

# 空間局在オーディオの視線追跡選択とそのエンタテインメント応用

清水 文也

藤代 一成

慶應義塾大学 理工学部情報工学科

## 1 背景と目的

近年、デジタルサイネージやテレビなど、さまざまなデバイスのインター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、視線追跡がユーザの興味対象や意図、思考を効率的に反映することができるからである。

Jainらは映像の再編集に視線追跡を導入した[1]。また、Zhangらは視線方向情報により画面を操作する手法を提案した[2]。さらに、Onoらは学習情報を用いて低解像度画像から視線方向を推定する手法を考案した[3]。視覚インターフェースには視線追跡専用の高価なデバイスや身近ではないものを使用することが多いが、コストや普及率を考慮すると発展は望めない。

一方で、視覚インターフェースはオーディオとの融合にも大いに可能性があると考えられる。本研究では、視線追跡をベースとしたオーディオインターフェースを開発することで、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し、ユーザの視線方向によって再生するオーディオを選択可能にする手法を提案する。本研究の視線追跡には一般的なカメラを1台だけ用い、美術館や博物館のオーディオガイドへの応用を提案する。

## 2 提案手法

図1のようにユーザの正面に対象物とウェブカメラを1台設置し、対象物はあらかじめ水平方向に3等分してオーディオを割り当てておく。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を切り替える。

### 2.1 鑑賞モード

ユーザを以下の3つの鑑賞モードに分類し、遷移する様子を図2に示す。

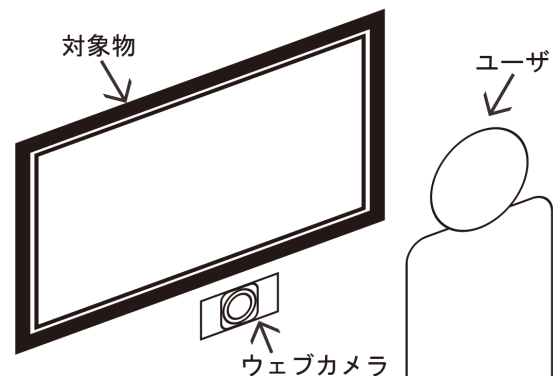


図1: 提案システムの概要

- 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を再生する。

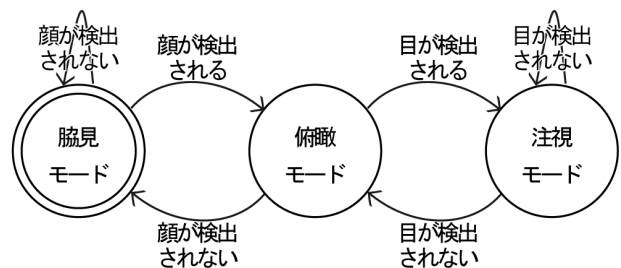


図2: 鑑賞モードの状態遷移図

Eye Tracking-based Selection of Localized Audio with Applications to Entertainment  
Fumiya Shimizu, Issei Fujishiro  
Department of Information and Computer Science, Keio University

## 2.2 視線追跡と注視領域判別

ユーザが注視している領域を判定する流れを図3に示す。まず、対象物を鑑賞しているユーザの体全体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し、グレースケール画像に変換する。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。トリミングされたグレースケール画像を判別分析法 [4] により2値化し、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。

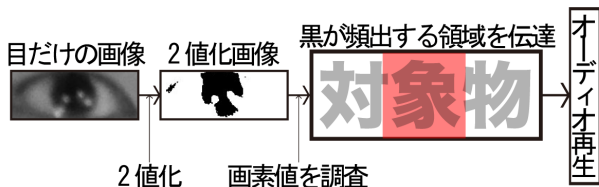


図3: ユーザが注視している領域を判定する流れ

## 2.3 オーディオ再生

対象物の各領域に、該当するオーディオ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を切り替える。

## 3 実装

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度において、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、マルチチャンネル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。

## 4 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試みた。図4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。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を再生できた。

本システムを20代の男女5名に使用してもらい、ユーザが自らの手でオーディオを選択し再生する既存のオーディオガイドと比較してもらった。その結果として、『より自然で対象物に集中しやすい』という評価が得られた。

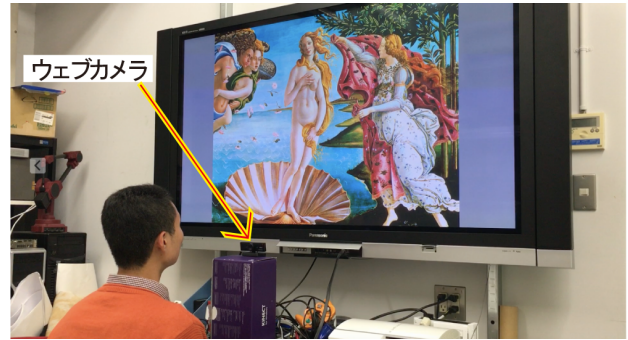


図4: 美術館のオーディオガイド適用例の想定実験

## 5 結論と今後の課題

本稿では、ウェブカメラだけで取得したユーザの視線方向を用い、再生するオーディオを選択可能とし、ユーザに臨場感を提供する手法を提案した。本手法により、対象物に集中しやすい環境をユーザに提供することができた。

美術館における本システムのもうひとつの応用として、絵画に合わせて音楽を自然に切り替える例も実装中である。共通する課題として、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。この課題を解決するために、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。また、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3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ることが望まれる。さらに現実には、3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適応させることも求められる。

## 謝辞

本研究の一部は、平成27年度科研費挑戦的萌芽研究15K12034の支援により実施された。

## 参考文献

- [1] 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.
- [2] 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.
- [3] Yasuhiro Ono, Takahiro Okabe, and Yoichi Sato: "Gaze estimation from low resolution images insensitive to segmentation error," *MIRU 2005*, No. 5, pp. 96-103, 2005.
- [4] 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, 2015