

ハードウェア実現した動的輪郭モデルによる唇の外側および内側の形状抽出について

Hardware Realization of Active Contour Model and its Application for Outer and Inner Lip Shape Extraction

佐々木 悠介[†]川村 尚生[‡]菅原 一孔[‡]

Yusuke Sasaki

Takao Kawamura

Kazunori Sugahara

概要

本稿では高雑音下でのロボットなどのシステムの制御に、読唇手法を用いるための手法について検討する。読唇の際、発話時の唇形状の抽出には、動的輪郭モデルをハードウェア化したものを用いる。動的輪郭モデルでは画像を記録しているメモリとのアクセスが他の手法に比べ極端に少なく済むため、高速な動作が可能であり、組み込み型のシステムに向いているなどの特徴がある。しかし他の手法と同様に、唇以外の部分、例えば鼻やあごの部分を唇の一部としてとらえてしまうなどの誤動作もあり、これが正確な読唇を行う際の妨げとなっていた。本稿では、同じ顔画像に動的輪郭モデルを2回適用することにより、唇の外側と内側の形状を抽出しそれによる読唇手法を提案する。実験結果によりその有効性を確認する。

1. はじめに

工場内など高雑音環境下でも、製造装置や搬送用ロボットなどを音声により制御したい場面がある。このような場面では、大語彙の音声認識は必要なくある程度の語数の単語を識別できれば十分な場合も多い。しかし、現在各所で研究が進んでいる音声認識手法では、周囲の雑音が少ない場合には有効であるが、高い雑音環境では認識率が極端に低下してしまう問題がある。この点、人の発話時の唇形状を認識する、いわゆる読唇手法では周囲の雑音の影響はまったくなく、また手話などを使う方法に比べると、人が特別な訓練をする必要がないなどの有利な点があり有効な手法のひとつと考える。

読唇システムを構築するためには、入力動画から高速に唇形状を抽出するための手法が不可欠である。画像中の色のヒストグラムを集計し、それにより唇領域を抽出するなどの従来手法では、画像を蓄えているメモリへのアクセス数は、画像中の画素数に相当しその数は膨大である。メモリへのアクセスには、最低でも1クロック必要であり、結果としてアクセス数の多い従来手法の動作速度は、低下してしまっていた。この点、動的輪郭モデルによる手法では、メモリへのアクセスは少なく済み、高速な唇形状抽出処理が期待できる。

本稿では読唇によるシステムの制御を目指し、動的輪郭モデルより唇の外側と内側の形状を抽出する手法を提案する。また、本手法の様々なシステムへの組み込みを目指し、FPGA上にハードウェアとして実現した。それによる唇形状抽出結果を示し、本手法の有効性を確認する。

2. 動的輪郭モデル

Kassらは画像中のエネルギーの最小化により、画像中の特定領域を抽出する動的輪郭モデル(Snakes)を提案し、その後様々な画像エネルギーを考慮した手法が提案された[1]。しかしこれらの手法では、エネルギーの最小化に多大な計算処理が必要であり、本稿で扱う組み込みシステムへの応用は困難である[2]。その後、エネルギーの最小化を、動的輪郭モデルに加わる力の釣り合いとして捕らえるSampled-ACMが橋本らにより提案され[3]、さらに菅原らにより振動項を考慮することにより耐雑音性能を向上させたものが提案された[4]。本稿では、振動項を考慮したSampled-ACMを用いて唇形状を抽出することにする。

振動項を考慮したSampled-ACM(以下、動的輪郭モデル)では仮想的な閉曲線上の複数の動作点に、圧力、引力、反力および振動項と呼ばれる4つの力が動作点に働くことにより、閉曲線が収縮し領域を抽出するが、本稿で抽出対象である唇の形状がおおむね全領域にわたり外側に凸であるという特徴を考慮し、圧力は考慮せず引力と反力および振動項により収縮動作をするものとした。

引力は図1に示すように隣り合う2つの動作点間に働く力であり、その間の距離に比例した大きさを持つものとした。振動項は引力の合力 F_a に対し直角方向に働く力であり、収縮のたびにその方向を反転する。なお、この振動項の大きさは一定の値 F_v を持つものとした。

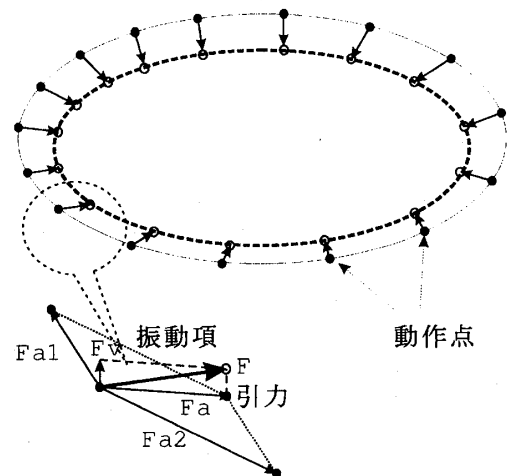


図1: 引力と振動項による収縮動作

図2に示すように、反力は動作点が対象の画像領域に接した際に働く力であり、引力 F_a と振動項 F_v の合力の

[†]鳥取大学大学院 工学研究科 知能情報工学専攻

[‡]鳥取大学 工学部 知能情報工学科

抽出領域に対する垂直成分を打ち消す働きをもつ。ただし、反力はその大きさに閾値を持ち、ある一定以上の垂直成分は打ち消すことができないものとする。これらの力の働きにより、画像中の雑音をすり抜けたり、あるいは突き抜けたりする動作を実現することが可能となり、画像中の雑音に強い領域抽出手法となる。

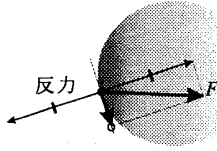


図 2: 反力の働き

3. 唇の外側および内側形状の抽出

入力画像に動的輪郭モデルを適用すると、図 3 に示すとおりに唇の外側形状を抽出することができる。なおこの場合、動的輪郭モデルの初期輪郭としては画像全体を囲むように配置した。動的輪郭モデルに基づく手法では画像が記録されているメモリへのアクセスは、動作点が移動する画素のみですみ、色による判別手法などの他の手法に比べ極端にメモリのアクセス数を低減させることが可能である。

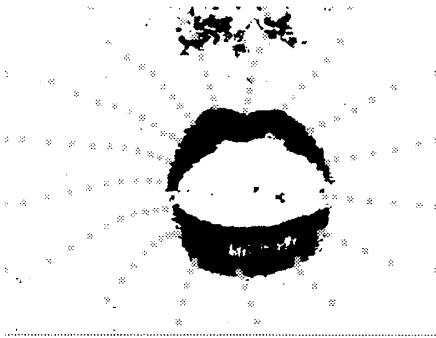


図 3: 唇の外側形状を抽出している様子

続いて、唇の内側の形状を抽出するために、顔画像の白黒を反転した画像に再度動的輪郭モデルを適用することを考える。ただし、2回目の動的輪郭モデルの動作点の初期位置は、1回目の収束結果を用いることにする。唇の内側の形状を抽出している様子を図 4 に示す。

このようにして得られた母音発話時の唇の外側および内側形状の抽出例を図 5 に示す。

しかし、唇の外側形状の抽出の際に画像の明るさや陰の影響により、鼻やあごの部分を唇の一部としてとらえてしまうこともある。その様子を図 6 に示す。このような抽出結果をもとに読唇を行うと、その識別率の低下を招くことが予想される。この点、図 6 のように、鼻やあごの部分を唇の一部としてとらえてしまった場合でも、顔の頬の部分については、それを唇ととらえてしまう動作例はほとんど無いことから、唇の幅はある程度正確に求められえていることが多い。

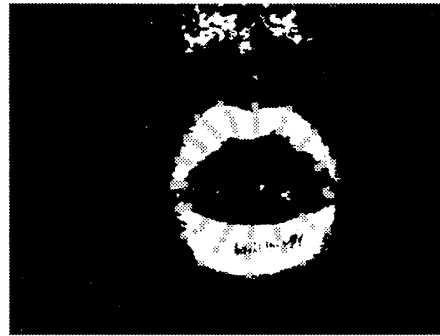


図 4: 唇の内側形状を抽出している様子

つぎに、唇の内側形状を抽出する場合、鼻やあごの部分に引っかかり正しく唇の外側形状を抽出できていなかった場合でも、あごと唇の領域は連続している場合が多いことなどから、唇内側下半分の形状は比較的正確に抽出できている。

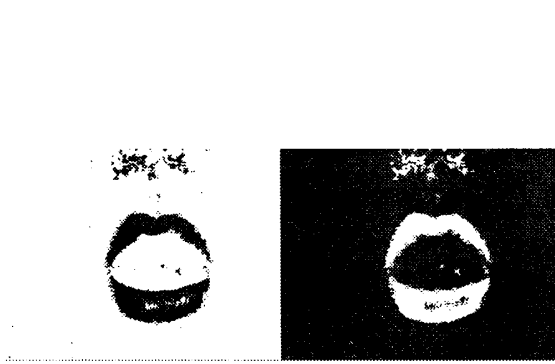
4. 唇形状抽出システムのハードウェア実現

3. で述べた唇形状抽出手法を、様々なシステムへ組み込むためにそのハードウェア化を試みた。ハードウェア化は FPGA 上に行うこととし、画像入出力回路、FPGA と画像メモリを搭載した装置を開発した。画像の入出力は NTSC 規格に準拠した信号を取り扱うこととし、また FPGA には ALTERA 社の APEX20KC EP20K200CF484C8 を用いた。この FPGA のロジックエレメント数は 8,320 であり、20 万ゲートに相当する規模の FPGA である。システム全体の構成を図 7 に示す。

3. で述べた読唇手法ならびに画像入出力回路の制御回路を実装した結果、65%のハードウェア量でシステムを実現できた。動的輪郭モデル部の 1 フレームの処理に必要なクロック数は約 267,000 であり、FPGA のクロックが 48MHz である条件下で毎秒 30 フレームの動画像を処理するに十分な処理速度を実現できた。またメモリアクセス数については、外側形状抽出の最大収束回数を 50 回、内側形状抽出の最大収束回数を 30 回としたとき、最高でも約 15,000 回となり、他の手法に比べ大幅に低減できた。

5. おわりに

本稿では、発話時の顔画像に動的輪郭モデルを 2 回適用することにより唇の外側および内側形状の抽出手法を提案した。提案した手法は FPGA 上にハードウェア実現し、唇形状の抽出実験をした結果、その有効性を確認した。今後の課題として、本稿で得られた唇形状を基に単語を認識するシステムの開発についての検討が残されている。



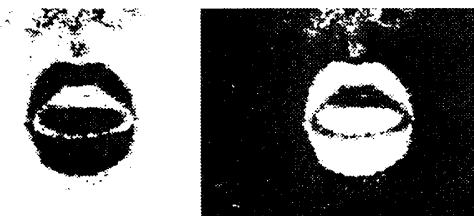
(a) あ



(b) い



(c) う



(d) え



(e) お

図 5: 母音発話時の唇の外側および内側形状の抽出例



図 6: 唇の外側形状の抽出の失敗例

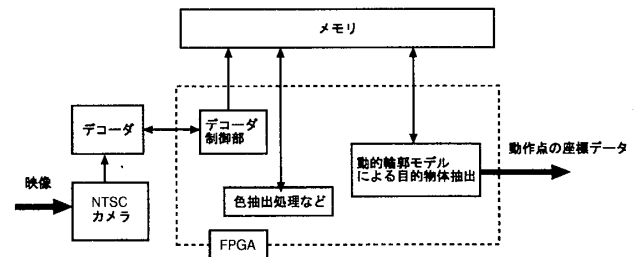


図 7: システム構成

参考文献

- [1] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active Contour Models," International Journal of Computer Vision, pp.321-331, 1998.
- [2] 須賀 弘道, 羽鳥 好律, 植松 明, "SNAKE を用いた顔画像からの構成部品の輪郭抽出," 電子情報通信学会論文誌 A Vol.J79-A, No2, pp.298-301, 1996.
- [3] 橋本 昌寛, 木下 宏揚, 酒井 善則, "Sampled Active Contour Model による輪郭抽出法," 電子情報通信学会論文誌 D-II, Vol.J77-D-II, No.11, pp.2171-2178, 1994.
- [4] 菅原 一孔, 新地 俊幹, 小西 亮介, "振動項を持つ動的輪郭モデル," 電子情報通信学会論文誌 D-II Vol.J80-D-II, No.12, pp.3232-3235, 1997.
- [5] T. Shinchu, Kazunori Sugahara and Ryosuke Konishi, "Vowel Recognition According to Lip Shapes by Using Neural Network," Proceedings of 1998 IEEE International Joint Conference on Neural Networks, pp.1772-1777 (1998). Anchorage Alaska