

# 記憶補助を目的とした装着型実時間顔認識システム

加藤 祐也<sup>1,a)</sup> 内海 ゆづ子<sup>1,b)</sup> Kai Kunze<sup>1,c)</sup> 岩村 雅一<sup>1,d)</sup> 黄瀬 浩一<sup>1,e)</sup>

## 1. はじめに

以前会ったことがある人に再会して、名前が思い出せない—このような経験を誰しもしたことがあるだろう。人は、仕事や普段の生活で数え切れない人と出会うが、記憶力には限界があり、そのすべてを記憶しておくことは不可能である。しかし、一度会ったことのある人の名前を本人の前で言えないのは、非常に恥ずかしく、また、相手にとっても気分のよいものではない。この問題を解決するため、人に代わって出会った人を記憶し、再会すると相手に悟られずに名前を提示するシステムを我々は提案した [7]。このシステムでは、顔認識を用いて個人認証を行っていたが、顔追跡の失敗により誤認識が多かった。

そこで本稿では、システムを改良し、認識精度の向上を図った。従来手法では、顔追跡に色情報を用いていたが、顔に色の近い首の部分や他の物体まで追跡を行うため、追跡精度が低く、認識にも影響を及ぼしていた。そこで、フーリエ変換を用いた手法 [3] を採用することで、顔追跡精度の向上を図った。実験の結果、提案システムの認識精度が 88.6% となり、顔追跡の精度向上によって認識の精度が向上した。

## 2. システム概要

今回作成したシステムを図 1 に示す。このシステムはカメラ (logicool 製 HD Pro Webcam C920)、ヘッドマウントディスプレイ (HMD, VUZIX 製 Wrap1200VR)、ノートパソコン (Panasonic 製 Let's note AX2) からなり、装置を装着したまま自由に動き回ることができる。カメラの解像度は 640×480 pixels であり、フレームレートは 15fps である。HMD の解像度は 1280×720 pixels である。ノートパソコンは intel Core i7-3687U CPU, 8GB メモリを搭載している。カメラから送られたデータはノートパソコンで

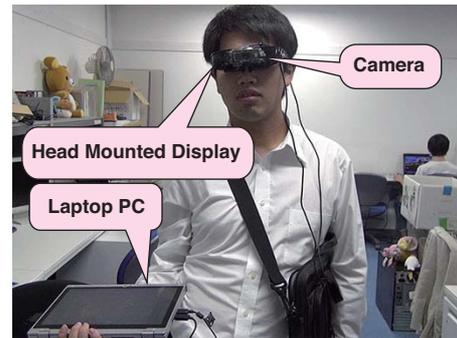


図 1 提案システムの概要

処理され、認識結果が HMD に表示される。

システム内での処理の流れを図 2 に示す。提案システムではまず、カメラで撮影された画像に対して顔検出、追跡を行い、顔画像を切り取る。続いて、得られた顔画像を認識し、最終的に認識結果を表示する。本システムは、ユーザに代わって出会う人を認識する必要があるため、できるだけ高速に顔の認識を行うことが必要とされる。そのため、顔検出、追跡は高速なものが望ましい。本システムでは、顔検出に Viola らの手法 [6] を用いた。この手法は AdaBoost をもとに作成された Cascade 構造の識別器により、顔を高速に検出可能である。また、追跡には Henriques らの物体追跡手法 [3] を用いた。この手法では、フーリエ変換を用いることにより、対象物体の追跡を 100fps 以上の速度で行うことが可能である。提案システムは、どのような場所でも利用でき、相手と自然に話などをしながら画像の撮影を行うことを想定しているため、撮影した画像の顔向き、表情や、照明は常に変化する。このようなアピランズが変化した画像から認識を行うために、提案システムでは複数の顔画像を用いて認識をする手法 [1] を用いて顔認識を行った。この認識では、複数の同じクラスに属す

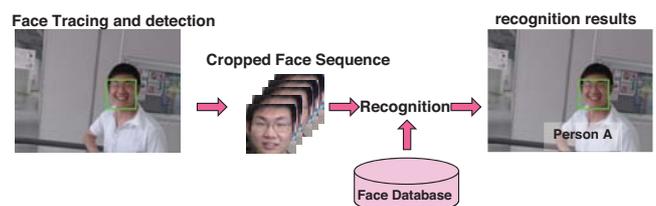


図 2 顔認識の処理の流れ

<sup>1</sup> 大阪府立大学 大学院工学研究科 〒 599-8531 大阪府堺市中区学園町 1-1

a) katou@m.cs.osakafu-u.ac.jp

b) yuzuko@cs.osakafu-u.ac.jp

c) kunze@m.cs.osakafu-u.ac.jp

d) masa@cs.osakafu-u.ac.jp

e) kise@cs.osakafu-u.ac.jp

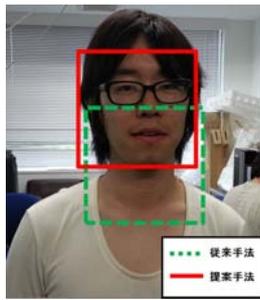


図 3 顔の追跡結果の一例

る顔画像を 1 つの凸包で表現し、凸包間の距離を用いて認識を行う。このことで、アピアランスが一致している顔画像同士の比較を自動的に行うことができる。提案システムでは、追跡の結果から得られた顔画像の輝度値を用いて凸包の計算を行った。

### 3. 実験

提案システムの顔の追跡、認識精度の評価を行った。

#### 3.1 顔追跡の評価

従来システムで用いていた色情報に基づいた顔追跡手法 [2] と、提案システムで用いた顔追跡手法について比較を行った。提案システムで撮影した画像に対して同じ顔検出手法を用いて顔検出を行い、従来システムでの手法と提案システムでの手法を用いて顔の追跡を行った。画像の解像度は  $640 \times 480$  pixels である。追跡結果の一例を図 3.1 に示す。図 3.1 より、従来システムで用いていた手法は、首まで顔領域として追跡していたが、提案システムの手法では、顔の領域を検出しており、検出の精度が向上したことが分かる。

#### 3.2 顔認識の評価

顔追跡の精度が顔認識に与える影響を調査した。実験には、VidTIMIT Audio-Video Database [5] を用いた。このデータベースは 43 人の人物の動画画像から構成される。撮影時にカメラは固定されており、動画画像中で被験者はそれぞれ短い文章を音読し、顔を左右上下に動かしている。動画画像の解像度は  $512 \times 384$  pixels でフレームレートは 25 fps、動画画像の長さはそれぞれ異なり、12 秒から 20 秒程度である。本実験では、35 人分の画像を用いて行った。データベースには session 1 の動画画像から各人物ごとに 200 フレーム分用いた。テストデータには session 2 の画像を用いた。テストデータは 1 セット 16 枚とし、1 人あたり 10 セットの合計  $10 \times 35 = 350$  セットを用いた。Session 1 と Session 2 は、撮影された日は 2 日から 7 日程度の間隔があり、被験者の髪型、化粧は異なっている場合がある。データベース、テストデータともに検出、追跡した顔を切り出して 8 bits グレースケールに変換したのち、 $20 \times 20$  pixels に正規化



(a) 従来システムで用いた追跡手法



(b) 提案システムで用いた追跡手法

図 4 VidTIMIT Audio-Video Database の追跡結果例

した画像の画素値を特徴量として用いた。提案システム、従来システムで追跡、認識を行って画像を切り出した例を図 4 に示す。実験の結果、提案システムでは 88.6%、従来システムでは 77.7% の認識率を示し、提案システムで顔追跡の精度が向上したことで認識精度が向上したことが明らかとなった。

### 4. まとめと今後の課題

本稿では、人の記憶を補助することを目的とした装着型顔認識システムを改良した。顔の追跡手法を従来のものに換えてフーリエ変換を用いた手法を用いることで、追跡精度と認識精度を向上させることができた。今後の課題として、顔画像のデータベースの拡張方法の検討と、さらなる認識率の向上が挙げられる。現在のシステムでは、出会った事がある人をデータベースに登録していた。これに加えて、Facebook などの SNS から得られた顔画像の情報を用いることで、データベースの拡張を図ることが考えられる。このことが実現できれば、初対面であっても、SNS 上で知り合いの人をすぐに認識することが可能となる。また、認識率の向上に関しては Kapoor ら [4] の提案する顔画像の特徴量以外の人の背景や情報などを複合的に用いることが考えられる。たとえば、ある仕事でのつながりのある人に出会った場合、仕事でつながりのある他の人とも出会う確率が高くなる。そこで、人とのつながりの情報も SNS から抽出し、顔特徴量と複合して用いることで顔認識の精度向上を図る。

#### 参考文献

- [1] Cevikalp, H. and Triggs, B.: Face Recognition Based on Image Sets, *Proc. of CVPR*, pp. 2567–2573 (2010).
- [2] Comaniciu, D., Ramesh, V. and Meer, P.: Real-time tracking of non-rigid objects using mean shift, *Proc. of CVPR*, Vol. 2, pp. 142–149 (2000).
- [3] Henriques, J. F., Caseiro, R., Martins, P. and Batista, J.: Exploiting the Circulant Structure of Tracking-by-Detection with Kernels, *Proc. of ECCV*, pp. 702–715 (2012).
- [4] Kapoor, A., Lin, D., Baker, S., Hua, G. and Akbarzadeh, A.: How to Make Face Recognition Work: The Power of Modeling Context, *Workshop on Activity Context Representation at AAAI 2012* (2012).
- [5] Sanderson, C.: Biometric Person Recognition: Face, Speech and Fusion, *VDM-Verlag* (2008).
- [6] Viola, P. and Jones, M. J.: Robust real-time face detection, *IJCV*, Vol. 57, No. 2, pp. 137–154 (2004).
- [7] 加藤祐也, 内海ゆづ子, 黄瀬浩一: カメラ付きヘッドマウントディスプレイを用いたウェアラブル顔認識システム, 画像の認識・理解シンポジウム (MIRU2012) 予稿集 (2012).