

教室での入退室管理用顔認証システム構築に向けた シヤムネットワークの性能評価に関する研究

棚町貴子¹ 廣重法道² 鶴田直之²

概要: 学校では、生徒の出欠確認や活動場所の特定に時間や手間を取られることがある。本研究では、出欠確認の負担軽減を目的として、日本人の生徒を主な識別対象としたネットワークカメラを用いたリアルタイム顔認証システムの開発に向け、照合用画像が1枚もしくは少ない枚数でも識別することが可能なシヤムネットワークの性能（認識率とリアルタイム性）の評価を行った。

キーワード: 画像認識・理解, 分類学習, 認証

A Research on Performance Evaluation of Siamese Networks for Face Recognition System for Access Control in Classrooms

TAKAKO TANAMACHI^{†1} NORIMICHI HIROSHIGE^{†2}
NAOYUKI TSURUTA^{†2}

Abstract: In schools, it sometimes takes a lot of time and effort to check the attendance of students and identify their activity locations. In order to improve the efficiency of access control, we are developing a real-time face recognition system using a network camera that mainly identifies Japanese students. In this paper, we evaluate the applicability of Siamese networks, which can identify faces even with one or a small number of images for matching.

Keywords: Image recognition/understanding, Classification, Authentication

1. はじめに

1.1 研究の背景

近年、セキュリティ意識の高まりから、ビルの入退室管理や銀行の本人確認、コンサート会場のなりすまし防止などに顔認証技術が活用されている^[1]。現在、新型コロナウイルス対策として、顔認証は非接触の個人認証を可能とする有効な手段としてのニーズがより一層高まっている。その一方、人種や性別により精度が異なるといった顔認証のバイアスが報告されている^[2]。

さて、学校では、教室の移動を伴う授業やクラスの枠を超えたグループ活動（総合的な探求の時間や課題研究等）において、生徒の出欠確認や活動場所の特定に時間や手間を取られることがある。そこで、教室への入退室管理に顔認証システムを活用することができれば、出欠確認の負担を軽減するだけでなく、感染症等が発生した際、教室利用者の情報提供を関係機関へ速やかに提供することができる。これに対し、インターネット上には、自作で顔識別を実装したプログラムが多数公開されている。ただし、学習に必要なデータセットに日本人の顔画像のものは少ない。さらには、通常の識別用ニューラルネットワークでは、学習のために1人に対して大量の顔画像が必要となるが、毎年入

れ替わる生徒1人1人に対して、大量の画像を用意することは困難であると考えられる。

1.2 研究の目的と概要

本研究の目的は、日本人学生を主な識別対象としたネットワークカメラを用いたリアルタイム顔認証システムの構築である。今回、入力された顔画像に対して個人を一意に特定する「顔識別」に、照合用画像が1枚もしくは少ない枚数でも識別することが可能なシヤムネットワーク（Siamese Network）[3]を用いて、リアルタイム顔識別を行うプログラムを作成した。なお、入力画像の中から顔を特定し、顔が存在する領域を推定する「顔検出」に、Haar-like特徴量ベースのCascade分類器[4]を用いた。プログラムは学習時間短縮のため、GPU（Graphics Processing Unit）を利用するよう改良を行った。また、日本人学生の学習用データセットを作成し、シヤムネットワークで顔の特徴空間を作成、既存の大量の外国人顔画像で作成した特徴空間との識別精度の比較を行った。本研究では、リアルタイム顔識別プログラムの作成に加え、大量の外国人顔画像で作成した特徴空間（学習モデル）が日本人学生を識別する一般モデルとして適しているかを明らかにするとともに、識別対象者を絞った学習用データセットから特徴空間を作成する

¹ 福岡県立城南高等学校
Fukuoka Prefectural Jonan High School
² 福岡大学工学部電子情報工学科

Dep. of Electronics Engineering and Computer Science, Fukuoka University

ことで、識別精度を向上させることを目標とする。本研究の概要を図1に示す。なお、実装はPythonを用いた。

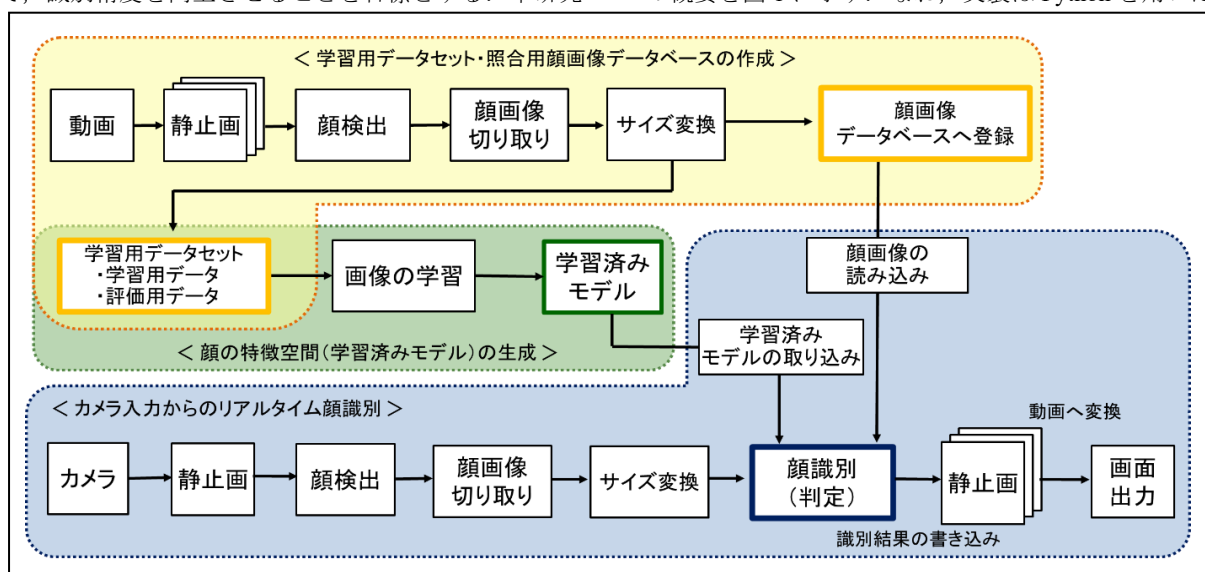


図1 研究の概要

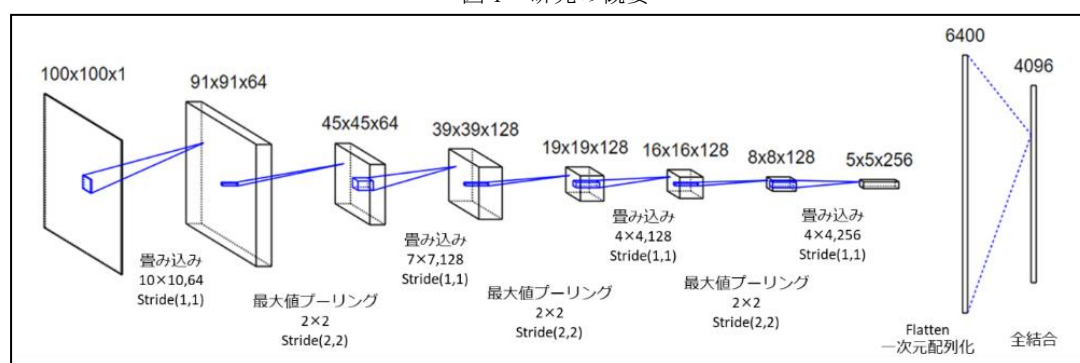


図2 畳み込みニューラルネットワークの構造

2. シャムネットワークを用いた顔識別

顔識別を行うために、畳み込みニューラルネットワーク (CNN: Convolutional Neural Network) で顔画像の特徴ベクトルを算出し、顔画像間の関係性をモデル化するのにシャムネットワークを用いて、顔の特徴空間を作成した。

また、特徴空間作成時の学習時間短縮のため、GPU (Graphics Processing Unit) を利用するプログラムに改良した。ここでは、シャムネットワークを用いた顔の特徴空間の作成と学習時間短縮のためのプログラムの改良、実行結果について報告を行う。

2.1 畳み込みニューラルネットワークによる特徴ベクトルの算出

入力画像に対する特徴ベクトルの算出のため、畳み込みニューラルネットワークを利用した。畳み込みニューラルネットワークの構造を図2に示す。入力は 100×100 に大きさを統一したグレースケール1層の画像である。図2の順に畳み込みと最大値プーリングを行う。Flattenで、特徴マップを一次元配列化してベクトルに変換する。全結合によって、Flattenで出力されたベクトルから、サイズ4096の

ベクトルを出力する。この出力が、入力画像の特徴ベクトルである。

2.2 シャムネットワークを用いた顔の特徴空間の作成

シャムネットワークとは、二つの画像を入力し、その画像が似ているのか似ていないのかの距離を算出するネットワークである。入力画像を学習させるために、事前に同一カテゴリ (同一人物) と非同カテゴリ (異なる人物) のペアにグループ化する。ラベルとして、同一カテゴリのペアには1, 非同カテゴリのペアには0を設定する (図3)。次に、シャムネットワークの構造を図4に示す。



図3 画像のグループ化とラベルの設定

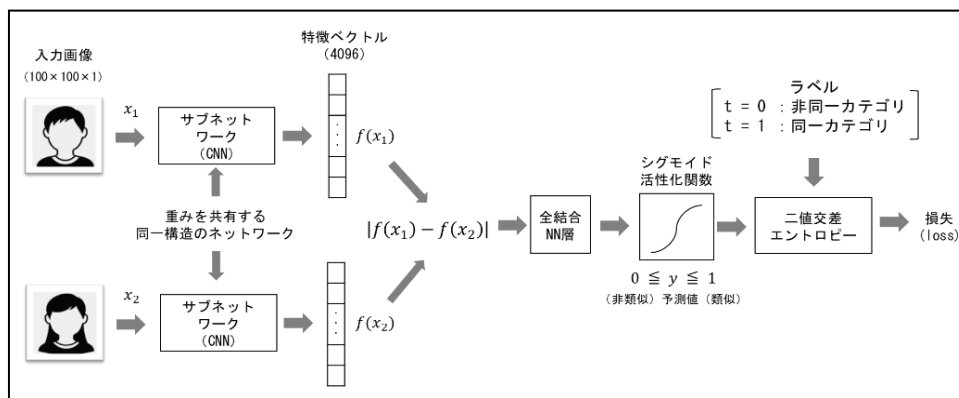


図4 シャムネットワークの構造

シャムネットワークでは、二つの入力画像を比較するために、同じ構造を持ち、同じパラメータと重みを共有するサブネットワークに入力する。本研究では、図2の畳み込みニューラルネットワークをサブネットワークとし、2枚の画像を各特徴ベクトルに変換する。2つの特徴ベクトルの差ベクトルのノルム(L1)を取ることで画像間の距離を求める。求めた距離をもとに、単一出力ニューロンを持つ全結合演算を行い、シグモイド活性化関数で類似性を判断する予測値を出力する。この予測値は0～1の確率値で出力され、0に近いほど類似しておらず、1に近いほど類似していることを表している。損失関数に二値交差エントロピーを用いて、出力された予測値とラベルの間における損失を求める。損失を求める式を式(1)に示す。ここで、 t はラベルで0または1を取り、 y は予測値である。

$$\text{loss} = -\log(y) - (1 - t)\log(1 - y) \quad (1)$$

学習中のネットワークは、損失をもとにパラメータと重みの更新を行う。この時、サブネットワークは、同一カテゴリである場合、2つの画像間の距離が小さくなるように、画像が非同カテゴリであれば距離が遠くなるようにパラメータと重みを学習する^[5]。

2.3 シャムネットワークと通常の識別用ニューラルネットワーク (NN: Neural Network) の違い

図5に、シャムネットワークと従来の識別用ニューラルネットワークの違いを示す。二つのネットワークの違いを識別部分(全結合)と特徴空間構成部分(CNN)に分けて考える。

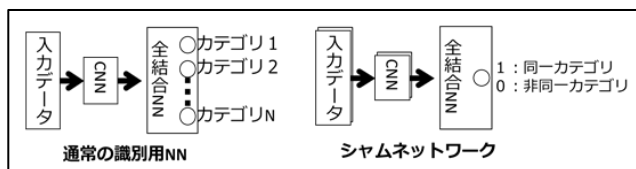


図5 通常の識別用NNとシャムネットワークの違い

通常の識別用ニューラルネットワークでは、画像を特定のカテゴリやラベルに分類し、識別したいカテゴリに対応するニューロンだけが1で他を0とするOne-hot-encodingで学習を行う。よって、カテゴリ数と同数の出力ニューロン数が必要となる。そのため、カテゴリを追加する場合、

出力ニューロンを増やし再学習する必要がある。これに対して、シャムネットワークでは、出力ニューロン数は一つで、同一カテゴリか非同カテゴリかのみで識別を行う。カテゴリを追加する場合でも、画像間の距離のみで比較するため、再学習する必要はなく、照合用のデータベースに少量の画像を追加するだけで、モデルは同じように機能する。

特徴空間構成部分の構造に違いはない。しかし、識別部分から伝播される誤差が異なるため、特徴空間は異なる。通常の識別用ニューラルネットワークでは、カテゴリごとの学習用画像が大量に用意できる場合にのみ性能が向上する。シャムネットワークでは、同一カテゴリは近く、非同カテゴリは遠くなるよう距離を学習するため、カテゴリごとにまとめた特徴空間になると期待される。

したがって、生徒1人に対して大量の画像を用意することが難しく、年度途中の転入による生徒の追加等を考えると、通常の識別用ニューラルネットワークよりシャムネットワークの方が本研究に適していると考えられる。

3. 事前実験

3.1 実行環境

Aditya氏が公開しているシャムネットワークによる顔識別プログラムと顔画像データセット[6]を利用し、必要なライブラリのインストールとデバッグを行い、実行可能な状態にした。しかし、大量の画像を処理するため、クライアントPCでは動作しなかった。そこで、実行環境をサーバに変更し、クライアントPCからリモート接続で作業した。クライアントPCの計算機環境を表1に、サーバの計算機環境を表2に示す。

表1 クライアントPC

OS	Windows 10 Pro (Microsoft)
CPU	Intel(R) Core(TM) i7-8650U
メモリ	16.0 GB
言語	Python 3.8.10
開発環境	Visual Studio Code 1.56.2 (Microsoft)

表 2 サーバ

OS	Ubuntu 20.04.2
CPU	Intel(R) Core(TM) i7-6700K CPU @ 4.00GHz
GPU	NVIDIA Corporation GP102 [GeForce GTX 1080 Ti]
メモリ	375.6GB
ディスク	22.0TB

3.2 外国人顔画像データセット

外国人顔画像データセットには学習用 140 名, 13,749 枚, テスト用に 60 名, 5,311 枚の画像が用意されている. それぞれのデータセットには, 人物名 (一意) のフォルダがあり, フォルダ内にはフォルダ名の人物の表情画像が複数枚用意されている. 学習用データセットの一部を図 6 に示す.



図 6 学習用データセット (上) 学習用画像フォルダ
(下) フォルダ内の表情画像

3.3 事前実験 1: GPU の利用による学習時間の短縮

大量の外国人顔画像データセットから特徴空間を作成する実験を行った. 以下, 大量の人物画像から作成した特徴空間を大特徴空間と呼ぶ. 特徴空間を作成するのに 200,000 回学習を行った結果, 完了までに約 45 時間かかった. そこで, プログラムを CPU のみで処理するのではなく, GPU も利用することで学習時間の短縮を図った. GPU を利用するために, google 社の「tensorflow-gpu」, NVIDIA 社の「CUDA」と「cuDNN」をインストールし, プログラムの改良を行った.

その結果, 学習時間は約 6 時間となり, CPU のみで処理した場合に比べ, 学習時間を約 8 分の 1 に短縮することができた.

3.4 事前実験 2: 外国人顔画像未知データによる顔識別

識別したい個人の画像 (1 枚) と名前が登録されている顔画像データベースと大特徴空間を用いて, 未知の外国人顔画像データで識別を行った. 実行結果を図 7 に示す. 図 7 では, 中央の女性のみ名前が表示されている. 名前の右に表示されている数値は予測値である. この人物は, データベースに登録されており, 名前も一致しているため識別ができています. Unknown の表示は予測値が 0.5 未満でデータ

ベース内に登録されていないと判断された結果である. しかし, Unknown と表示された右端の男性はデータベースに登録されており, 誤識別を起こしていることが分かった. また, 矩形が表示されていない人物は顔自体が未検出である.

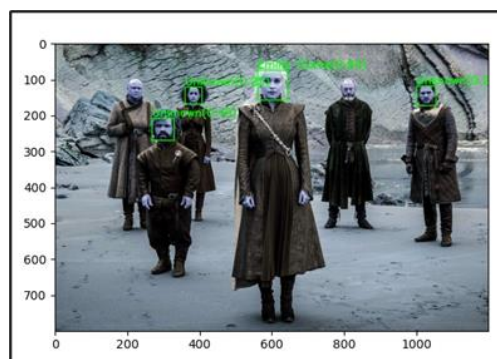


図 7 未知データでの顔識別結果

3.5 事前実験 3: 大特徴空間での日本人顔画像での顔識別

顔画像データベースに日本人の識別対象者の画像データと登録名をファイル名として追加し, 大特徴空間による顔識別実験を行った. 顔画像データベースに追加した識別対象者のデータ, 顔識別を行う未知データ (静止画) および出力結果を図 8 に示す.

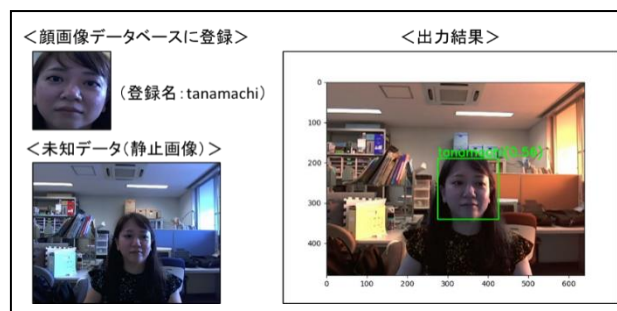


図 8 大特徴空間での日本人画像の顔識別

図 8 の出力結果より, 表示名が登録名と一致しており識別ができていることがわかる. 同様に, 日本人学生 9 名のデータを新たに顔画像データベースへ登録し, 各個人の未知データで顔識別の実験を行った. その結果, 9 名のうち正しく識別できたのは 1 名, 誤識別が 8 名であった. このことから, 日本人学生の顔識別を行う場合, 一般モデルとして, 大量の外国人顔画像で作成した大特徴空間を利用しても十分な精度は期待できないと考えられる.

そこで, 日本人の識別対象者にデータを絞って, 顔画像データセットから特徴空間を作成することにより精度向上を図る. 以後, 識別対象者に絞って作成した特徴空間を小特徴空間と呼ぶ.

4. 日本人学生の顔画像による小特徴空間の作成

4.1 学習用データセットと照合用の顔画像データベースの作成

識別対象者の学習用データセットの作成には、Intel 社が開発・公開しているコンピュータビジョン向けライブラリ「OpenCV」を利用した。

顔画像のデータは、Microsoft 社の Teams で行った会議動画から 10 秒程度を切り出し、収集した。まず、動画からフレームを切り出して、静止画で保存する。静止画をグレースケールに変換し、矩形領域の明暗差を利用して物体検出を行う Haar-like 特徴量ベースの Cascade 型分類器で正面顔の検出を行った。検出した顔部分を切り出して、入力サイズの 100×100 ピクセルに変換し、識別対象者名のフォルダに保存した。1 人当たり学習用 200 枚、テスト用 50 枚の顔画像データを 20 名分収集し、事前実験と同じフォルダ構成で学習用データセットを作成した (図 9)。また、未使用のデータを 1 枚、照合用の顔画像データベースに識別対象者名とともに登録した (図 10)。



図 9 日本人学生の学習用データセット (一部)

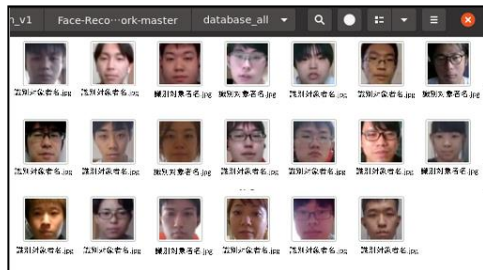


図 10 顔画像データベース

4.2 シヤムネットワークによる小特徴空間の作成

日本人学生の学習用データセットをシヤムネットワークで学習し、小特徴空間を作成した。学習回数は事前実験と比較するため、200,000 回とした。

学習過程の特徴空間の変化をみるために、シヤムネットワークの中間層である畳み込みニューラルネットワークから出力される 4096 次元の顔の特徴ベクトル (図 2, 図 4) を抽出し主成分分析を行った。図 11 に学習前の第 1 主成分と第 2 主成分の散布図、図 12 に 20 万回学習時の第 1 主成分と第 2 主成分の散布図を、図 13 に 20 万回学習時の第 1 主成分から第 3 主成分の散布図を示す。

図 11 より、学習前の特徴ベクトルの第 1 主成分と第 2 主成分は 0 の交差する付近に集中して分布している。図 12 より、学習を行うことでカテゴリ (識別対象者) ごとの距離

が離れ、カテゴリごとにまとまった分布となっていることが分かる。その中で、識別対象者 G と N、識別対象者 B と T、識別対象者 A と H と R は、データ重なりや距離が十分に取れていないように見える。そこで、第 3 主成分を追加した (図 13)。すると、図 12 ではデータの重なりや隣接しているように見えていた識別対象者 G と N、識別対象者 B と T、識別対象者 A と R は距離が離れていることが分かる。しかし、識別対象者 H と R は離れ切れていない。これは、第 3 主成分までの特徴ベクトルが類似しているためだと考えられる。

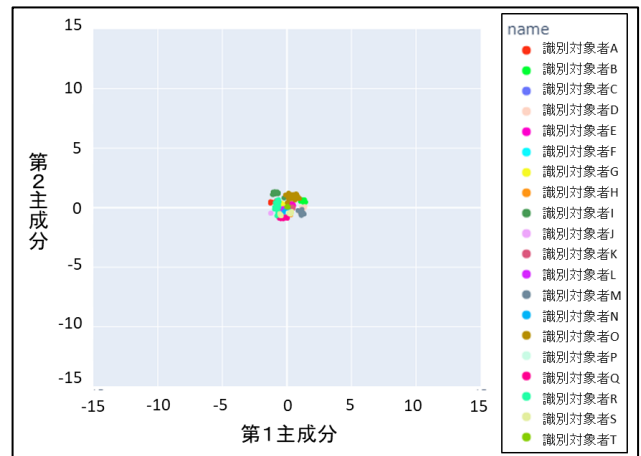


図 11 学習前の小特徴空間 (2 次元)

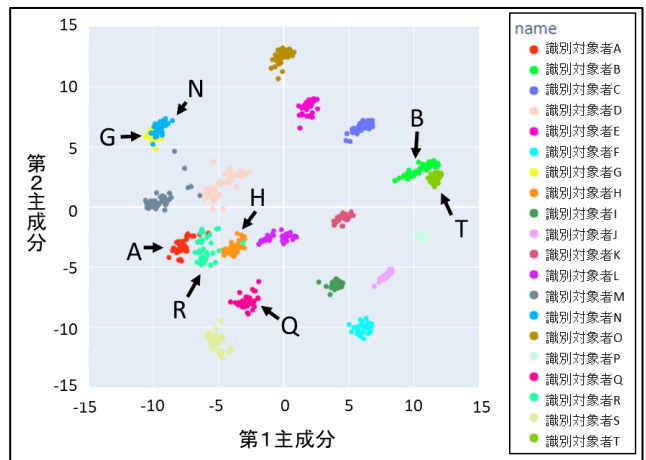


図 12 20 万回学習時の小特徴空間 (2 次元)

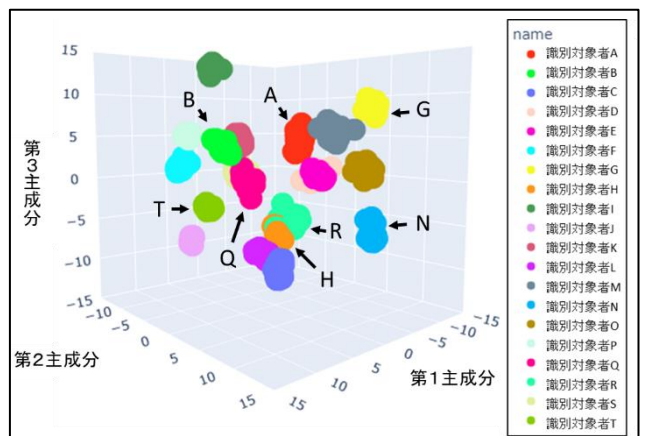


図 13 20 万回学習時の小特徴空間 (3 次元)

5. 実験

5.1 小特徴空間での顔識別実験

事前実験 3 の顔画像データベースと未知データ (図 8) を用いて, 小特徴空間で顔識別を行った結果を図 14 に示す. 図 14 より, 識別結果は登録名と一致しており, 識別ができていたことがわかる. また, 大特徴空間での予測値は 0.56 であったのに対して, 小特徴空間では 0.98 と予測値が高いことがわかる.

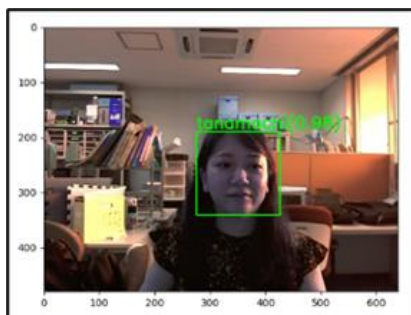


図 14 小特徴空間での顔識別

5.2 大特徴空間と小特徴空間の識別精度比較

大特徴空間と小特徴空間の識別精度の比較を行った. 図 15 に大特徴空間での識別結果, 図 16 に小特徴空間での識別結果を示す.

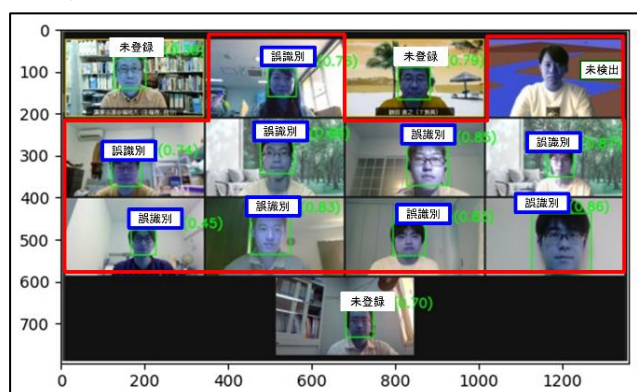


図 15 大特徴空間による顔識別結果

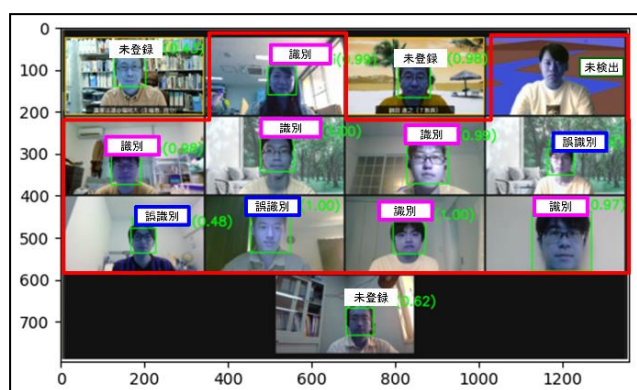


図 16 小特徴空間による顔識別結果

照合用の顔画像データベースは, 図 10 の識別対象者 20 名が登録されているものを用いた. 20 名の識別対象者のうち, 10 名の識別対象者が含まれている複数人画像で比較を

行った. なお, 学校現場での利用を想定し, 識別対象者が全員登録されていることとし, Unknown の表示は行わない変更を加えている. ここでは赤線で囲まれている識別対象者のみの結果について考える.

図 15 より, 大特徴空間では識別対象者 10 名のうち 9 名が誤識別, 1 名が未検出となっている. それに対して, 図 16 の小特徴空間では識別対象者 10 名のうち 6 名が正しく識別できており, 誤識別 3 名, 未検出 1 名であった. これにより, 日本人学生の顔識別を行う場合, 大量の外国人画像で作成した大特徴空間を一般モデルとしても用いるよりも, 日本人学生の識別対象者に絞った小特徴空間を用いた方が識別精度は高いことがわかる. よって, 小特徴空間の作成により識別精度を向上させることができた. なお, 未検出者 1 名は, 顔識別以前の顔検出ができていない. 顔画像データセットを作成する際, 背景によって顔検出がうまくいかず, 背景の異なる画像を用意する事例があった. したがって, 顔検出の際に背景の影響を受けることがあるため, カメラを設置する際は, 顔を検出しやすい環境かどうかを検討する必要があると考えられる.

5.3 ネットワークカメラとサーバ利用によるリアルタイム顔識別

顔識別プログラムと小特徴空間, 顔画像データベースを用いて, カメラ画像からリアルタイムに顔識別を行う環境の構築を行った. 画像入力にネットワークカメラ, 顔識別を行う機器としてサーバ, 識別結果の出力にディスプレイを用いた.

なお, USB カメラと Raspberry Pi 4B で入力から顔識別までを行う環境の検討も行ったが, 静止画 1 枚の識別結果を得るのに約 191.9 秒 (約 3 分 10 秒) かかった. よって, Raspberry Pi 4B でも顔識別プログラムを実行することはできるが, リアルタイムに処理を行うには十分な性能がないとし, サーバを利用することとした.

ネットワークカメラの IP アドレスを固定し, OpenCV を利用して, プログラム内に RTSP (Real Time Streaming Protocol) プロトコルを使用したストリーミング URL を記述することで, ネットワークカメラの映像をプログラム内に取り込んだ. 取り込んだ画像から, Haar 特徴ベースの Cascade 型分類器で顔検出を行い, 顔識別をした. 識別結果は, 静止画上に顔の領域, 予測した登録名, 予測値を描画し, 動画に変換してディスプレイに出力した. サーバを利用したことで, ネットワークカメラへ顔を入力し, 識別結果を得るまでの処理時間は 1 フレームあたり約 0.125 秒となった. 表 4 にネットワークカメラの仕様を, 図 17 にネットワークカメラからの入力をリアルタイムに処理し, 出力している実験の様子を示す. 図 18 は, 識別対象者 R の識別結果 100 回中, 第 1 候補から第 3 候補として予測されたカテゴリの出現回数を積み上げ棒グラフで表したものである.

表4 ネットワークカメラの仕様

型番	TS-WPTCAM2 (IO)
カメラ	100 万画素 CMOS センサーカメラ
フォーマット (動画)	H.264
解像度	VGA (640×480)
フレームレート	15 (fps)
有線規格	100BASE-TX/10BASE-T



図17 リアルタイム顔識別実験

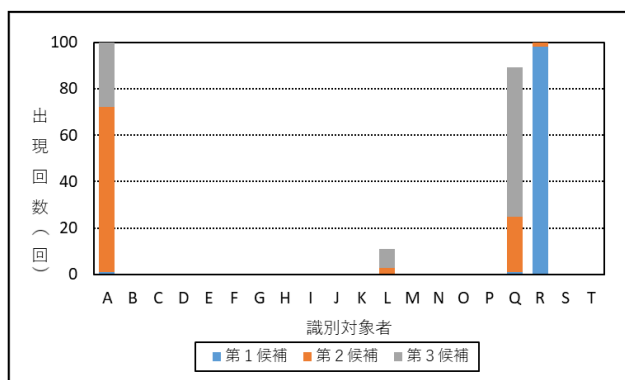


図18 第1候補から第3候補までの出現回数

図18より、100回の識別結果の中で、識別対象者Rが第1候補として最も多く出現していることが分かる。第2候補は識別対象者A、第3候補は識別対象者Qである。図12より、第2主成分までで見ると識別対象者RとAは隣接しており、特徴ベクトルの距離が近い第2候補となっていると考えられる。しかし、第3候補として挙げられている識別対象者Qはカテゴリ間の距離が離れているように見える。これは、入力画像と識別対象者Qの特徴ベクトルが類似している、もしくは、照合用の顔画像データベースに登録した識別対象者Qと識別対象者Rの特徴ベクトルが類似しているためではないかと考えられる。

6. おわりに

本研究では、日本人の学生を主な識別対象としたWebカメラを用いたリアルタイム顔認証システムの開発を目的に、カメラから入力した顔画像を、シャムネットワークを用いて顔識別し、リアルタイムに識別結果を画面出力した。その中で、GPUを利用したプログラムへ改良することで、学

習時間を約8分の1に短縮することができた。ネットワークカメラからの入力に対する識別時間は、1フレームあたり約0.125秒となった。また、日本人学生の学習用データセットを作成、シャムネットワークで小特徴空間を作成した。作成した小特徴空間を主成分分析し散布図にすると、学習前の特徴空間に比べ、シャムネットワークで学習した小特徴空間は、カテゴリごとが集中し、全てではないものの互いの距離が離れて分布している様子を見ることができた。作成した小特徴空間と大量の外国人顔画像で作成した大規模特徴空間の比較を行った結果、日本人学生の顔画像を識別するのに外国人顔画像を用いた大特徴空間では十分な識別精度が得られず、小特徴空間により精度は向上した。

今後の課題は、特徴空間の改善、処理速度の向上、照合用顔画像の選定の3つを挙げる。

特徴空間の課題として、カテゴリ同士の距離が離れ切れていない部分があった。カテゴリ同士の距離ベクトルが類似していると考えられる。改善のため、学習回数を増やす、学習時のパラメータを変更する、様々な表情画像を用意することが挙げられる。今回、収集した顔画像データは動画からの切り出したため似たような表情が多かった。様々な表情の画像を収集、学習することで、より個人の特徴が反映された分布になるのではないかと考える。処理速度の向上のため、照合用画像の特徴ベクトルでの作成を挙げる。現在、画像で保存している照合用顔画像データベースを特徴ベクトルで保存することで、識別時のニューラルネットワークでの特徴ベクトルの算出が未知データのみとなるため、処理時間の短縮につながると考える。最後に、照合用画像の選定を挙げる。今回、顔画像データベースに登録した画像は、必ずしも各カテゴリの中心に位置する画像であるとはいえない。これは、図18で示した実験のように離れて見えるカテゴリが上位の候補として挙がってくる要因の一つであると考えられる。今後、学習の過程で、カテゴリの中心となる画像を選定する工夫が必要だと考える。

謝辞

本研究を行うにあたり、産業・理科教育教員派遣研修事業として研修の機会を与えてくださった独立行政法人教職員支援機構に深く感謝いたします。

参考文献

- [1] 入退室管理・顔認証コラム顔認証システムがいま注目を集める理由とは？メリットと活用法を解説！、<https://www.nec-solutioninnovators.co.jp/ss/kaoato/column/02/> (参照 2021.8.23)
- [2] Joy Buolamwini, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, *Proceedings of Machine Learning Research* 81:1-15 (2018)
- [3] Koch, G., Zemel, R., and Salakhutdinov, R., Siamese neural networks for one-shot image recognition, In *Proceedings of the 32nd International Conference on Machine Learning*, 37 (2015). Available at *Siamese Neural Networks for One-shot Image*

Recognition on the ICML'15 website.

- [4] Haar Cascades を使った顔検出, http://labs.eecs.tottori-u.ac.jp/sd/Member/oyamada/OpenCV/html/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html
(参照 2022-01-05)
- [5] Math Work, シヤム ネットワークの学習とイメージの比較, <https://jp.mathworks.com/help/deeplearning/ug/train-a-siamese-network-to-compare-images.html#TrainASiameseNetworkToCompareImagesExample-13>(参照 2021.8.23)
- [6] Aditya Jain, <https://github.com/adityajn105/Face-Recognition-Siamese-Network> (参照 2021-05-26)