

# 遠隔発話訓練システム用顔画像処理アルゴリズムの開発

○ 中島 雄平 上野 歩美 中川 彰 山川 仁子 荳木 禎史 宇佐川 毅 (熊本大学)

## 1 はじめに

現在，社会の国際化や 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は年々増加する傾向にある。その中で，日本語学習者の数も増加しており，国内で約 9500 人（平成 12 年度文化庁調査），海外で約 210 万人もの外国人が日本語を学習している（平成 10 年，国際交流基金調査）。

このように日本語学習者の増加が進んでいるが，日本語教育の現場では教師と学習者が一対多の関係となっており，かつ授業は主に文法と筆記が中心で，聴取訓練は学習者の個人的な習得度とは独立に実施されている。特に発話訓練については，教師が学習者全員に対して十分な訓練時間を確保することは事実上不可能に近い。

本稿は，日本語発話訓練システムの開発を目標とした一連の研究 [1][2] の一部で，本システムの学習者の発話訓練時の動画処理部を構成するための顔画像処理アルゴリズムについて述べる。

提案のシステムでは，発話時の学習者の顔画像をディスプレイ上部に取り付けたカメラで取得し，発話画像を学習者に提供することにより，発話の訓練に活用する。その際，教師の発話画像と音韻毎に同期させて，口唇の動きを学習者に提示することを予定している。提案システムの利用環境は，CALL 教室や学習者の自宅など，背景画像は非常に複雑であることが想定される。

本稿では，複雑な背景画像から高速に顔画像を抽出できるアルゴリズムを提案する。

## 2 提案アルゴリズム

入力画像は学習者から送られてきた 256 階調の RGB 画像（320x240 画素）を想定している。抽出アルゴリズムは主に色情報処理と矩形生成処理の 2 つの処理で構成されている。

### 2.1 色情報処理

色情報処理では学習者から送られてきた画像から，学習者の顔の候補領域を抽出する。人間の肌の色は，人種・民族の多様性に関わらず色相がほぼ一定で，一般的な背景などに比べ比較的色彩度が高いという特徴がある。

そこで，YIQ 変換と HSV 変換のうち，I 信号画像，Q 信号画像，H 画像，S 画像の 4 つの変換画像を用いて人の肌色を抽出した。

YIQ とは，NTSC(National Television System Committee) 方式のカラーテレビ放送で使われるカラーモデルであり，Y 信号は輝度信号，I 信号は肌色を含むオレンジからシアンまでの色調，Q 信号は I 信号以外の色調を示している。HSV 変換における H は色相 (Hue)，S は彩度 (Saturation)，V は明度 (Value) を表す。

RGB 値から YIQ 値に変換する式と RGB 値から HSV 変換式を以下に示す [3] [4]。

$$Q = 0.211R - 0.523G + 0.312B \quad (1)$$

$$I = 0.596R - 0.274G - 0.322B \quad (2)$$

$$H = \cos^{-1} \left\{ \frac{(R - G) + (R - B)}{2\sqrt{(R - G)^2 + (R - B)(G - B)}} \right\} \quad (3)$$

$$S = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)} \times 100 \quad (4)$$

ここで， $\max()$ ， $\min()$  はそれぞれ  $(R, G, B)$  の最大値，最小値を返す関数である。

### 2.2 矩形生成処理

矩形生成処理では，色情報処理によって得られた顔領域候補の中から顔領域を選択し，矩形で囲む処理を行う。

ここで，学習者はディスプレイの前に座っているものと想定しており，予備実験により顔領域の最小の大きさを画像全体の 5 分の 1 であるとし，これを閾値とした。まず画像全体に水平方向の走査線を設け，走査線上で閾値以上連続して顔領域候補が存在する線分を探索し，かつその中で最小と最大の x 座標 ( $x_{\min}$ ， $x_{\max}$ ) を取得する。

次に得られた x 座標の範囲内で，垂直方向の走査線を設け，さきほどの水平方向と同様に，閾値以上連続して顔領域候補が存在する線分を探索し，その中で最小と最大の y 座標 ( $y_{\min}$ ， $y_{\max}$ ) を取得する。

最後に 4 つの座標から矩形の 4 端点を求め，その情報をもとに矩形を生成する。

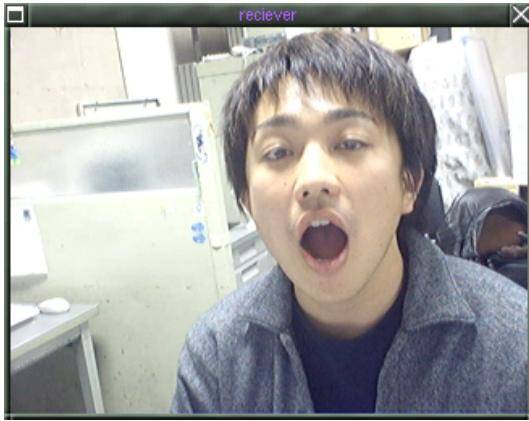


図 1: 原画像

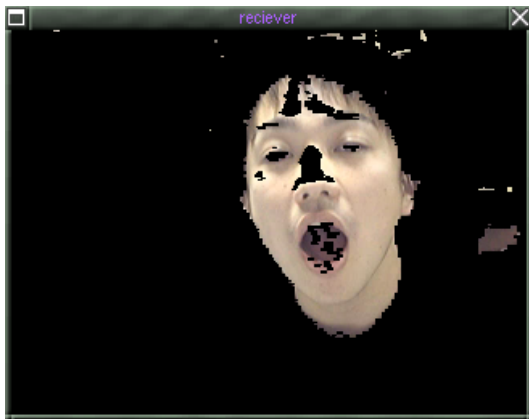


図 2: 色情報処理画像

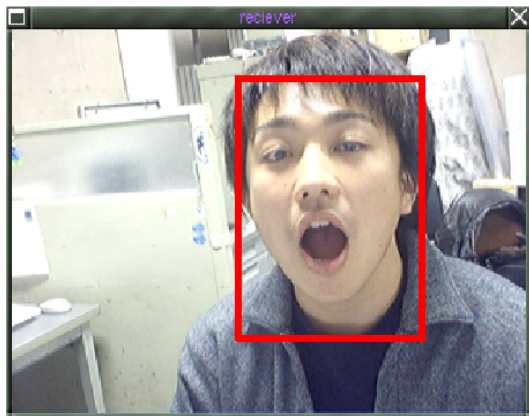


図 3: 矩形画像

### 3 実験とその結果

本提案アルゴリズムの有効性を確認するために実験を行った。実験は CPU が 1.7G Hz の Linux 環境下で、カメラには Logitech 社の”Q-cam 4000 Pro”を使用した。入力画像は、複雑背景において撮影された、320x240 画素、各色 256 階調のカラー画像である。

入力画像の例として任意の時刻でのフレームを図 1 に示す。図 2 は、同フレームにおける入力画像から色

情報処理によって得られた顔領域候補である。色情報処理において背景など顔領域以外のほとんどの領域が排除されているのが分かる。ただし、鼻は光の影響を他の部位よりも受けやすいので、顔領域候補としては抽出されなかった。図 3 は、矩形生成処理において得られた顔領域である。矩形で囲まれている部分が顔領域となる。

図 3 は任意の時刻でのフレームの抽出結果であるが、視察によりほぼ全てのフレームにおいて顔領域と思われる領域を抽出することができた。また、10.10 fps で動作したことが確認できた。

### 4 まとめ

本報告では、すでに提案してしているシステムにおいて、学習者に対し拡大した口唇領域の画像を提示することを目的として、学習者側から送られてきた画像を用いて顔領域の抽出を行うアルゴリズムについて報告した。

本報告における抽出アルゴリズムは、エッジ情報ではなく色情報をもとに顔領域を抽出しているため、複雑な背景下においても精度良く顔領域を抽出できた。また、色情報のみを用いていることから実時間での動作も確認できた。

今後の課題として、学習者により効率良く学習してもらうために、動画の時間伸縮を行い、教師画像と学習者の画像の時間軸を変化させ同期をとった口唇画像の提示や、目や口唇領域の抽出を行うことが挙げられる。

### 参考文献

- [1] 中川 彰, 上野 歩美, 中島 雄平, 山川 仁子, 菅木 禎史, 宇佐川 毅, ”日本語初習者を対象とした日本語遠隔学習システム構築の試み”, 信学技報, EA2003-114, 2003.11
- [2] 上野 歩美, 中川 彰, 菅木 禎史, 宇佐川 毅, “音声認識システムを用いた発話訓練システム構築の試み”, 第 5 回 学生のための研究発表会, 講演論文集, pp.25-28 (2003).
- [3] <http://www.cs.unr.edu/looney/cs674/ch6html.htm>
- [4] <http://zope.lab.tkikuchi.net/edu/image/07Color.html>

問い合わせ先:

熊本大学 菅木 禎史, 宇佐川 毅

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

Tel: 096 - 342 - 3621