Vol.2017-CG-168 No.23
Vol.2017-DCC-17 No.23
Vol.2017-CVIM-209 No.23
2017/11/8

視線の隠的追跡手法とそのエンタテイメント応用

清水 文也[1,a)]　藤代 一成[1,b)]

概要：デジタルサイネージやヘッドマウントディスプレイなど, さまざまな日常生活レベルのデバイス操作のインタフェースとして視線追跡を利用する手法が数多く研究されている．これは, 視線を追跡することでユーザの意図や思考を効率的に反映することができるからである．しかし専用デバイスを用いて視線を追跡する場合, 各ユーザに対して事前にキャリブレーションを行う必要がある．本研究では, 一般的なカメラを 1 台だけ用いて, 事前にキャリブレーションを必要としない，陰的視線追跡の手法を開発することで, ユーザが対象物に集中しやすい環境を作り出し, 臨場感を提供することを目的とする．さらに, その応用として, ユーザの視線方向によって再生するオーディオデータを選択可能にする手法を提案し, 美術館や博物館のオーディオガイドへの適用を前提とした予備実験を行う．

キーワード：視線追跡, オーディオインタフェース, ヒューマンインタラクション

Hidden Eye-Tracking Technologies with its Application to Entertainment

Fumiya Shimizu[1,a)]　Issei Fujishiro[1,b)]

Abstract: Many methods using gaze tracking have been proposed for realizing everyday interfaces for such devices as digital signage and HMD. This is because gaze expeditiously reflects the intention and thoughts of the veiwer. When using eye-tracking devices, we need to calibrate those devices for each viewer. We herein present a calibration-free method which allows the viewer to select the audio data through the detection of his/her gaze with a single webcam. As a result, we will be able to provide the viewer with an immersive environment where he/she can focus on the object easily. Also, we perform a preliminary experiment on the system with a motivation to apply it to audio-based guidance at a museum or an art gallery.

Keywords: Gaze tracking, audio interface, human interaction.

1. 背景と目的

近年, デバイス操作のインタフェースとして視線追跡を利用する手法が数多く研究されている．Jain ら [2] は映像の再編集に視線追跡を導入し, Zhang ら [7] は視線方向情報により画面を操作する手法を提案した．スマートフォンや VR コンテンツなど, デバイスの操作に視線が導入されていたり, 現実世界での人間の行動観察に高度な眼球運動が計測されていたり, 様々な場面で視線追跡が用いられている．これは, 視線を追跡することでユーザの意図や思考を効率的に反映することができるからである．

一方で, 事前にユーザにキャリブレーションを行うと, ユーザに視線を追跡されることを理解されてしまう．ユーザに悟られずに視線を追跡することでしか有用性を発揮できない分野や事例も, また数多く存在すると考えられる．本研究では, 一般的なカメラを 1 台だけ用いて, 事前にキャリブレーションを必要としない，陰的視線追跡の手法を開発することで, ユーザの視線をユーザに気付かれずに追跡することを目的とする．

[1] 慶應義塾大学 大学院理工学研究科
Graduate School of Science and Technology, Keio University, 3-14-11 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan
[a)] f.shimizu@fj.ics.keio.ac.jp
[b)] fuji@ics.keio.ac.jp

Vol.2017-CG-168 No.23
Vol.2017-DCC-17 No.23
Vol.2017-CVIM-209 No.23
2017/11/8

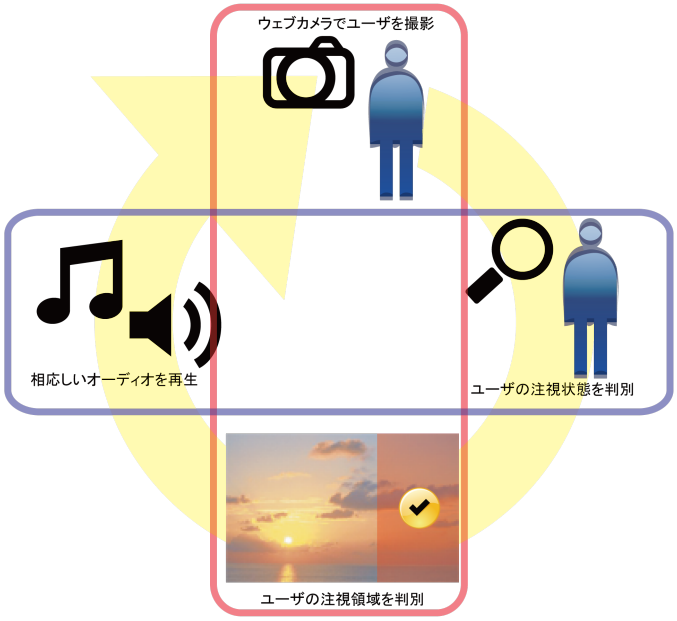
2. 関連研究

Zhan ら [7] は一般的なカメラ 1 台と画像処理だけを用いた視線追跡手法を開発した．ユーザによる事前のキャリブレーションが不要であるが，この手法を用いたインタフェースは左右どちらを見ているかしか判別できない．Wang ら [6] は一般的なカメラ 1 台だけを用いて，視線方向を含めた表情をキャプチャする手法を開発した．視線方向をキャプチャすることで，瞳の位置を正確にアニメーションに反映することができるが，ユーザの見ている場所を推定することはできない．また，Toyama ら [5] はオーディオガイドへの応用を想定して，視線追跡をベースとした聴覚インタフェースを開発した．対象物への集中が目的であるが，ユーザごとに事前のキャリブレーションが不可欠である．さらに，Deguchi ら [1] はユーザが注視している物体を認識し，ユーザに対してオーディオガイドを局所的に再生する手法を開発した．ユーザの対象物への集中を妨げない手法であるが，注視している物体を，視線ではなく位置情報で認識するため，対象物が限定されてしまう．

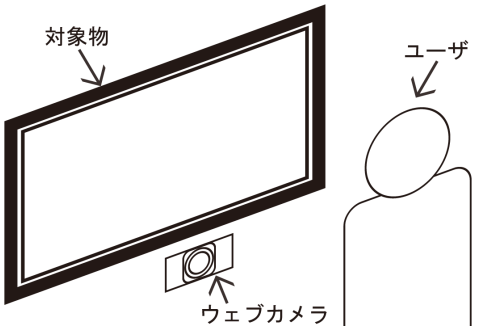
このように，ユーザの認識している対象物に相応しいオーディオデータの再生を目標としている関連研究は存在するが，本システムは，視線追跡に一般的なカメラ 1 台だけを使用し，ユーザに事前学習を必要としない．また，対象物全体ではなく，対象物に局在するオーディオデータを選択可能とする点が特徴的である．本稿では，本研究の先行成果 [4] から，視線追跡の手法により頑健性をもたせ，想定実験の評価を行った．

3. 提案手法

提案システムの概要を**図 1** に示す．処理の流れを示した**図 1(a)** において，赤枠の部分で視線の陰的追跡手法を，青枠の部分でオーディオガイドを想定したエンタテイメント応用を提案する．**図 1(b)** のようにユーザの正面に対象物とウェブカメラを 1 台設置する．はじめに，対象物はあらかじめ領域分割してオーディオデータを割り当てておく．ウェブカメラでユーザを撮影し，対象物に対するユーザの注目度を，顔と目の検出の有無から 3 つのモードに分類し，各モードに応じたオーディオデータを再生する．また，ユーザの視線を追跡し，いくつかに分割された対象物の領域のどこを注視しているかを判定し，その領域に相応しいオーディオデータを再生する．ここで，ユーザの視線方向が変わるたびに再生するオーディオデータを切り替える．



(a) 処理の流れ



(b) ユーザと対象物，ウェブカメラの位置関係

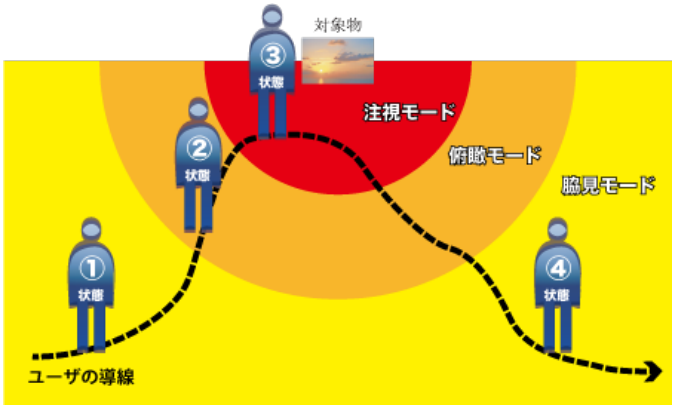
図 1 提案システムの概要

3.1 鑑賞モード

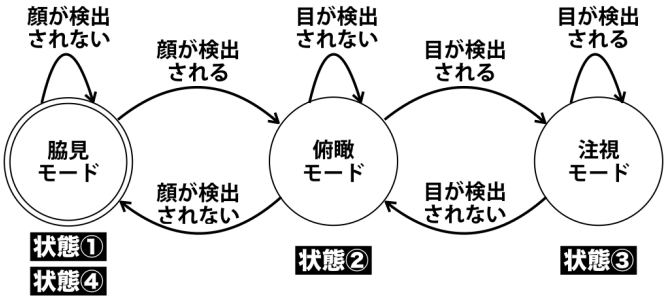
本システムでは，対象物からユーザの距離によってユーザの注視状態が異なると仮定する．ユーザが対象物から遠ければ，ユーザの対象物に対する集中度が小さく，ユーザが対象物から近ければ，ユーザの対象物に対する集中度が大きいとする．ユーザを以下の 3 つの鑑賞モードに分類することで，ユーザの視線を追跡するエリアを限定する．

- 脇見モード: ウェブカメラから取得した画像にユーザの顔が検出されず，ユーザが対象物を認識していない状態で，オーディオデータは再生しないか，変更しない．
- 俯瞰モード: ウェブカメラから取得した画像にユーザの顔が検出されるが，対象物からのユーザまでの距離が遠く，ユーザが対象物の特定の領域を注視せず，対象物全体を見ている状態で，対象物全体に関わるオーディオデータを再生する．
- 注視モード: ウェブカメラから取得した画像にユーザの顔と目が検出され，対象物からユーザまでの距離が近く，ユーザが対象物の特定の領域を注視している状態で，その特定の領域に該当するオーディオデータを再生する．

Vol.2017-CG-168 No.23
Vol.2017-DCC-17 No.23
Vol.2017-CVIM-209 No.23
2017/11/8



(a) 距離によるモード分類



(b) モード遷移

図 2 ユーザの動きによる鑑賞モードの変化

まず, ユーザの動きによる鑑賞モードの変化の様子を**図 2** に示す. 次に, 距離による定義を**図 2(a)** に示す. **図 2(a)** において, 破線に従ってユーザが対象物を鑑賞するとすると, 状態 1, 状態 4 は脇見モード, 状態 3 は俯瞰モード, 状態 4 は注視モードに分類される. さらに, その鑑賞モードが遷移する様子を**図 2(b)** に示す.

また, ユーザは初期状態では対象物を認識していない脇見モードとし, それぞれの距離の変化に応じて状態を遷移させる. 提案手法において, 距離の変化は, ウェブカメラでユーザの顔と目の検出ができるかどうかで判断する.

3.2 視線追跡と注視領域判別

ここで, ユーザと対象物の距離は充分に近く, 鑑賞モードは注視モードに分類されていると仮定する. あらかじめ水平方向 4 等分, 鉛直方向 2 等分された 8 ブロックからユーザの注視している領域を判別する. まず, 8 ブロックのうち左右 6 ブロックのどちらに注視領域があるか判別する流れを**図 3** に示す. はじめに, 対象物を鑑賞しているユーザを正面からキャプチャした画像をウェブカメラから取得する. さらに, 取得した画像をグレースケール画像に変換する. 得

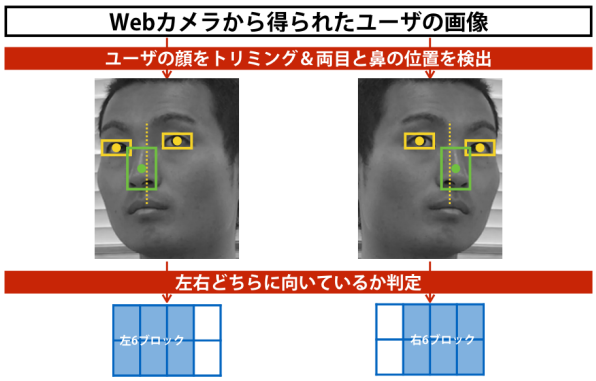


図 3 注視領域判別の流れ（6 ブロック）

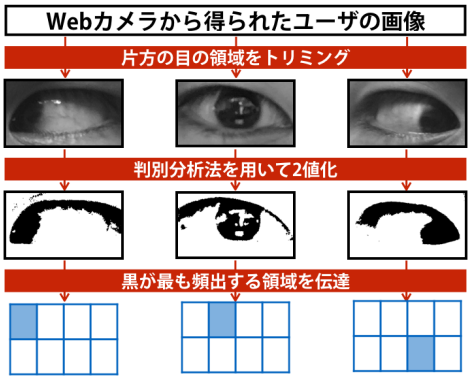


図 4 注視領域判別の流れ（1 ブロック）

られた画像からユーザの両目と鼻の位置を検知し, 左右どちらに顔が向いているかを検出することで, 左右どちらの 6 ブロックかを判別する.

その 6 ブロックのうち, ユーザが注視している領域を判別する流れを**図 4** に示す. グレースケール画像からユーザの目の領域を検出し, どちらか片方の目の部分だけを一定のサイズにトリミングする. 片方の目だけを用いることによって, ユーザの目を遮蔽する前髪の影響を小さくする狙いがある. さらに, トリミングされたグレースケール画像を判別分析法 [8] により 2 値化する. 2 値化された画像の領域のうち, 黒のピクセルが頻繁に出現する領域を注視している対象の領域として判別する.

以上の注視領域判別を**図 5** に示すように 30fps で処理し, 10 フレームごとの最多値を採用する. このことにより, オーディオデータの再生の遅延をより小さくし, まばたきや視線のちらつきによる判別の揺れを抑制する.

Vol.2017-CG-168 No.23
Vol.2017-DCC-17 No.23
Vol.2017-CVIM-209 No.23
2017/11/8

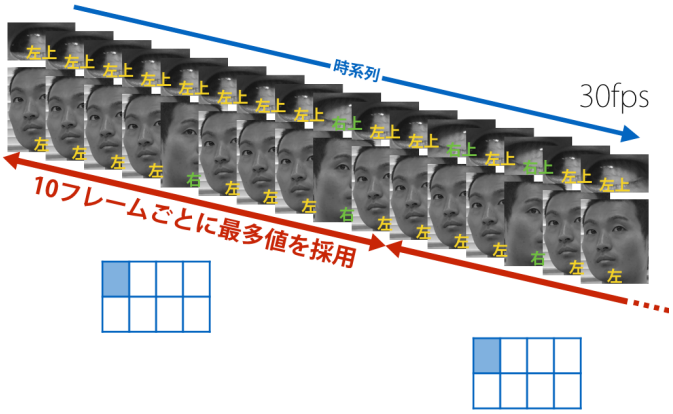


図 5 注視領域判別の頻度

3.3 領域分割

ウェブカメラだけから得られる画像を用いて視線を安定的に追跡する頑健性を確保するため, あらかじめ対象物の局所特徴に応じてユーザが領域を分割する. ここで, 対象物の領域分割は, 横型は**図 6** のように水平方向 4 等分, 鉛直方向 2 等分まで許容する. 分割数に制約はあるが, 分割位置の任意性は残す. これにより, 絵画のような芸術作品やポスターのような文字情報を含む画像など, 多くの対象物に対応することができる.

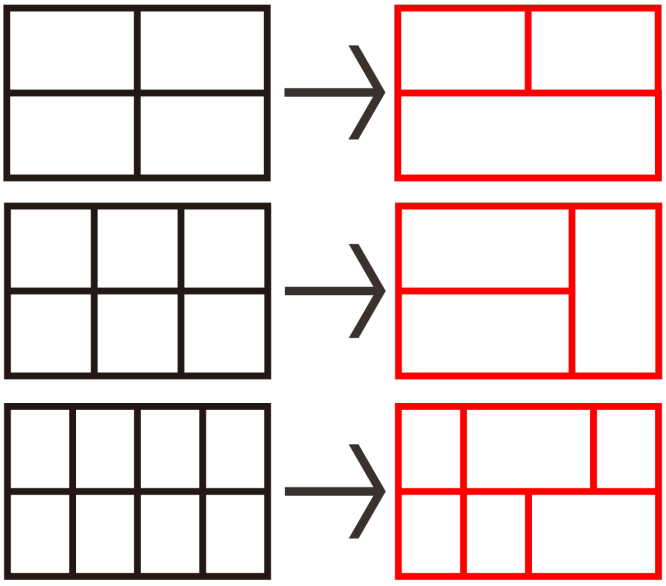


図 6 対象物の領域分割の例

3.4 オーディオデータ再生

オーディオデータを再生する様子を**図 7** に示す. 対象物の各領域に, 該当するオーディオデータをあらかじめ割り当てておく. 俯瞰モードに切り替わったら, 対象物全体に関わるオーディオデータを再生する. 注視モードに切り替わったら, 視線追跡によってユーザが注視していると判定された領域に該当するオーディオデータを再生する. 以降は, 鑑賞モードが変更されたときとユーザが注視している領域が変更されたときに再生するオーディオデータを切り替える. また, オーディオデータを切り替えるとき, ユーザの集中をできる限り削がないよう, クロスディゾルブを適用する.

図 7 オーディオデータ再生の流れ

4. 実装

図 8 に示すようにユーザから対象物の距離は 0.8〜1.0m, ウェブカメラの画角 30 度とした. 開発環境として, プロセッサ: Intel Xeon E5540 2.53GHz CPU, 実装メモリ: 12.0GB を用いて本システムを動作させた. 視線追跡の実装には, Itseez によって開発, 公開されている, コンピュータ・ビジョン向けライブラリ OpenCV (http://opencv.org) を使用した. また, オーディオデータ再生の実装には, Creative Technology, Ltd. によって開発, 公開されている, マルチチャネル 3 次元定位オーディオが表現可能なライブラリ OpenAL (http://www.openal.org) を使用した.

Vol.2017-CG-168 No.23
Vol.2017-DCC-17 No.23
Vol.2017-CVIM-209 No.23
2017/11/8

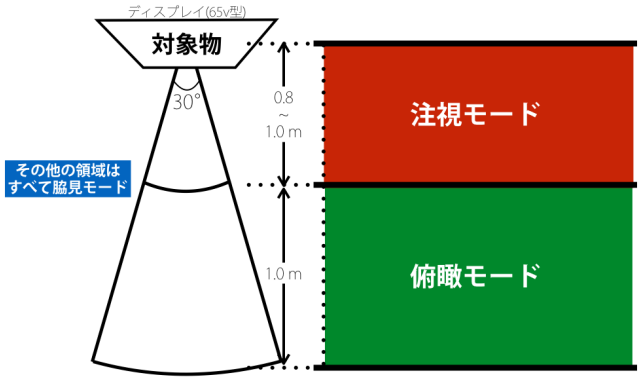


図 8 提案システムの動作環境

5. 結果と評価

本システムの適用例として, 美術館におけるオーディオガイドへの適用を想定した予備実験を試みた. **図 9** に示すようにディスプレイに絵画を表示させ, ディスプレイの中央前方にウェブカメラを設置した. 脇見モードにはオーディオデータは割り当てず, 俯瞰モードには全体の雰囲気に合わせたオーディオデータを, 注視モードにはそれぞれの領域に合わせたオーディオデータを割り当てた. その結果, ユーザの鑑賞モードと注視領域に相応しいオーディオデータを再生できた.

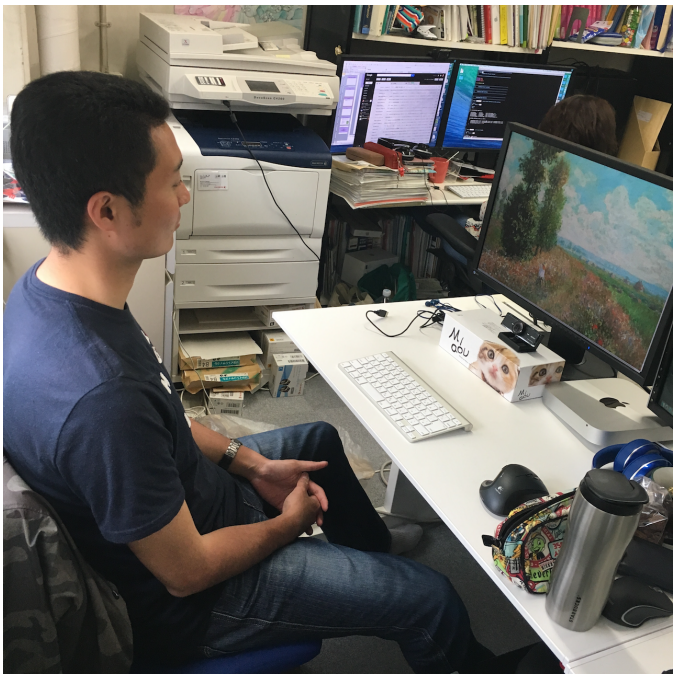


図 9 予備実験の様子

図 10 提案システムの視線追跡に関する精度評価環境

また, 本システムを提案手法により視線追跡した場合と既存の視線追跡デバイスにより視線追跡した場合の精度を比較した. 既存の視線追跡デバイスとして, The Eye Tribe 社によって開発された, The Eye Tribe (http://theeyetribe.com/) を用いた. 1 名の被験者, 1 枚の絵画に対して, 注視モードにおいて, **図 10** に示すようにユーザが対象物を鑑賞する視線の動きを同時に追跡した. その結果を表 1 に示す. このことから, 提案システムの陰的視線追跡手法が, 充分ではないものの高い精度をもつことを示すことができた. しかし依然として, さらに高い精度への改善が求められる.

表 1 提案システムの視線追跡に関する精度評価

計測時間（秒）	60
計測フレーム数（枚）	1,800
一致フレーム数 （枚）	1,324
一致率（%）	73.6

6. 結論

本稿では, ウェブカメラだけで取得したユーザの画像を用い, ユーザに事前にキャリブレーションを必要としない, 陰的視線追跡手法を提案した. さらにその応用として, ユーザの視線方向によって再生するオーディオデータを選択可能にする手法を提案し, 美術館や博物館のオーディオガイドへの応用実験を試みた. 本手法により, ユーザが対象物に集中しやすい環境を作り出し, 臨場感を提供するための視線追跡精度を示すことができた.

Vol.2017-CG-168 No.23
Vol.2017-DCC-17 No.23
Vol.2017-CVIM-209 No.23
2017/11/8

7. 今後の課題

本研究では, 視覚芸術において画面の構図を決定する際に用いられる三分割法に水平方向だけ対応させたが, 鉛直方向にも対応させ, **図 11** に示すように最低でも 4 × 4 へと対応させることが望まれる. 現実には, 三分割法に則っていない対象物が数多く存在するため, それらに領域分割を増加させることも求められる. また, 人の感性を表す AV 空間 [3] に基づいて, 絵画の局所的特徴に対する感じ方とオーディオデータに対する感じ方の対応をとることで, システムがユーザに相応しい対応付けを提案することも課題である. そのなかで, 絵画の局所的特徴に応じて, 音量の大小やテンポの変化など, オーディオデータの特徴を変化させ, より詳細な対応付けを実現することも可能であると考える.

本稿では, 視線追跡の性能だけを測る評価実験について報告した. しかし, オーディオインタフェースとしての快適さを測るための評価実験も求められる. 視線の移動に応じて, オーディオデータをシームレスかつ的確に切り替え, ユーザにとって対象物への集中度が高まっているかを評価しなければならない. その結果をもとに, オーディオデータを切り替えるクロスディゾルブの時間幅や, ユーザの鑑賞モードの遷移を見直す必要がある.

また課題として, 現状では複数のユーザに対応していないことが挙げられる. これを解決するために, 顔認識システムを導入し, 各ユーザに対する視線追跡を行う必要がある.

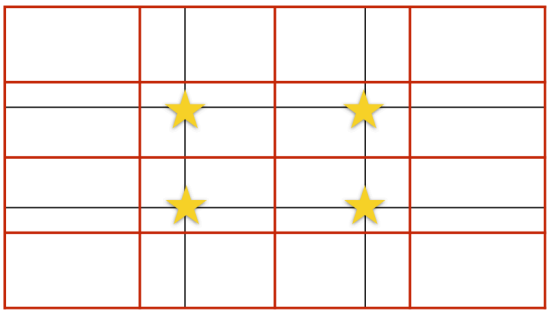


図 11 三分割法

謝辞 本研究の一部は, 科研費挑戦的萌芽研究 15K12034 および 16K12459 の支援により実施された.

参考文献

[1] Akiko Deguchi, Hiroshi Mizoguchi, Shigenori Inagaki, and Fusako Kusunoki: “ A Next-Generation Audio-Guide System for Museums“SoundSpot”: An Experimental Study,” *Knowledge-Based Intelligent Information and Engineering Systems,* pp. 753-760, 2007.

[2] Eakta Jain, Yaser Sheikh, Ariel Shamir, and Jessica Hodgins:“ Gaze-driven Video Re-editing, ”*ACM Transactions on Graphics,* Vol. 34, No. 2, Article No. 21, 2015.

[3] James Russell: “ A Circumplex Model of Affect, ” *Journal of Personality Social Psychology,* Vol. 39, No. 6, pp. 1161-1178, 1980.

[4] Fumiya Shimizu and Issei Fujishiro: “ Selection of Localized Audio Track Based on Eyetracking Technologies with Application to Musical Art Gallery, ” *5th Image Electronics and Visual Computing Workshop (IEVC2017),* 5C–1, Da Nang, 2017.

[5] Takumi Toyama, Thomas Kieninger, Faisal Shafait, Andreas Dengel: “ Museum Guide 2.0—An Eye-Tracking based Personal Assistant for Museums and Exhibits,” *Re-Thinking Technology in Museums,* 2011.

[6] Congyi Wang, Fuhao Shi, Shihong Xia, and Jinxiang Chai: “ Realtime 3D Eye Gaze Animation Using a Single RGB Camera, ” *ACM Transactions on Graphics,* Vol. 35, No. 4, Article No. 118, 2016.

[7] Yanxia Zhang, Andreas Bulling, and Hans Gellersen: “ SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays, ” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* pp. 851-860, 2013

[8] 大津展之: “ 判別および最小 2 乗規準に基づく自動しきい値選定法 ”, 電子情報通信学会論文誌, Vol. J63- D, No. 4, pp. 349-356, 1980