

仮想物体に息を吹きかける動作の認識とその AR 応用

Recognition of blowing on the virtual objects and its AR application

佐々木 隆稀* 高井 昌彰† 飯田 勝吉‡
 Ryuki Sasaki Yoshiaki Takai Katsuyoshi Iida

・1. はじめに

近年、拡張現実(AR)や仮想現実(VR)が広く一般に注目されており、実際にコンシューマ向けのコンテンツが多数配信されている。

これらの技術は、現実空間には“存在しない”仮想的な存在を、人に“存在する”と錯覚させるものであるため、現実空間に則した整合性を持つ必要がある。これらの整合性は、例えば仮想物体の描画に遅延が生じないようにする時間的整合性など、最終的な出力結果の視覚的表示に関するもの他に、現実空間での物理的な動作に応じて知覚する仮想空間の情報に変化し、また仮想空間での事象が現実空間に干渉するという物理的整合性のような、システムを介したインタラクションに関するものがある。つまり、仮想空間との多様な入出力の振る舞いも、人にとって現実空間における自然なものであることが望ましい。

VRにおける入力として、手や頭など体の動き、音声認識、視線などが考えられており、出力はHMDへのディスプレイの他、座った椅子が連動して動く等が実現されている。

一方、ARにおいては空間認識・画像認識による手や体の動きの認識・追跡が実現されているものの、コンシューマ向けARアプリの多くがスマートフォンで動作するという点から、入力は画面のタッチ等の二次元的なものが多い。

そこで、本研究では現実空間から仮想空間への作用のインターフェースとして“息(風)”に注目し、現実空間で吹かれた息を認識し、認識した息による風を仮想空間上に再現するARアプリケーションを実現することを目指す。

・2. 息の認識手法

人の口から吹き出される息を認識し、これを仮想空間で再現するためには、息の判別の他に、息の発生位置・強さ・吹きかけ方向の認識を行う必要がある。

息を判別する手法として、Jellyfish Party[1]のような実際の風を風量計で計測する方法と、Kirifuki[2]のように息の音声データや口唇の画像データなどから吹かれた息を推測する方法の2つがある。本研究では、システムの汎用性を考慮し後者のアプローチを採用し、音声データを用いて息の判別と強さを認識するとともに、顔画像データから息の発生位置と吹きかけ方向の認識を並行して行うシステムを実現する。

・3. 息の認識・再現システム

・3.1 システム概要

息の認識から仮想空間での息再現までのシステム処理の流れを図1に示す。

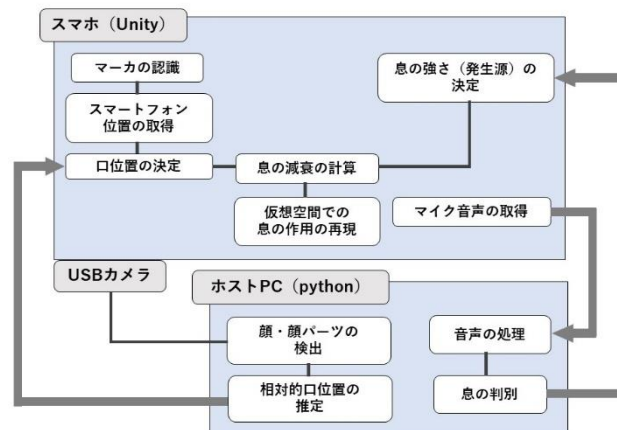


図1 処理の流れ

本システムでは、ホストのPCとクライアントのスマートフォンでそれぞれ音声情報・位置情報について処理を行い、TCP通信によってデータの送受信を行う。ホストPC上ではpythonが、スマートフォン上ではゲームエンジンであるUnityが動作している。また、現時点の実装上の理由により、顔画像の撮影にはスマートフォン本体に装着した小型USBカメラを用いるが、最終的にはスマートフォン内蔵の内カメラと外カメラの同時使用を想定している。

ホストPC側では、スマートフォンの内側に取り付けられたUSBカメラの画像データからスマートフォンと口の相対的位置関係を推定し、同時にスマートフォンから送られてきた音声データから息の判別を行う。

スマートフォン側では、内蔵マイクによる音声の取得と、スマートフォンの外カメラを用いたマーカ認識による空間上での位置取得を同時に行い、ホストPCから送られるデータと組み合わせて仮想空間上で息の作用を再現する。

・3.2 ホストPC側の実装

・3.2.1 息の判別

本システムでは、音声データの周波数分布に注目して息の判別を行う。システム動作時にマイクから入り込む可能性が考えられる環境音例の周波数分布と、あらかじめ用意した息の周波数分布モデルを図2に示す。

*北海道大学大学院情報科学院 Graduate school of Info. Sci. Technology, Hokkaido University

†北海道大学情報基盤センター Information Initiative Center, Hokkaido University

‡北海道大学情報基盤センター Information Initiative Center, Hokkaido University

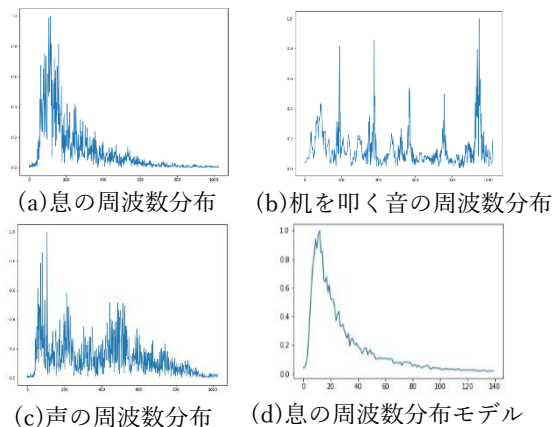


図 2 各々の周波数分布と息の周波数分布モデル

息の判別は周波数分布間の類似度 (L2 ノルム) に基づいて行う。また、息と判別された場合には音声データの最大振幅を息の強さとして用いる。音声データの取得と判別は約 0.2 秒周期で繰り返される。この実装において、息を吹きかけている際の息認識率は 90%程度であり、息を吹きかけていない際の息認識率 (誤認識率) は 0.1%を下回った。

・3.2.2 カメラからの相対的口位置の推定

スマートフォンの内向きに装着された USB カメラから得られる顔画像データに対し、画像処理ライブラリである OpenCV のカスケード分類器を用いることで顔パーツの検出を行う。その検出結果からカメラと口の相対的位置関係 (撮影画像上における口唇位置の 2 次元座標) を推定する。

本システムでは、検出成功確率を高めるために、口だけではなく顔全体と鼻の検出も行い、口唇領域の検出を段階的に行っている。

・3.3 スマートフォン側の実装

・3.3.1 マイク音声の取得

Unity の OnAudioFilterRead 関数によって得られた float 型の配列をバイト列に変換し、ホスト PC に送信する。

・3.3.2 マーカ認識

マーカ認識及び AR 表示を行うためのツールである Vuforia を用いて現実空間におけるマーカの認識を行い、仮想空間上におけるカメラ位置の取得と仮想物体の AR 表示を行う。マーカ認識にはスマートフォン内蔵の外カメラを使用する。

・3.3.3 仮想空間での息の再現

マーカ認識により空間上のカメラ位置が取得でき、さらにスマートフォンに装着した USB カメラの画像データからカメラと口の相対的位置関係が推定できるので、これらの位置情報を組み合わせることにより、3 次元空間における口の位置を導出することができる。

また、息が画面中央に向かって吹きかけられていると仮定することで、3 次元空間における息の向きが得られる。息の強さはマイク音声データから推定される。

以上により、息の発生位置・強さ・向きに関する基本情報が得られるが、仮想空間での息の再現にあたって、風の減

衰の効果も考慮する。本システムでは、息方向と口-物体方向の角度のずれ (θ) と口-物体間距離 (dis) に応じて息が減衰する効果を次式でモデル化する。

$$power_o = power_m \times \cos^n \theta \times \frac{a}{dis - b}$$

ここで、 $power_o$ は仮想物体位置での風の強さ、 $power_m$ は推定口位置での息の強さであり、 n, a, b は減衰の度合いを表す定数である。このモデルによる減衰の様子を図 3 に示す。

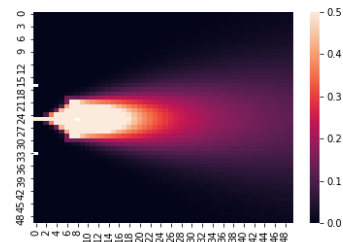


図 3 減衰式を用いた減衰の様子

・3.3.4 アプリケーション

再現された息を用いて、揺れる旗、転がるボール、回る風車のアプリケーションを実装した。また、各地点における息による風の影響を見るため、風のベクトル表現の可視化ツールを実装している。図 4 に揺れる旗の実行例を示す。

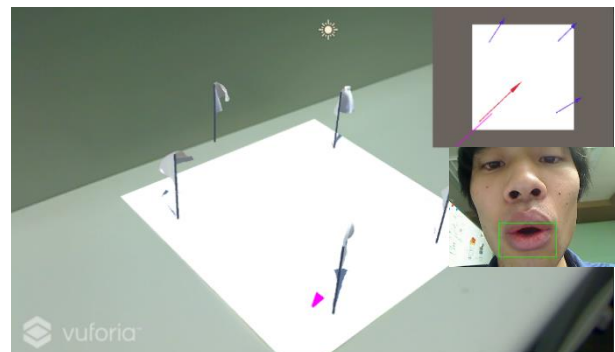


図 4 揺れる旗のアプリケーション

・4. おわりに

本稿では、現実空間で吹かれた息をマイク音声と顔画像データから認識し、仮想空間上で息による風の流れを再現するシステムについて述べた。息判別精度向上のため音声データの機械学習を導入することや、顔画像からの向き推定に基づいて任意方向への息の吹きかけ動作に対応することは今後の課題である。

参考文献

- [1] Yashiro Okuno, Hiroyuki Kakuta, Tomohiko Takayama, Kazuhiro Asai: "Jellyfish Party: Blowing Soap Bubbles in Mixed Reality Space", (2003)
<https://ieeexplore.ieee.org/document/1240759>
- [2] 伊賀聡一郎, 樋口文人: "Kirifuki: 呼吸・吸気によるエンターテインメントシステム", 日本バーチャルリアリティ学会論文誌, vol.7, no.4, p.445-451, (2002)