

## 日本語発話訓練のための顔画像処理

### 顔データベースの構築

中島 雄平<sup>†</sup> 梅田 耕佑<sup>†</sup> 菖木 禎史<sup>†</sup> 宇佐川 毅<sup>†</sup>

<sup>†</sup> 熊本大学工学部 〒860-8555 熊本県熊本市黒髪2丁目39番1号

E-mail: †{yuhei.kousuke}@hicc.cs.kumamoto-u.ac.jp, ††{chisaki,tuie}@cs.kumamoto-u.ac.jp

あらまし 顔特徴抽出は対象となる人物の認識や、口の動きの抽出などに非常に重要である。本稿では、遠隔発話訓練システムのための顔特徴抽出を目的とした、顔データベースの構築法を提案する。発話訓練のための重要な要素の一つである口唇領域の抽出は、その色情報が肌色情報と酷似しており、個人性もあるため困難である。利用者の照明環境等により、口唇を含む領域の決定は容易ではない。これらを考慮して口唇と瞳との比率の関係から口唇を含む標準的な領域を推定するため、本手法は、様々な照明環境において口唇よりもロバストに検出できる瞳を利用した。本稿では、構築したデータベースの有効性を、口唇領域の推定精度から検討した。

キーワード 発話訓練, 口唇領域, 顔データベース, 顔特徴, 瞳位置

## Image processing of learner's face for an utterance training system

### Construction of a face database

Yuhei NAKASHIMA<sup>†</sup>, Kousuke UMEDA<sup>†</sup>, Yoshifumi CHISAKI<sup>††</sup>, and Tsuyoshi USAGAWA<sup>†</sup>

<sup>†</sup> Faculty of Engineering & Graduate School of Science and Engineering  
Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan

E-mail: †{yuhei,kousuke}@hicc.cs.kumamoto-u.ac.jp, ††{chisaki,tuie}@cs.kumamoto-u.ac.jp

**Abstract** The extraction of the facial feature is very important for various purposes such as user's identification, lip movement extraction, and so on. In this paper, the method of building the face database aiming at the extraction of the facial feature for a distance training system of Japanese utterance is introduced. Since lip color information has very close resemblance to skin-color information and differs person by person, the extraction of the lips area which is very important for a distance training system is difficult. Because the lighting condition at learner's side of a distance learning system is very hard to control, the determination of area including lips is not easy. Based on these considerations, the rectangular area which includes lips is estimated from the proportional relationship between lips and a pair of pupils. This strategy is chosen because the robustness of the detection of pupil is easy comparing with one of the lips under various lighting conditions. In this paper, in order to determine the proportional relationship between lips and a pair of pupils, the face database is constructed, and the ratio of inclusion of lips in the estimated area is evaluated on closed-set data.

**Key words** utterance training system, lips area, face database, facial feature, pupils position

### 1. ま え が き

現在、国際化の進展や1990年の入管法(出入国管理及び難民認定法)の改正に伴い、外国から日本への入国者が年々増加する傾向にある。法務省の統計によると、平成14年度の入国者数は577万人で過去最高を記録している[2]。日本語学習者の数も増加しており、国内で約126000人(平成14年度文化庁調

査)、海外で約235万人(平成15年、国際交流基金調査)の外国人が日本語を学習している。このように、様々な国からの外国人の流入を受け、日本語学習者の多国籍化が進み、学習目的が多様化したことで様々な種類の日本語教材の需要が高まっている。

本研究では、特に教員が学習者全員に対して十分な訓練の時間を確保しづらいとされる、発話訓練の要素を含んだ、発話初

級日本語学習者を対象とした遠隔日本語発話訓練システムの構築を行っている [3].

一般に発話訓練においては、口唇領域を学習者に提示することで、効率的に学習できるとされており [4], [5], 口唇領域の抽出に関して、これまで相当数の先行研究がなされている。その中には、色情報を用い閾値を決め抽出する方法 [1], [6], [7] や、エッジ情報を用いて抽出する方法 [8] がある。その色情報が肌色情報と酷似しているため、学習者の個人性、訓練環境に十分対応することは困難である。また、顔の傾きによる顔特徴の形状変化に対し、ロバストに、かつ安定して口唇領域を検出するのは困難になる。これに対応するために、部分的テンプレートをを用いる方法 [9], 円形の分離度フィルタを用い、各特徴の端点を検出する方法 [10] が提案されているが、いずれもその大きさを変更しながらのマッチングとなるので、その演算量が大きくなってしまふ。他にも発話時のフレーム間差分により、口唇領域を抽出する方法も考えられるが、口唇領域の動きと同時に他の部位も不確定で動く可能性が高く、遠隔学習支援システムでの応用を考えた場合、その実用性は低い。発話訓練システムでは、そのレスポンスによる負担を学習者にかけないためにも、実時間で抽出する必要がある。そこで、口唇領域を抽出するのではなく、複雑な背景下でも照明条件の影響を受けずに比較的安定に検出できる両目を先に検出し、その位置情報から口唇領域を推定することで演算量の削減が見込める。従って、本稿では、口唇領域推定のために予め顔各部の相対位置情報をデータベース化することを試みた。

以下、口唇領域を推定するためのデータベースの構築法について述べる。さらに、構築したデータベースの有効性を、口唇領域の推定精度から検討する。

## 2. データベースの構築法

前述した通り、本稿では両目の重心の位置から口唇領域を推定するため、予め顔特徴を取得しデータベースとして保持することを目的としている。流れとしては、まずデータとなる被写体を撮影し、その後顔特徴点の位置情報を目視によりマウスを利用してデータとして入力する。次にその位置情報を利用してデータベース化したものを利用して口唇領域を推定するために各顔特徴パラメータの評価を行う。以下、この流れについての詳細を示す。

### 2.1 撮影環境

データベース構築において、照明条件など撮影する上で考慮すべき項目がいくつかある。今回データベースを構築するにあたり考慮した項目は、1) 背景、2) 照明条件、3) カメラ設定の統一性、4) 純粋な顔の保持の4つである。

第一の項目として、撮影における背景は、その均一性やその後のマウスによる各顔特徴を検出し、入力するため、色相環において人の肌色と対称な位置にある青色に設定し、人物の各顔特徴が明白になるようにした。第二の項目における照明条件においても同様に、人物の各顔特徴が明白になるよう、一般的な部屋の明るさといわれている 1200 Lx 前後に統一した。第三の項目として、各被写体においてカメラのズーム機能などカメラ



図1 撮影風景

表1 撮影に使用した器具

器具名	メーカー	個数
Qcam Orbit	Logicool	3
Qcam 4000 pro	Logicool	2
照度計 (ANA-F9 型)	テックジャム	1

表2 カメラ設定

	設定
画像サイズ	320 × 240[pixel]
ホワイトバランス	AUTO
水平方向	0[度]
上下方向	0[度]

表3 撮影用カメラの配置

上下方向 [度]	水平方向 [度]
20	0
10	5
0	10
-10	15
	20
	30

の設定を一定にし、カメラの設定の統一性を図った。カメラの位置については後述する。最後に第四の項目として、純粋な顔を保持するために、被写体の各顔特徴が髪や服、照明の影響などにより隠れてしまわないよう、髪を耳にかけるなどしてその影響の除去を行った。また、眼鏡や帽子による影響も除去した。

図1に、撮影風景を示す。

### 2.2 撮影カメラ

データベースの作成のために、撮影に使用した器具を表1に、その際のカメラの設定を表2に示す。画像サイズについては、本研究における発話訓練時のサイズと同様になるよう設定した。

### 2.3 カメラの位置について

撮影には3台の Qcam Orbit、被写体の頭部固定のための調整用には、2台の Qcam 4000pro を使用した。それぞれのカメラは図2に示す中心 O に、取得された画像の中心が来るように配置する。ここで中心 O とは、被写体の頭部の中心を想定し設定した点のことである。

撮影用カメラ (Qcam Orbit) の配置は、人の顔が左右対称で

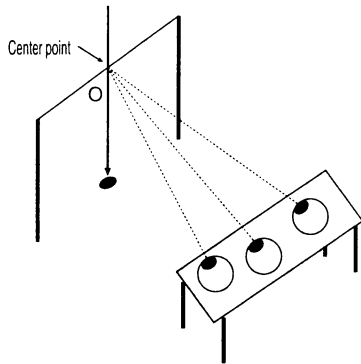


図2 撮影用カメラの調整

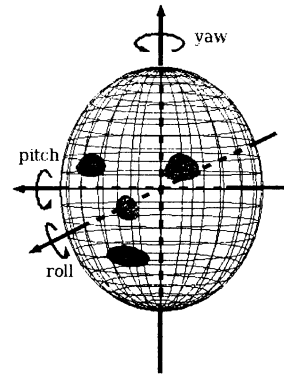


図5 3軸による頭部の動き

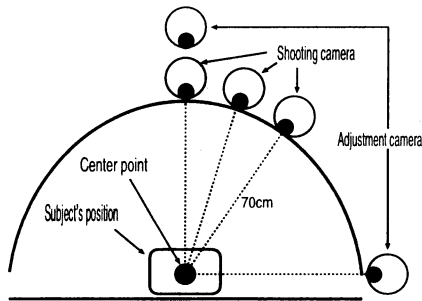


図3 上から見たカメラの配置図

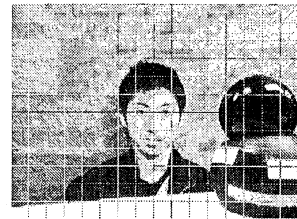


図6 正面の調整の様子

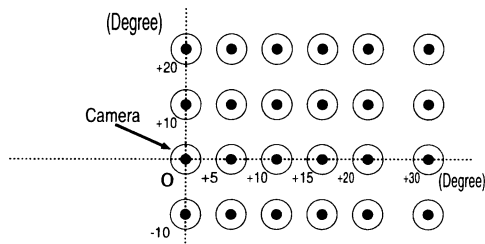


図4 被写体から見たカメラの配置図

あると仮定し、被写体から見て右方向のみ撮影した(図3、4)。また、中心Oまでの距離が一定になるよう配置し、表3に示す全ての組み合わせにおいて配置した。仮に上下方向が0度でない場合にも、常に中心Oが画像の中心になるように配置した。ここで中心Oまでの距離は、発話訓練時におけるディスプレイ上部に取り付けられるカメラからの距離を想定し70cmとした。調整用のカメラは、被写体から見て正面を水平方向0度とすると、それぞれ0度、90度になるようにし、なおかつ、被写体から見た仰角を上下方向としたとき、上下方向0度になるように配置した。

#### 2.4 顔の向きについて

被写体の頭の傾きには個人差があり、仮に被写体に正面をまっすぐ見るよう指示したとしても、それが必ずしも正面を向いているとは限らない。同様に首の動きによる上下方向の傾きも個人差が大きく影響している。被写体が正確に正面を向く手

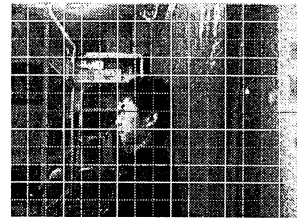


図7 90度方向からの調整の様子

法として、レーザポインタを被写体の頭部に取り付けて撮影する手法 [11] があるが、個人差による影響が大きくなってしまふ。そこで、被写体の正面を調整用カメラで撮影し、カメラの撮影画により実験者が確認し、被写体が正面を向いていることを確認することで、個人差を抑制した。その際、被写体の頭を壁につけることで固定をはかった。

図5に示すように一般に人の頭の動きは3軸からなっている。頭部の調整において、roll角及びyaw角は正面の調整用カメラを用いる。またpitch角においては、一般的に耳が頭を横から見て中心の垂直線上にあり、その高さが目尻のラインをそのまま後ろに伸ばした先が耳の付け根の上部になる [12]。従って、目尻と耳の付け根の上部の2点水平になるように90度の位置にある調整用カメラで調整する。正面の調整の様子を図6に、90度の位置にあるカメラでの調整の様子を図7に示す。

#### 2.5 撮影方法

被写体に、頭部が中心Oに来るように座ってもらうために、椅子は高さが可変なものを選択した。撮影は調整用カメラで頭

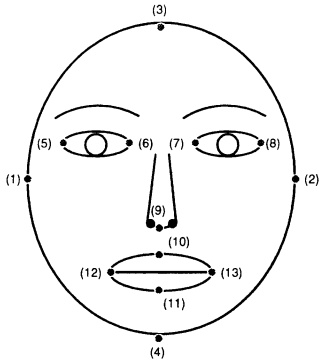


図 8 目視により抽出する顔の各特徴点

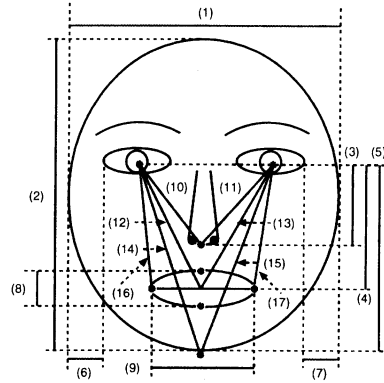


図 10 瞳の間の距離を 1 としたときの各顔特徴点間距離の比率

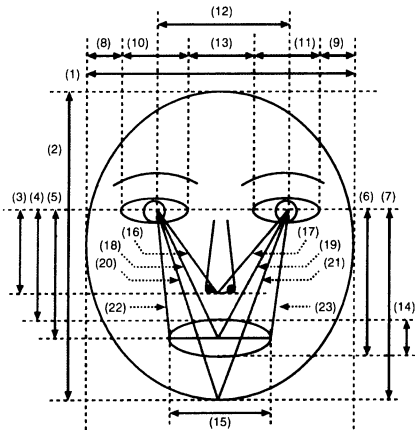


図 9 各顔特徴点間の距離

表 4 図 8 の各特徴点に対応する目視による抽出位置

番号	特徴点の抽出位置
1 ~ 4	顔輪郭において上下左右方向で最も端の部分
5 ~ 8	目尻、目頭部分において白目瞳(眼球)の端点
9	鼻の下と鼻との接点
10	上唇の M の部分の下端
11 ~ 13	唇の端点

部を固定後、撮影用カメラ 3 台の繋がった PC に同時にソケットを送ることで同時刻に行う。

## 2.6 顔特徴点

今回位置情報を取得する各顔特徴点を図 8 に示す。その際、再現性のあるデータベースとするため、抽出する特徴点を予め決めておく必要がある。特に目や鼻、口の位置情報は非常に重要な要素であるため注意しなければならない。各特徴点において、今回定義した顔特徴点の抽出位置を表 4 に示す。また、取得した顔特徴点の位置座標から距離を求め(図 9)、さらに瞳間(目尻及び目頭の中心を瞳とする。)の距離を 1 としたときの、各特徴点間距離の比率(図 10)を求め、これを利用し口の推定を行う。



図 11 予備実験用原画像

## 3. 予備実験

データベース構築をするにあたり、撮影条件などの有効性を示すため予備実験を行った。構築にあたり考慮した項目は、1) マウスのカーソルの変更、2) データ再現性、3) 画像サイズの検討、4) 左右対称性である。以降、この項目について検討する。図 11 は予備実験で使用した画像である。

### 3.1 マウスのカーソルの変更

クリックによる特徴点入力の際、デフォルトでのカーソルを用いると、クリックを行う人物の意識外の座標を入力することが多かった。そこで、マウスのカーソルを図 12 に示すものに変更してクリックすることで、より正確に特徴点を捉えることができるようにした。図 12 は入力する際に用いたマウスのカーソルである。図 11 に使用した画像を、図 12 に従来と新規カーソルによる入力した特徴点の座標を原画像にそれぞれプロットした画像を示す。図 13 において、従来のカーソルではクリックした座標が真値よりも若干右下に位置しているのに対し、新規カーソルでは、真値を正確にプロットしているのがわかる。表 5 に、従来のマウスのカーソルと提案のマウスのカーソルを使用して、2 人の実験者が男性画 2 枚、女性画像 1 枚に対して、それぞれ 3 回(計 18 回)の特徴点抽出を行った際の、特徴点座標位置のピクセル単位での分散を示す。

### 3.2 データの再現性

今回考案したデータベース構築法の再現性をみるため、3 人の被験者による入力について、



図 12 使用したマウスカーソル

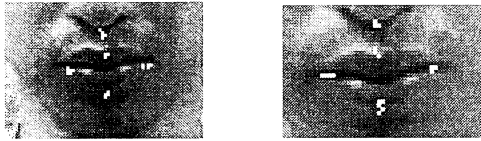


図 13 クリックした結果 (左: 従来, 右: 新規カーソル)

表 5 マウスカーソルによる分散 単位 (pixel)

カーソル	顔輪郭	目	鼻	口	計
従来カーソル	1.72	0.64	0.74	0.48	0.89
新規カーソル	1.97	0.29	0.18	0.49	0.73

表 6 特徴点を抽出する実験者による座標の分散 単位 [pixel]

人物	顔輪郭	目	鼻	口	計
A	1.19	0.33	0.24	0.62	0.6
B	3.44	0.29	0.15	0.37	1.06
C	1.28	0.25	0.15	0.47	0.54
計	1.97	0.29	0.18	0.49	0.73

表 7 被写体による座標の分散 単位 [pixel]

被写体	顔輪郭	目	鼻	口	計
a	1.97	0.29	0.18	0.49	0.73
b	0.47	0.51	0.14	0.39	0.37
c	0.38	0.57	0.39	1.47	0.70
計	0.94	0.46	0.24	0.78	0.6

表 8 画像サイズによる分散 単位 [pixel]

サイズ	顔輪郭	目	鼻	口	計
320x240	1.97	0.29	0.18	0.49	0.73
1280x960	5.81	4.63	1.38	2.91	3.68

- (1) 特徴点を抽出する実験者による座標の分散
- (2) 被写体による座標の分散

の二つをそれぞれ比較した。結果を表 6, 7 に示す。なお、被験者には予め入力する特徴点の場所の指導を行っている。この結果より、特徴点を抽出する実験者及び被写体による抽出特徴点の座標位置は 1 画素以下の分散で、安定に抽出できることが確かめられた。

### 3.3 画像サイズの検討

使用した画像サイズの適合性をみるため、1280x960 の画像でのクリックポイントの分散と比較・検討を行う。結果を表 8 に示す。これにより、画像サイズに比例した分散が出ていることから、画像サイズを小さい場合でも今回のデータベース構築には問題がないと言える。

### 3.4 左右対称性

データを有効に活用するため、人の顔が左右対称であると仮定し、撮影は片側だけの撮影としたが、本節ではその検証実験

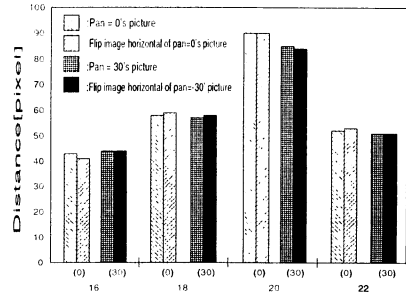


図 14 水平方向に対称な角度から撮影した画像を用いての図 10 の各パラメータにおける左右対称性の検討 ( () 内は角度)

を行う。実験には、上下方向を 0 度に設定し、水平方向の 0 度、± 30 度に配置した撮影用カメラで同時刻に撮影を行った。被写体は、成人大学生 5 名 (男性 3 名、女性 2 名) である。

まず、取得した一人あたり 3 枚の画像から、水平方向 0 度の画像 (h0) とそれを左右反転させた画像 (h0\_flip)、水平方向 30 度の画像 (h30) とそれを左右反転させた画像 (h30\_flip) を得る。

次に得られた 4 枚の画像に対して左右対称性を検証するため有効だと思われる、図 10 における 16~23 の距離情報を用いて、5 人の平均をとり、h0 と h0\_flip、h30 と h30\_flip の比較を行った。図 14 に結果を記す。

結果をみると、左右の対称性の最も低い部位で約 0.73 pixel であり、本研究での口唇領域の推定をする上では、十分な対称性があると結論される。このことから、データベース作成において、片側のみを撮影したデータでも十分に有効であるといえる。

## 4. データベースの評価

本章では、取得した各顔特徴の位置情報をもとに、口唇領域の切り出しサイズを決定すること、及びその適用結果を示すことにより、データベースの有効性を明らかにする。今回は両目からの相対的な口唇位置推定の妥当性検討のために閉口で検討するが、実際には閉口での口唇領域推定を行う。データベース作成における被写体は、成人の大学生 16 名 (男性 10 名、女性 6 名) であり、一人あたり 24 枚、計 384 枚の撮影を行った。その後、この画像を用いて各顔特徴をクリックした。表 9 に今回取得したデータベースの一部である比率についてのパラメータを示す。各パラメータの番号は図 10 に対応している。

### 4.1 口唇領域のサイズ

発話訓練への利用を考慮した場合、その口唇領域を囲む矩形のサイズは的確に学習者の口唇領域を捉えていなければならない。そのため今回得たデータベースの情報から、任意の角度から撮影した画像に対して適度なサイズの矩形を生成する。今回は、全データの 95% が含まれる値 (95 パーセンタイル) を求め、それを基準とした。各方向における 95 パーセンタイルを表 10、表 11、表 12 に示す。それぞれの表における値は、瞳の間の距離を 1 としたときの比率を表している。

### 4.2 適用結果

本節では、任意の角度から撮影した画像に対して、口唇領域

表 9 顔特徴パラメータ

Pan/Tilt	4	5	8	9	14	15
20/0	1.02	1.46	0.28	0.58	1.54	1.52
20/5	1.01	1.43	0.27	0.56	1.49	1.51
20/10	1.01	1.46	0.27	0.58	1.5	1.56
20/15	1.05	1.47	0.27	0.58	1.49	1.59
20/20	1.07	1.52	0.3	0.58	1.52	1.65
20/30	1.17	1.66	0.32	0.59	1.62	1.8

表 10 撮影角度による眉間から  
口唇領域の中心までの距離の比率 [%]

Tilt \ pan	0	5	10	15	20	30
20	1.15	1.07	1.17	1.18	1.23	1.34
10	1.19	1.16	1.17	1.27	1.31	1.35
0	1.11	1.16	1.17	1.21	1.19	1.44
-10	1.07	1.16	1.11	1.18	1.2	1.42

表 11 撮影角度による口唇の縦幅の比率 [%]

Tilt \ pan	0	5	10	15	20	30
20	0.39	0.36	0.34	0.33	0.35	0.4
10	0.4	0.43	0.36	0.34	0.36	0.42
0	0.34	0.37	0.36	0.36	0.39	0.45
-10	0.37	0.42	0.4	0.45	0.42	0.49

表 12 撮影角度による口唇の横幅の比率 [%]

Tilt \ pan	0	5	10	15	20	30
20	0.66	0.64	0.64	0.66	0.68	0.68
10	0.62	0.72	0.74	0.71	0.74	0.72
0	0.7	0.71	0.75	0.76	0.71	0.76
10	0.78	0.73	0.8	0.8	0.76	0.76

のサイズの決定と位置の推定が可能であるかを適用した結果を示す。本稿は、データベースの構築に着眼しているため、画像処理による瞳の検出で得た情報から、推定すべき口唇領域のサイズや位置を取得して推定を試みた場合、画像処理の抽出精度に依存した推定結果になってしまう。そこで、特徴点入力によりその適用結果をみる。具体的には、3次元で顔の方向が推定できるように仮想的に顎が検出されたとして、両目と顎をクリックし、そこからデータベースを参照する。その後、特徴点入力した位置情報と最も近い値をもつ撮影角度を取得する。その撮影角度において、開口時の口唇領域のサイズを決定し、その画像に対し適用した。適用した画像を図 15 に示す。上下ともデータベースを適用し、口唇領域がほぼ囲まれていることがわかる。従って、データベースの有効性が示された。

## 5. まとめ

本稿では、遠隔発話訓練システムのための顔特徴抽出を目的とした、顔データベースの構築法を提案するとともに、構築したデータベースの有効性を、口唇領域の推定精度から検討した。構築した顔データベースは、瞳などの顔特徴の位置情報を持ち、さらに、本データベースを基に瞳位置から口唇領域推定が可能

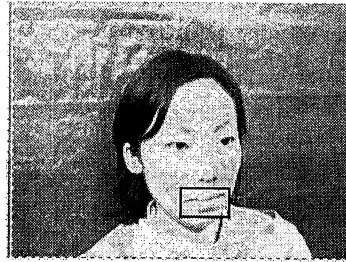


図 15 上下方向 10 度、水平方向 20 度の画像に対し、  
口唇領域推定を適用した画像

であることをデータを用いて示した。

今後の課題としては、学習者の発話時における口唇領域のサイズの決定、学習者の瞳の検出から口唇領域を抽出すること、加えて 3次元モデルとの融合も考慮した瞳以外の第 3 点の検出が考えられる。

## 文 献

- [1] 森本 広頼, 李 亜兵, 亀井 且有, 井上 和夫, “個人差や撮影環境を考慮した濃淡画像からの口唇領域の抽出”, 信学論 D-II, Vol. No.1 J79-D2, pp.139-141, 1996
- [2] 法務省大臣官房司法法制部(編),「第 41 出入国管理統計年報(平成 14 年版)」, 国立印刷局, (2003).
- [3] 中川 彰, 中島 雄平, 上野 歩美, 山川 仁子, 菅木 禎史, 宇佐川 毅, “音声認識を応用した遠隔日本語発話訓練システムに関する研究”, 熊本大学 大学院 修士論文 (2004.2).
- [4] 文化庁, “日本語教育指導参考書 1 音声と音声教育”, 大蔵省印刷局, 東京, 1998
- [5] 竹内 理, “認知的アプローチによる外国語教育”, 松柏社, 東京, 2002
- [6] 寺田 賢治, 山中 理聖子, 大患 俊一郎, “口のカラー動画像を用いた音韻認識”, 電学論 D, vol.119, No.1, pp.37-43, 1999
- [7] 中田 康之, アンドウ 護俊, “色抽出法と固有空間法を用いた読唇処理”, 電学論 D-II, Vol. J85-D-II, No.12, pp.1813-1822, 2002
- [8] Brunelli and T.Possio, “Face Recognition: Features versus Templates”, IEEE Trans PAMI, 15.10, 10, pp.1042-1052(1993)
- [9] 宋欣光, 李七雨, 徐剛, 辻三郎, “部分特徴テンプレートとグローバル制約による顔部品特徴の自動抽出”, 信学論, J80-D-II, 8, pp.2178-2185(1997)
- [10] 林 健太郎, 橋本 学, 鷺見 和彦, “頑強性と精緻性を備えた顔特徴点追跡による顔方向推定”, 信学論, D-II, Vol. J84-D-II, No. 8, pp. 1762-1771, 2001
- [11] 菊地 貴行, 小柳 剛, 佐藤 美恵, 春田 正男, “個人認証のための正面顔検出に用いる特徴量の検討”, 芸術学論, Vol.1, No. 3, pp. 132-133, 2002
- [12] 飯島 貴史, “CG デザイナーのための「人体のしくみ」”, ワークスコーポレーション, 東京, 2003