読唇のための色特微量に基づくエネルギーを用いた動的口唇輪郭抽出法

正員 小越 康宏* 非会員 井出 寿登*,**

非会員 荒木 睦大* 正員 木村 春彦***

Active Lip Contour Using Hue Characteristics Energy Model for A Lip Reading System

Yasuhiro Ogoshi*, Member, Hisato Ide*,**, Non-member, Chikahiro Araki*, Non-member,
Haruhiko Kimura***, Member

Active contour model (SNAKES) is very used as one of the powerful technique in a contour extraction that utilizes principle of energy-minimizing. Performing extraction of lip contour with the lip image that has strong edges or noises on the lips and oral cavity is an important problem. This paper proposes a new energy model of SNAKES based on hue characteristics of lip images.

キーワード：動的輪郭モデル (SNAKES), 初期輪郭, エネルギー最小化法, HSV 表色系

Keywords: SNAKES: active contour model, initial contour, energy-minimization method, hue characteristics

1. まえがき

発話時の映像から口形変化を分析し発話内容を推測する読唇システムがあり、発声が困難な人の意思伝達手段、騒音下での音声認識補助手段として期待が持たれている⁽¹⁾⁽²⁾。

読唇システムでは高速かつ正確な口唇輪郭抽出が要求されるが、発話時に口唇形状が大きく変化するので難しい。

輪郭抽出において代表的な動的輪郭モデル (SNAKES: Active Contour Model)⁽³⁾ は、対象物付近の閉領域を安定して抽出でき、対象物の形状や濃度勾配に応じて様々なエネルギー関数を導入した派生モデル⁽⁴⁾ が提案されてきた。

SNAKESに共通する問題として以下の3つが考えられる。

- (1) 目的の輪郭外にあるノイズを捉える (図1(a))。
- (2) 目的の輪郭内にある強いエッジを捉える (図1(b))。
- (3) 初期輪郭は、収束ステップ数と抽出結果に影響を与える。

本研究では、これらの問題点を解決すべく、新しいエネルギー関数と初期輪郭の設定方法を提案する。

2. 一般的な SNAKES の定義

SNAKESは制御点 $v(s) = (x(s), y(s))$, $s = 1, 2, \dots, S$ により閉ループで構成される輪郭モデルである。目的とする画像上の輪郭に対し、外側に初期輪郭となる SNAKES を与え、下式に与えるエネルギー関数の総和 E_{snakes} を最小化しながら変形を繰り返して目的の輪郭に近づける。

$$E_{snakes} = \sum_{s=1}^S \sum_{i=1}^I \{W_i \times E_i(v(s))\}$$

* 福井大学大学院工学研究科

Graduate School of Engineering, Fukui University

** 現在, (株) ミマキエンジニアリング 勤務

Mimaki Engineering Co., Ltd.

*** 金沢大学大学院自然科学研究科

Graduate School of Science and Technology, Kanazawa University

(a) Influence of noise (b) Excessive action

Fig. 1. Lip contour image (poor control)

一般に目的とする輪郭の性質に応じ、様々なエネルギー関数 $E_i(v(s))$, $i = 1, 2, \dots, I$ が定義される。また、各エネルギー関数 $E_i(v(s))$ に、重み W_i を付加し作用を調整する。代表的なエネルギー関数 $E_i(v(s))$ を以下に列挙する。

1次微分のエネルギー関数 $E_1(v(s))$ を下式に示す。これは曲率を表し直線のとき最小となる。

$$E_1 = \sum_{s=1}^S (\|v(s+1) - v(s)\| + \|v(s-1) - v(s)\|)$$

2次微分のエネルギー関数 $E_2(v(s))$ を下式に示す。これは曲率の変化を表し、直線や円のとき最小となる。

$$E_2 = \sum_{s=1}^S \sqrt{(v(s+1) + v(s-1) - 2 \times v(s))^2}$$

エッジエネルギー関数 $E_3(v(s))$ を下式に示す。画像に対し Sobel フィルタを用い水平微分 $Sobel_h(x, y)$, 垂直微分 $Sobel_v(x, y)$ を考慮したものであり、各制御点が強いつまみとなる輪郭上にあると負に大きな値となる。

$$E_3(x, y) = -\sqrt{Sobel_h(x, y)^2 + Sobel_v(x, y)^2}$$

3. 提案するエネルギー関数の概要

〈3・1〉 新しいエネルギー関数を求めるための前処理

原画像 (図2(a)) を RGB から HSV 表色系に変換し、HSV の値を基に2値化画像を求め口唇領域を抽出する (図2(b))。2値化画像では口唇領域外に口唇に近い色特徴を持った部分が点在しているため、広く知られている輪郭線長追跡を用いノイズ除去する (図2(c))。

(a) Original image (b) Binary image (c) Noise reduction

Fig. 2. Binary image of the lip

(a) SD_y vs Height (b) SD_x vs Width (c) SD_y vs Width

Fig. 3. SD vs lips size

〈3・2〉新しいエネルギー関数の定義

口唇領域の画素を $p_1, p_2, \dots, p_N$ とおき、これらの x, y 成分から標準偏差 SD_x, SD_y 、重心 GP_x, GP_y を求め、口唇の横、縦のサイズを表すパラメータ $Width, Height$ を近似式から求める。

日本語 5 母音の発話を撮影した 146 フレームにおける $Width, Height$ と SD_x, SD_y の相関関係を図 3 に示す。

$Height$ と SD_y には強い相関 (0.9784) があり (図 3(a))、下式の様に近似式を求め、 t -検定の結果 $P=0.9998$ となった。

$$Height = -0.0188 \times SD_y^2 + 4.4444 \times SD_y - 4.3327$$

また、 $Width$ と SD_x の相関は低い (0.3550) が (図 3(b))、 $Width$ と SD_y の相関は比較的高く (-0.5644) (図 3(c))、近似式を α とおき、 SD_x を考慮し下式の 2 変数関数の様に近似式を求め、 t -検定の結果 $P=0.9818$ となった。

$$\alpha = (-0.0230 \times SD_y^2 + 1.1026 \times SD_y + 169.90)$$

$$Width = \alpha \times SD_x / 41.0$$

なお、これらの近似式は人物の違いにより再調整を要する。

次に、口唇外周輪郭の形状に近い楕円を想定し、重心から楕円上までの距離を $Dist(x, y)$ を下式に与える。

$$Dist(x, y) = \left\{ \frac{x - GP_x}{Width/2 + d} \right\}^2 + \left\{ \frac{y - GP_y}{Height/2 + d} \right\}^2$$

$Dist(x, y) = 1.0$ で口唇外周輪郭とほぼ重なるが、楕円が口唇外周輪郭外に出てしまうと、輪郭形状が失われるため、楕円サイズを調整するパラメータ d を考え、 $d = -10$ 、 $Dist(x, y) \leq 1.0$ と設定し、経験的に口唇外周輪郭から 10 画素分だけ内側に楕円領域を求める。この楕円領域を用い下式の様に新しいエネルギー関数 $E_4(v(s))$ を定義する。

$$E_4(v(s)) = (255 - l) \times (1.0 - Dist(v(s))) + l$$

l はエネルギーの値を調整するパラメータであり、重心位置で輝度値 255、楕円外の口唇領域で輝度値 l とすることで、重心から外側にエネルギーが作用する。内外で濃淡勾配を大きくとり、かつ、外側である程度のポテンシャルを持たせた方が良いため、経験的に l の値を 100 と設定する。

以上に従い、図 4(a) に楕円モデル、図 4(b) に 2 値化画像と楕円モデルを合成したエネルギー関数 $E_4(v(s))$ を示す。

〈3・3〉初期輪郭の設定

初期輪郭となる制御点 $v(s)$ は下式により楕円状に与える。

$$x(s) = GP_x + (Width/2 + d) \times \cos(\theta)$$

(a) Ellipse model (b) Energy model (c) Initial contour

Fig. 4. New energy model and initial contour

Fig. 5. Result of the lip contour extraction

$$y(s) = GP_y + (Height/2 + d) \times \sin(\theta)$$

SNAKES の収束ステップ数を考慮すると、初期輪郭は口唇外周輪郭に近いほど良く、経験的に $d = 6$ とし想定される輪郭より 6 画素外側に与える (図 4(c))。

また、制御点が等間隔に並ぶように最大制御点間距離 f を考え、変形中に制御点同士が重なるか交差したら一方を削除、 f よりも離れたら中間に制御点を追加する。 f は小さいほど細かな輪郭形状の変化に対応できる。

4. 実験結果

SNAKES のパラメータは、 $W_1=100$ 、 $W_2=80$ 、 $W_3=2$ 、 $W_4=100$ 、 $f=5$ と設定した。2 台の蛍光灯を顔に対し上下斜め 45° から照射し、顔から 30cm の位置にカメラを固定し正面顔を撮影する。日本語 5 母音発話時の代表的な 6 フレームの輪郭抽出結果を図 5 に示す。左から順に、閉口時、母音「あ」～「お」の最大開口時を示す。図の上段の様に適切な SNAKES の初期輪郭を与えることができ、図の下段の様に良好な SNAKES の収束結果が得られた。

SNAKES の変形に要したステップ数は平均 6.7 となった。

5. まとめ

提案手法により、SNAKES の高速化と精度の向上が期待できるようになった。得られた SNAKES からアスペクト比や口唇面積を求めることで読唇への応用が期待できる。(平成 19 年 10 月 11 日受付、平成 19 年 12 月 6 日再受付)

文 献

- (1) K. Kiyota and K. Uchimura: "An Uttered Word Recognition Using Lip Image Information", *Trans. IEICE*, Vol. J76-DII, No. 3, pp. 812-814 (1993-3) (in Japanese)
清田公保・内村圭一:「口唇周辺画像情報を用いた発話単語認識」, 信学論, **J76-DII**, 3, pp. 812-814 (1993-3)
- (2) Y. Sasaki, T. Kawamura, and K. Sugahara: "Hardware Realization of Active Contour Model and Its Application for Vowel Recognition by Lip Reading", *Trans. IPSJ*, 2005-SLDM-118(16), pp. 87-91 (2005-1) (in Japanese)
佐々木悠介・川村尚生・菅原一孔:「動的輪郭モデルのハードウェア化とその読唇母音認識への応用」, 情処研報, 2005-SLDM-118(16), pp. 87-91 (2005-1)
- (3) M. Kass, A. Witkin, and D. Terzopoulos: "Snakes: Active Contour Models", *International Journal of Computer Vision*, pp. 312-331 (1988)
- (4) T. Wakasugi, M. Nishimura, O. Yamaguchi, and K. Fukui: "Lip Contour Extraction Using Separability of Color Distributions", *Trans. IEICE*, Vol. J89-DII, No. 9, pp. 2025-2032 (2006-9) (in Japanese)
若杉智和・西浦正英・山口 修・福井和広:「色分布間の分離度を用いた唇輪郭抽出」, 信学論, **J89-DII**, 9, pp. 2025-2032 (2006-9)