(d) 情報

動的輪郭モデルによる唇形状抽出と母音認識のハードウエア実現について

佐々木 悠介 †　(鳥取大学大学院工学研究知能情報工学専攻)
川村 尚生 †　(鳥取大学工学部知能情報工学科)
菅原 一孔 †　(鳥取大学工学部知能情報工学科)
†{sasaki,kawamura,sugahara}@ike.tottori-u.ac.jp

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．しかし現在の音声認識手法では，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全くなく，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．

本稿では読唇によるシステムの制御を目指し，動的輪郭モデルにより唇の形状を抽出し，その形状情報を用いた母音認識の手法を提案する．また，本手法の様々なシステムへの組み込みを目指し，FPGA 上にハードウエアとして実現した．それによる母音認識結果を示し，本手法の有効性を確認する．

2　動的輪郭モデル

動的輪郭モデル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手法である．引力は隣り合う 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．

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 の合力の抽出領域に対する垂直成分を打ち消す働きをもつ．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．本稿では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．

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が動作点の移動する画素のみで済み，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．

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 2 に示すとおり唇の外側形状を抽出することができる．なお

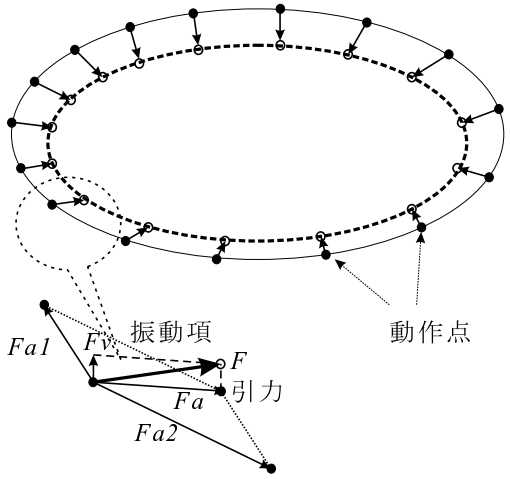

図 1: 引力と振動項による収縮動作

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．

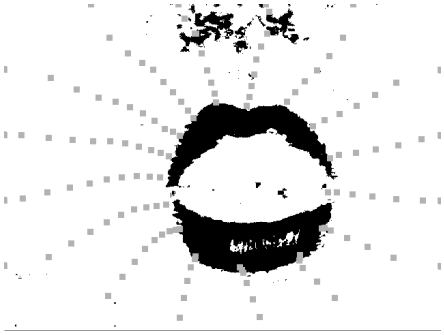

図 2: 唇の外側形状を抽出している様子

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 3 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 4 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 4 のように，

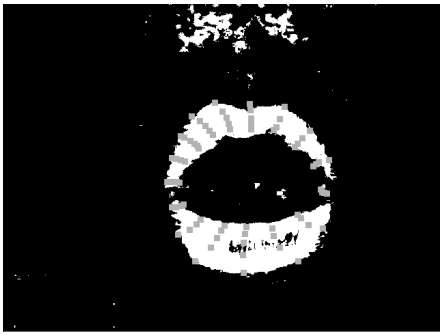

図 3: 唇の内側形状を抽出している様子

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．

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．

図 4: 唇の外側形状の抽出の失敗例

4 母音認識の手法

動的輪郭モデルの収縮により得られた動作点の座標を基に，母音を認識する手法を提案する．

本手法では，各母音ごとにテンプレートを用意し，収縮した動作点との距離を計算することによって母音を認識する．あらかじめ登録しておくテンプレートの作成手順としては，はじめに母音を発話している時の画像に対し動的輪郭モデルを適用し，収縮した動作点の座標データを取得する．続いて座標データを正規化する．最後に正規化したデータをそれぞれの母音ごとに平均し，二次曲線に近似した．テンプレートを曲線にすることで，距離計算の手間を軽減できると考えた．

5 唇形状抽出システムのハードウエア実現

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エア化を試みた．ハードウエア化は 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 5 に示す．

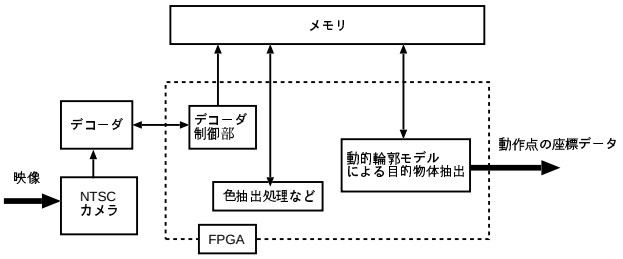

図 5: システム構成

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% のハードウエ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．

次に母音認識の実験を行った．母音のテンプレート曲線は，それぞれ 20 枚の画像の動的輪郭モデルを適用した結果を平均したものを使用して作成した．この認識システムにそれぞれの母音 20 枚づつ，計 100 枚の画像を与えた時の認識結果を表 1 に示す．

表 1: 母音認識結果

母音	誤って認識された数
あ	1
い	0
う	1
え	1
お	1

6 おわりに

本稿では，発話時の画像に動的輪郭モデルを適用し，取得した唇形状を基に母音を認識する手法を提案した．提案した手法は FPGA 上にハードウエア実現し，認識実験をした結果，その有効性を確認した．今後の課題として，本稿で示した母音認識の手法を，単語認識システム開発へと応用する事を考えている．また，カメラと映る人物の位置により唇の大きさが変化するので，大きさに依らないシステムを開発する事を考えている．