]も提案されている．しかし，このような追従性眼球運動に基づいてターゲットを選択する手法は複雑な軌道を伴うターゲットの場合，またターゲットの移動速度が遅いもしくは速い場合にターゲット選択の精度が低くなる．そこで本研究では，ターゲットの軌跡や速さに関わらず高い精度でターゲット選択が行えるシステムの開発を目指す．

2.3 動画のアノテーションに関する研究

浦谷ら[23]は，奥行きを考慮した AR 上の注釈情報の提示手法を提案した．飛田ら[24]は，技能学習用のマニュアル動画において再生速度や再生時間に基づいて注釈情報の詳細度合いを調整し提示する手法を提案した．しかし，動画の一つのシーンに対して複数の物体やキャラクタが存在しその物体やキャラクタ毎に注釈情報を付与したい場合，物体や注釈情報同士の重畳が生じる．これにより注釈情報の可読性の低下や動画が見にくくなるという問題が起こる．このような問題を解決するために注釈情報を再配置する手法[25]が提案されている．しかし，この手法は画面上において注釈情報の提示領域が十分あることが必要である．そのため付与したい注釈情報が長文である場合や画像などの場合においても物体や注釈情報の重畳が生じやすい．そこで本研究では，ユーザが注釈情報を取得したい物体を選択し，ユーザにとって見たい注釈情報のみを GUI 上の注釈情報提示用スペースに提示することでこの問題を解決する．

3. 視線を用いたターゲット選択に基づく動画内物体の注釈情報提示システム

本研究で提案するシステムについて述べる．本研究では，素早いポインティングタスクが可能である視線を用いて動画内の物体を選択し，その注釈情報を提示するシステムを実装した．本システムでは画面上に図 1 のようなディスプレイ右端にテキストボックスを用意し，ここに注釈情報を提示する．そのため本システムを用いることでユーザは注釈情報を取得したい物体の注釈情報のみを取得できる．故に本システムは動画の一つのシーンに対して複数の物体やキャラクタが存在し，その物体やキャラクタ毎に注釈情報を付与した場合，または付与した注釈情報が単語や短文ではなく長文や画像であっても注釈情報の重畳は起こらず，動画の見やすさは保たれる．また本システムはマウスやタッチ操作より素早いターゲット選択が可能であると報告のある視線のみを用いて動画内の物体を選択する．これまで視線を用いた様々なターゲット選択手法が研究されてきた．本システムのターゲット選択には，これまで我々が研究を行ってきた，ユーザのサッケード（跳躍性眼球運動）[17]を検出したとき移動しているターゲットをある間だけ疑似的に静止させ，その間に疑似的に静止しているターゲットを視線のみで選択する手法[26]を用いる．移動するターゲットの選択はユーザにとって難しいタスクである．しかしこの手法を用いることでターゲットの軌道の複雑さや速さに関係なく容易にターゲットを選択できる．

本システムの全体の処理の流れについて述べる．まず前処理では，動画内の物体を検出・追跡し，各追跡物体に対して注釈情報を登録する．そして本処理ではユーザの視線を検出し，ユーザが前処理で注釈情報を登録したターゲットに対して視線を用いて選択した場合その注釈情報を提示する．本システムの視線を用いたターゲット選択は，ユーザのサッケードを検出したとき移動しているターゲットをある間だけ疑似的に静止させ，その間に疑似的に静止しているターゲットを視線のみで選択する．サッケードとは，多くの物から見たい物を見つけるための飛び石を渡るような

素早い眼球運動である.そしてサッケードが生じている間,視覚的な認知能力が低下するサッケード抑制という特性[27]がある．この特性からサッケードが起こる際における視覚的な情報の提示は,ほとんど意味がないと考えられる.そこで，サッケードを検出した際に視線移動に必要な時間だけ動画内のターゲットを疑似的に静止させ，ユーザは疑似的に静止しているターゲットに対してポインティングタスクを行うことで移動するターゲットの選択を容易にする.

3.1 注釈情報の登録

本システムにおける前処理の具体的な処理について述べる．まず前処理として動画ファイルを入力し物体の検出を行う．本システムは，深層学習に基づく物体検出技術である YOLOv4[28]を用いて物体検出を行った．また本システムにおいてサッケードを検出すると同時にターゲットを疑似的に静止させる際，物体検出の精度が原因でサッケードを検出したときのフレームでターゲットを検出していない可能性がある．そこで本研究では，物体検出した後にカルマンフィルターを用いた SORT[29]というアルゴリズムに深層学習を組み込んだ DeepSort[30]というフレームワークを用いてその物体を追跡する．そして各動画フレームに対して追跡した物体バウンディングボックスの座標を記録する．また追跡した物体毎にラベルを割り当て，各ラベルに提示したい注釈情報を登録する.

3.2 ユーザインタフェース

本システムのユーザインタフェースについて述べる．本システムは，ユーザは前処理で入力された動画を視聴すると同時にディスプレイ上におけるユーザの視線座標を検出する．この際，ヒトは静止物体を凝視しているつもりでも実際には細かい眼の揺れが生じてしまう固視微動という特徴[17]がある．本システムではユーザに視線カーソルを提示するため加重平均の式による視線座標の平滑化[31]を行った．また動画閲覧の際，選択可能なターゲットにはバウンディングボックスを提示する．そしてサッケードを検出した際にターゲットを疑似的に静止させ，ユーザは疑似的に静止したターゲットに対して視線カーソルを移動させ，ポインティングタスクを行う．本システムでは，サッケードを検出した際にその動画フレームに対応する前処理で記録した検出物体のバウンディングボックスの座標を呼び出し，ターゲットを疑似的に静止させる．ユーザがターゲットを選択した場合，選択した追跡ターゲットのラベルを特定し，このラベルに対応する注釈情報を図 1 のように提示する．また図 1 の「モード切り替え」のボタンについて述べる．ユーザはこのボタンを一定時間注視すると視線カーソルが表示され，ターゲット選択が可能になる．またターゲット選択が可能な状態でこのボタンを一定時間注視すると視線カーソルは非表示になり，ターゲット選択を行わない．この機能により，単に動画を見たいユーザに対してはサッケードの検出を行わず，ターゲットの疑似的な静止が起こらないようにした.

3.3 視線を用いたターゲット選択

図 2　視線を用いたターゲット選択手法

本システムにおけるターゲット選択手法について述べる．本システムにおけるターゲット選択の概要図を図 2 に示す．本システムでは，サッケードを検出した際に視線移動に必要な時間だけターゲットを疑似的に静止させ，ユーザは疑似的に静止しているターゲットに対してポインティングタスクを行うことでターゲットを選択する．本システムでは，取得した最も新しい視線座標とそれ以前に取得したいくつかの視線座標とのユークリッド距離がすべてある閾値以上のときサッケードを検出する手法[31]を用いた.

そしてサッケードを検出したときのフレームに映る動画内のターゲットを疑似的に静止させる．このとき疑似的に静止したターゲット以外の視覚情報をグレースケール化することでユーザに目立たないように加工して動画を再生する．またユーザの視覚的認知能力に対する影響を最小限にするために，ターゲットを疑似的に静止させる時間はポインティングタスクに必要最低限な時間に設定する．本研究では，ヒトのポインティングタスクのモデルであるフィッツの法則[32]を用いてターゲットを疑似的に静止させる時間を定めた．フィッツの法則は，ポインティングタスクの開始点からターゲットまでの距離とターゲット幅からポインティングタスクにかかる時間を推定できる．D をポインティングタスクの開始点からターゲットの中心までの距離，W をターゲット幅，a と b をユーザとデバイスに依存する定数とおくとポインティングタスクの開始点からターゲットまでの移動に要する時間 MT は$MT = a + blog2(DW / +1)$で表される．本研究では，サッケードを検出したときに用いた視線座標のうち最も新しい視線座標以外のいくつかの視線座標の重心をポインティングタスクの開始点と置く．またサッケードの検出に使用した２番目の視線座標をポインティングタスクの開始時間に設定する．また本研究ではYOLOv4 で検出した物体のバウンディングボックスをターゲット幅に設定した．そしてユーザは，疑似的に静止しているターゲットを一定時間凝視することで選択を行う．ここでもユーザの視覚的認知能力に対する影響を最小限にするために磯本らの手法[33]を用いてターゲット選択のための凝視時間を低減させる．この手法では，フィッツの法則によって推定したポインティングタスクにかかる時間と実際にかかったポインティングタスクの時間の差が十分小さくなったときターゲットを選択する．しかし，ヒトは物体を視覚から認識する際に物体の輪郭を見るという特徴[34]により，ユーザはターゲット選択の際にバウンディングボックスの枠を見てしまう．そこで本研究では，上記の処理に Bubble Gaze Cursor[19]を組み込んだ．この手法を用いることで，ターゲット選択の当たり判定においてターゲットが選択される領域がターゲット幅ではなくボロノイ領域に増え（図 3），ユーザはバウンディングボックスを見てしまった場合でも選択しやすくなると考えられる．また実際のユーザインタフェースでは図 3 のようなボロノイ分割の図をユーザに提示しないよう実装した．

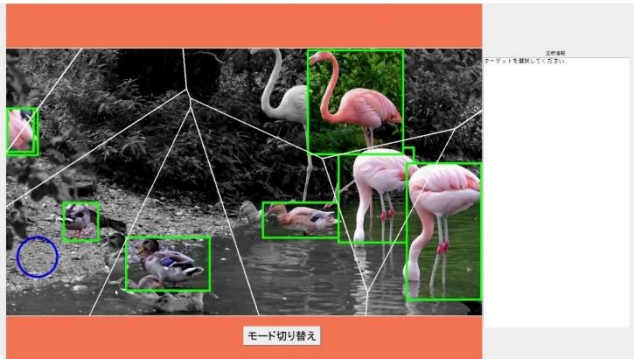

図 3　ターゲット選択領域

4. 評価実験

提案システムの性能を評価するための実験とその結果，考察について述べる．本実験の調査内容は，本システムに組み込んだターゲット選択手法の性能とユーザビリティである．

4.1 実験方法

本システムにおけるターゲット選択の性能を評価する実験では，実験参加者には本システムを用いてターゲットの移動する速さ，移動方向，ターゲット幅，ターゲット数を考慮したポインティングタスクを行ってもらった．本実験で使用したターゲットの移動する速さ，移動方向，ターゲット幅を以下にまとめる．またこのタスクのターゲットの配置例を図 4 に示す．

- 速さ[inch/ms]: 2.50×10^{-2}, 5.00×10^{-2}, 7.50×10^{-2}, 7.50×10^{-2}, 1.00×10^{-1}, 1.25×10^{-1}, 1.50×10^{-1}
- 移動方向: 縦，横
- ターゲット幅[inch]: 0.500, 0.750, 1.00, 1.25, 1.50
- ターゲット数[個]: 3, 4, 5, 6, 7, 8

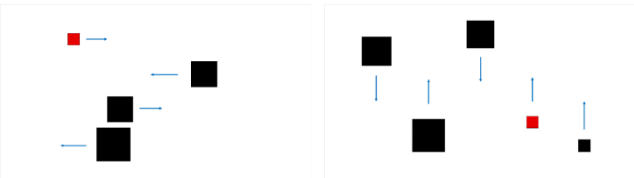

図 4　本実験で用いた移動するターゲット例．
（左）横方向の移動．（右）縦方向の移動．

このタスクでは図 4 のようにターゲットを配置し，赤いターゲットを選択することが求められる．実験参加者が赤いターゲットを選択するとディスプレイ上のすべてのターゲットの幅と速さが変わり，一つだけ赤いターゲットが提示される．そして実験参加者は再び赤いターゲットを選択する．またこのタスクではターゲットのオクルージョンを考慮して，ディスプレイの幅をターゲットの数で均等に分割するように配置した．例えば図 4 の左の図では，ディスプ

レイ上を縦方向に 4 分割し，各ターゲットはその分割した領域のみを移動する．またターゲットがディスプレイの端に到達した場合，その反対側のディスプレイの端に提示させた．このタスクでは実験協力者一人当たり，ターゲットの速さ 6 種×ターゲットの移動方向 4 種×ターゲット幅 5 種×ターゲット数×6 種 ＝720 試行を行った．

本システムにおけるターゲット選択の性能評価実験が終了後，本システムを使用した際のユーザビリティを調査する実験を行った．前処理ではこちらであらかじめ注釈情報を登録しておき，実験参加者には本システムにおける本処理の部分のみを自由に使用してもらった．本実験ではオンライン水族館やオンライン動物園を想定し，2 種類の鳥が歩き回る動画を使用した．次に 5 段階のリッカート尺度の System Usability Scale(SUS)[35]を用いたアンケートと自由記述アンケートを行った．その後 5 段階のリッカート尺度を用いて，サッケード時のターゲットの疑似的な静止の処理に対するストレスの程度を調査した．

4.2 実験環境

本実験では計 7 名の実験参加者を集めた．そのうち 6 名は男性で 1 名は女性であった．また実験参加者の平均年齢は 22.1 歳（標準偏差：1.35）であった．実験参加者の眼の状態は 3 名が裸眼，3 名がコンタクトであり，1 名が眼鏡を装着していた．また実験参加者全員が普段の眼の状態と同じ状態で実験に参加した．実験参加のうち 3 名が視線認識デバイスの使用歴があった．実験参加者の全員の平均使用歴は 7.71 カ月（参加者全員の標準偏差：10.5，視線認識デバイスの使用歴がある参加者の標準偏差: 8.49）であった．実験参加者の平均睡眠時間は 7.00 時間（標準偏差：1.41）であった．実験参加者の眼と身体の疲労度に関して，1 が「疲れていない」，5 が「疲れている」とした 5 段階のリッカー度尺度を用いたアンケートを実施した．結果，目の疲労度が 2.14（標準偏差：0.350），身体の疲労度が 1.57（標準偏差：0.495）であった．

本システムは Python の GUI 構築用ライブラリである Tkinter を用いて実装した．またユーザの視線検出の際には Tobii 社の Tobii Eye Tracker 4C を使用した．本システムの実装環境は Intel(R) Core i7-8750H CPU と NVIDIA GeForce GTX 1060 の GPU を搭載した PC を使用した．本実験で用いた PC のディスプレイは 19.5[cm] × 34.5[cm]のサイズで 1440[pix]×810[pix]のものであった．また実験参加者にはディスプレイから約 50[cm]離れた位置に座ってもらい，実験を行った．また本実験では，実験参加者毎に視線計測機器のキャリブレーションを行った．次にフィッツの法則におけるユーザとデバイスに依存する定数 *a, b* を求めた．本研究では ISO 9241-411 に記載されてあるマルチディレクショナルポインティングタスク[36]を用いた．このタスクでは凝視によってターゲットを選択するのに最適な凝視時間である 400[ms]以上凝視することでターゲットを選択するよう設定した．このタスクでは，0.500[inch]，0.750[inch]，1.00[inch]の 3 種のターゲット幅を使用した．また，1.50[inch]，2.00[inch]，2.50[inch]の 3 種のターゲット間距離を使用した．そのためこのタスクでは実験協力者一人当たり，ターゲット間距離 3 種×ターゲット幅 3 種×ターゲット数 13 個＝117 試行行われた．このタスクの終了後，フィッツの法則におけるターゲットの選択困難度を示す $ID = log2(DW+1)$ と実際にそのターゲットを選択するのにかかった時間をそれぞれ（x, y）の二次元座標にプロットした．最後にこの二次元座標データに対して回帰分析を行い，求めた一次式の傾きと切片からそれぞれフィッツの法則の係数を求めた．

4.3 実験結果

移動するターゲットを選択するタスクでは，ミダスタッチ[37]の割合と赤いターゲットが提示されてからターゲットを選択するのにかかった時間を測定した．ミダスタッチの割合の算出方法を式 1 に示す．

$$ミダスタッチ割合 = \frac{ミダスタッチ回数}{ミダスタッチ回数 + 選択回数} \quad (式1)$$

本実験における全体のミダスタッチの割合は 4.54%であった．またターゲット数毎のミダスタッチの割合を表 1 に示す．また Shapiro-Wilk の検定を行ったところ正規性は見られなかった．($p = 1 \times 10^3$）そこで Steel-Dwass 検定を行ったところ $p < 0.05$，$p < 0.01$ ともにどのターゲット数の間でも有意差は見られなかった．

表 1　ミダスタッチ割合

ターゲット数	3	4	5	6	7	8
ミダスタッチ[%]	2.21	3.56	3.23	5.00	6.15	6.98

またターゲット幅とターゲットの速さそれぞれに対して，実験参加者が選択するターゲットが提示されてから選択されるまでにかかった時間をそれぞれ表 2, 表 3 に示す．Tukey 法による多重比較をおこなったところ，表 3 では有意水準 $p < 0.05$，$p < 0.01$ ともに有意差は見られなかった．表 4 では，2.50×10^{-2} [inch/ms]とそれ以外のすべての速さに対してのみ $p < 0.05$ において有意差が見られた．

SUS は表 4 の質問 10 問で構成されており，1 が「そう思わない」，5 が「そう思う」としたリッカート尺度のアンケート結果から 100 点満点のスコアを算出する．スコアは，$SUSスコア = \{(奇数番号の質問結果の合計 - 5) + (25 - 偶数番号の質問結果の合計)\} \times 2.5$で求められる．

表 2　ターゲット幅×選択にかかった時間

ターゲット幅 [inch]	平均[ms]	標準偏差
0.500	1.02×10^{3}	6.78×10^{2}
0.750	1.02×10^{3}	6.54×10^{2}
1.00	1.08×10^{3}	7.40×10^{2}
1.25	1.09×10^{3}	9.38×10^{2}
1.50	1.08×10^{3}	8.14×10^{2}

表 3　ターゲットの速さ×選択にかかった時間

速さ[inch/ms]	平均時間[ms]	標準偏差
2.50×10^{-2}	1.23×10^{3}	1.23×10^{3}
5.00×10^{-2}	1.06×10^{3}	6.79×10^{2}
7.50×10^{-2}	9.70×10^{2}	5.64×10^{2}
1.00×10^{-1}	1.01×10^{3}	6.40×10^{3}
1.25×10^{-1}	1.02×10^{3}	6.04×10^{2}
1.50×10^{-1}	1.06×10^{3}	7.16×10^{2}

表 4　SUS アンケート結果

	質問内容	平均	標準偏差
1	このシステムをしばしば使いたいと思う	3.29	0.881
2	このシステム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	2.29	0.700
3	このシステムは容易に使いこなす事ができると思った	3.57	1.05
4	このシステムを利用するのに専門家のサポートが必要だと感じる	2.57	0.728
5	このシステムは十分に統一感があると感じた	3.71	0.700
6	このシステムでは一貫性のないところが多々あったと感じた	2.43	1.18
7	たいていの人は、このシステムの利用方法をすぐに理解すると思う	3.00	1.07
8	このシステムはとても操作しづらいと感じた	2.43	1.40
9	このシステムを利用できる自信がある	3.86	1.25
10	このシステム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	3.57	0.728
SUS スコア		60.4	14.4

5 段階のリッカート尺度を用いてサッケード時のターゲットの疑似的な静止の処理に対するストレスの程度を調査した.「ターゲットの疑似的な静止処理にストレスを感じた」という質問に対して 1 が「そう思わない」, 5 が「そう思う」とした．結果，平均 2.29（標準偏差：0.700）であった．

自由記述アンケートで得られた意見をまとめる．実験環境における意見として,「実際に見ている位置と違う位置にカーソルが現れた」や「視線カーソルの表示にラグを感じた」という意見があった．注釈情報の提示方法に関する意見として「注釈情報を提示する領域から遠い位置でターゲットを選択してから注釈情報を見るのが疲れる」等の意見が得られた.

4.4 考察

本実験において，3〜8 個のターゲットが一度に提示されている場合であっても低い確率でミダスタッチが生じる．そのため本システムのターゲット選択手法は，一度のシーンに複数のターゲットが存在する場合でもミダスタッチを起こさずにターゲットが選択できると考えられる．また全てのターゲット幅と速さに対して，ユーザの選択したいターゲットが現れてから約 1000[ms]で選択できた．この結果から本手法のターゲット選択手法を用いることで，一般的に選択することが難しいとされる短いシーンしか映らないターゲットや移動の速いターゲットも十分選択可能であると考えられる．しかし，最も移動する速さの低いターゲットは選択するのに時間がかかるという結果が得られた．この原因は, サッケード検出の精度によるものと考えられる．選択したいターゲットが現れた時点のターゲットと視線の位置が近い場合，実験参加者はサッケードを起こさずそのまま追従性眼球運動をしたためサッケードが検出されずターゲットの疑似的な静止処理が起こらなかった.

SUS のスコアは，標準平均の 68 より下回った．これは自由記述アンケートで得られた「実際に見ている位置と違う位置にカーソルが現れた」や「視線カーソルの表示にラグを感じた」という意見が得られたことから，視線認識デバイスの精度とシステムの実行環境がユーザにとって満足するほど十分でなかったと考えられる．また，サッケードはユーザにとって意図的な眼球動作ではない．そのため高い精度でターゲットの選択が可能であるにも関わらず，疑似的な静止処理が起こるタイミングが予測できないことが SUS スコアの低さの原因の一つであると考えられる．しかし，サッケード時のターゲットの疑似的な静止の処理に対するストレスは比較的低い結果が得られた．このことから移動するターゲットを疑似的に静止させても問題ないと考えられる．また「注釈情報を提示する領域から遠い位置でターゲットを選択してから注釈情報を見るのが疲れる」等の意見が得られたからより大きなディスプレイで本システムを使用する場合，ユーザにとってはより疲労感を与える

と考えられる．また本実験では図 1 のように画面の左端に注釈情報を提示する領域を設置したが，この意見から注釈情報の提示デザインを再検討する必要があると考えられる．

5. 課題

本システムにおける前処理の課題について述べる．前処理では深層学習に基づく物体検出技術である YOLOv4 を用いて物体検出を行った. また YOLOv4 を用いて物体を検出した後に，DeepSort というフレームワークを用いて物体追跡を行った．しかしこの手法では，物体を検出する際に学習データセットが必要となる．そのため本システムにおいて，学習データセットのない物体に対しては注釈情報を登録できず，またターゲット選択も行えない．この解決策として近年研究が進んでいる完全教師無し学習による物体追跡技術が挙げられる．また本システムの前処理では動画内の各追跡物体に対して注釈情報を登録するが，長時間の動画の場合注釈情報を登録するタスクは手間である．そのため検出した各追跡物体に対して効率的な注釈情報の登録を可能にするユーザインタフェースの実装が考えられる．

本システムにおけるユーザビリティの課題について述べる．本システムを用いてストーリー性のある動画を視聴する場合，本システムの疑似的な静止を行うターゲット選択処理がユーザのストーリーの理解に影響を及ぼす可能性がある．そのためストーリー性の動画に対する本システムの有効性も調査する必要がある．またサッケードは意図的な眼球運動ではない．自由記述アンケートでは得られなかったが，サッケード時にターゲットを疑似的に静止する処理が予測できない点は，ユーザが使いにくいと感じる原因の一つであると考えられる．また本システムではフィッツの法則による推定時間だけターゲットを疑似的に静止させた．しかし，視線を用いたターゲット選択の場合フィッツの法則への適合を疑問視している報告もある[38]．本実験では，サッケード時のターゲットの疑似的な静止の処理に対するストレスは比較的低い結果が得られたが，最適なモデルを使用することでよりストレスの軽減やユーザビリティの向上が期待できる．

6. おわりに

本研究では視線を用いて動画内の物体を選択し，その注釈情報を提示するシステムを提案し, 実装した. ユーザは，サッケードを検出したとき移動しているターゲットをある間だけ疑似的に静止させ，その間に疑似的に静止しているターゲットを視線のみで動画内の物体を選択する．そして本システムでは画面上に注釈情報を提示するためのテキストボックスを用意し，ここに注釈情報を提示する．このようにすることでユーザは注釈情報を取得したい物体の注釈情報のみを取得できる．故に本システムは動画の一つのシーンに対して複数の物体やキャラクタが存在し，その物体やキャラクタ毎に注釈情報を付与した場合，または付与した注釈情報が単語や短文ではなく長文や画像であっても注釈情報の重畳は起こらず，動画の見やすさは保たれる．また提案システムは，一般的に難しいとされる移動するターゲットの選択をターゲットの幅や速さに関わらず容易にする．評価実験の結果，様々なターゲット幅や速さ，また一つのシーンに対して存在するターゲットの数に対して高い精度でターゲットを選択できたが，ユーザビリティの面においては低い評価が得られた．

謝辞　本研究の全過程を通じて，担当教授の関西学院大学理工学部人間システム工学科　河野恭之　教授には熱心なご指導ご鞭撻を賜り，常日頃より多大なる助言をいただきました．ここに感謝の意を表します．

本研究へのご意見，ご助言を賜りました河野研究室所属の同研究室のメンバーに心より感謝いたします．研究面のみならず日常生活においても，長きに渡り支えて頂きました．ここに感謝の意を表します．

参考文献

[1] “患者教育：新型コロナウイルス感染症（COVID-19）の概要（簡易）”. https://www.uptodate.com/contents/1126696, (参照 2022-1-08).
[2] “うごく！どうぶつ図鑑”. https://www.tokyo-zoo.net/movie/mov_book/index.html, (参照 2022-1-08).
[3] “千葉市動物公園”. https://www.youtube.com/channel/UCk9Huw8uFABc1qRhp20uU9w, (参照 2022-1-08).
[4] “Sunshine aquarium オンライン水族館”. https://sunshinecity.jp/aquarium/information/online.html, (参照 2022-1-08).
[5] “PORT OF NAGOYA PUBLIC AQUARIUMAPARTMENT ～名古屋港水族館ライブ supported by CTC”. https://csr.ctc.co.jp/aqua/index.html, (参照 2022-1-08).
[6] “ニコニコ動画”. https://www.nicovideo.jp/, (参照 2022-1-08).
[7] “YouTube”. https://www.youtube.com/, (参照 2022-1-08).
[8] Hoffmann, E. R.. Capture of Moving Targets: a Modification of Fitts’ Law. Ergonomics. 1991, vol. 34, no. 2, p. 211–220.
[9] Jagacinski, R J., Daniel W. R., Sharon L. W., and Martin S. M.. A test of Fitts’ law with moving targets. Human Factors. 1980, vol 22, no. 2, p. 225-233.
[10] Tresilian, J. R. and Lonergan, A.. Intercepting moving objects: effect of temporal precision and movement amplitude. Experimental Brain Research. 2002, vol. 142, p. 193-207.
[11] Sibert, L. E. and Jacob, R. J. K.. Evaluation of Eye Gaze Interaction. CHI'00. 2000, p.281-288.
[12] Chuang, W. Y., Yung, H. H. and Wen, H. C.. AttachedShock: Facilitating Moving Targets Acquisition on Augmented Reality Devices using Goal-crossing Actions. MM’12. 2012, p. 1141-1144.
[13] Khalad, H., Tovi, G. and Pourang, I.. Comet and Target Ghost: Techniques for Selecting Moving Targets. CHI’11. 2011, p. 839-843.
[14] Hajr, A. H., Fels, S., Miller, G. and Ilich, M.. Moving Target Selection in 2D Graphical User Interfaces. INTERACT2011. 2011, p.141-161.
[15] Hild, J., Perterson, P. and Beyerer, J.. Moving Target Acquisition by Gaze Pointing and Button Press using Hand or Foot. ETRA’16. 2016, p. 257-260.
[16] Esteves, A., Velloso, E., Bulling A. and Gellersen, H.. Orbits:

Gaze Interaction for Smart Watches using Smooth Pursuit Eye Movements. UIST'15. 2015, p. 457-466.
[17] 鵜飼一彦. 眼球運動の種類とその測定. 日本光学会. 1994, vol. 23, no. 1, p. 2-8.
[18] Nayyar, A., Utkarsh, D., Ahuja, K., Rajput, N., Nagar, S. and Dey, K.. OptiDwell: Intelligent Adjustment of Dwell Click Time. IUI'17. 2017, p. 193-204.
[19] 崔明根, 坂本大介, 小野哲雄. Bubble Gaze Cursor：バブルカーソル法の視線操作への適用. 情報処理学会論文誌. 2020, vol. 61, no. 2, p. 221-232.
[20] Grossman, T. and Balakrishnan, R..The Buble cursor: enhancing target acquisition by dynamic resizing of the cursor' activation area. CHI'05. 2005, p. 281-290.
[21] Vidal, M., Bulling, A. and Gellersen, H.. Pursuits: Spontaneous Interaction with Displays based on Smooth Pursuit Eye Movement and Moving Targets. UbiComp'13. 2013, p. 439-448.
[22] Bâce, M., Becker, V., Wang, C. and Bulling, A.. Combining Gaze Estimation and Optical Flow for Pursuits Interaction. ETRA'20. 2020, Article No. 2.
[23] 浦谷謙吾, 町田貴史, 清川清, 竹村治雄. 拡張現実環境における奥行きを考慮した注釈提示手法の評価. 日本バーチャルリアリティ学会論文誌. 2005, vol. 10, no. 3, p. 343-352.
[24] 飛田優, 長松隆, 鎌原淳三, 石井裕. 技能習得ビデオにおけるユーザーの操作に適応したリアルタイム・アノテーション提示. 第 4 回データ工学と情報マネジメントに関するフォーラム. 2012, A8-2.
[25] Azuma, R. and Furmanski, C.. Evaluating Label Placement for Augmented Reality View Management. ISMAR'03. 2003, p. 66-75.
[26] Shimasato K. and Kono Y.. Gaze-Basedd Moving Target Acquisition, Pseudo Stopping for the Time predicted via Fitts' Law. AVI'20. 2020, article no. 84.
[27] Bridgeman, B., Hendry, D. and Stark, L.. Failure to detect displacement of the visual world during saccadic eye movements. Vision Research. 1975, vol. 15, no. 6, p. 719-722.
[28] Bochkovskiy, A., Wang, C. Y. and Liao, H. Y. M.. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint. 2020, arXiv: 2004.10934.
[29] Bewley, A., Ge, Z., Ott., L and Upcroft, B.. Simple Online and Realtime Tracking. arXiv preprint. 2016, arXiv: 1602.00763.
[30] Wojke, N., Bewley, A. and Paulus, D.. Simple Online and Realtime Tracking with a Deep Association Metric. arXiv preprint. 2017, arXiv: 1703.07402.
[31] Manu K., Jeff K., Rohan P., Terry W., and Andreas P.. Improving the Accuracy of Gaze Input for Interaction. ETRA'08. 2008 p. 65-68.
[32] Fitts, P. M.. The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. Experimental Psychology. 1954, vol.74, p. 381-391.
[33] Isomoto, T., Ando, T., Shizuki, B. and Takahashi, S.. Dwell Time Reduction Technique using Fitts' Law for Gaze-Based Target Acquisition. ETRA'18. 2018, article no. 26.
[34] 仲田仁, 田村仁. 視線追跡を用いた騙し絵の認識の解析. 情報処理学会第 78 回全国大会講演論文集. 2016, p. 307-308.
[35] Brooke, J.. SUS - A quick and dirty usability scale. Usability Evaluation in Industry. 1996, vol. 189, no. 194, p. 4–7.
[36] International Organization for Standardization. Ergonomics of human-system interaction – Part 411: Evaluation methods for the design of physical input devices. ISO ISO/TS 9241-411:2012.
[37] Jacob, R. K. J.. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At is What You Get. ACM Transactions on Information Systems. 1991, vol. 9, no. 2, p.152-169.
[38] Schuetz, I., Murdison, T. S., Mackenzie, K. J. and Zannoli, M. An Explanation of Fitts' Law-like Performance in Gaze-Based Selection Tasks Using a Psychophysics Approach. CHI'19. 2019, article no 535.