

エネルギー関数とオプティカルフローを用いた 口唇輪郭の抽出・補完と追跡†

光本浩士†† 岡崎耕三††† 梶見直樹††††
田村進一‡ 河合秀夫‡‡ 副井裕†††

口形は、人間の視覚では簡単に抽出・認識できるが、これを画像処理により行うとすると、単純なエッジ処理だけでは困難である。これは唇と皮膚の濃淡レベルの差があまりなく、かつ境界がはっきりしない部分がある上、口の周辺の髭や唇のしわ、歯、影などによるノイズが多いためである。われわれは、先験的知識として口唇自体の持つ拘束条件：(a)滑らかな輪郭、(b)左右上下の対称性、(c)画像特徴への適合度、を評価関数とした。それらをエネルギー関数で表現し、エネルギーの最小化を8近傍探索法で解くことにより、輪郭線の途切れや雑音の影響をなくする口唇輪郭抽出手法を提案する。唇の動き追跡に対してはオプティカルフローを用いる。実験においては、母音 /a/, /i/, /u/, /e/, /o/ を発声している自然な状態の口形画像から特に抽出しにくい唇外側輪郭の抽出を行った。

1. はじめに

画像処理においては多くの場合、雑音やぼけの影響を受ける。そのため境界線が連続にならず途中で切れるなど、抽出特徴点が曖昧になることが多い。それに対して人間の視覚は、それをうまく補完する機能を持っている。口唇輪郭の抽出は、その一例である。口形の理解は、また「読唇」の自動化、高次コミュニケーションの手段としてコンピュータ技術の発展とともに新しい重要な課題となってきている。

口形は、人間の視覚では簡単に抽出・認識できるが、これを画像処理により行うとすると、単純なエッジ処理だけでは抽出は困難である。これは皮膚の濃淡レベルの差があまりなく、かつ境界がはっきりしない部分がある上、口の周辺の髭や唇のしわ、歯、影などによるノイズが多いためである。

雑音環境下でも相手の唇の動きを見ることができると言葉の認識率が上がることが確かめられている¹⁾。これに着目し、Petajan²⁾ はカメラによる口唇情報を

音声認識の後処理に用いることによりその認識率を高めようと試みている。齊田ら³⁾ は唇の上下左右に発光ダイオードを装着し X-Y トラッカでその動きを検出し、単語認識の試みを行っている。松岡ら⁴⁾ は被験者の唇に黒色の口紅をつけ、唇の外側の輪郭線を抽出した。さらにパターン認識の手法を適用して母音の識別、単語の認識を行った。この方法は明瞭な唇輪郭が得られるため、安定かつ高速な処理を行うことができ、読唇の可能性を示したものである。しかし、より汎用性を持たせるためには口紅なしで口唇輪郭抽出可能な手法の開発が望まれる。内村ら⁵⁾ は、口形として、唇に口紅をつけずに、照明方法の工夫により口の内部領域を暗くし、適切なるしきい値処理によって口唇内側の輪郭の内部領域を抽出した。そして、この口形領域の計測から母音識別をする手法を実験的に示した。腎超音波像を対象とはしているが、山田ら⁶⁾ は不完全な画像からの輪郭抽出に有効なモデルを利用した DP マッチング法を提案している。この手法はモデルが比較的固定しているので、これを口唇抽出に適用した場合、発声時の口唇の大きな変動や多様な変形にどのように対処するかの問題が残る。これらの研究は騒音環境下での、あるいは聴覚障害者の音声認識の補助手段、もしくは発声メカニズム解析や顔画像の知的通信のための口唇運動解析手段として用いることができる。

Kass ら⁷⁾ は、Snake という会話型の画像輪郭抽出システムを開発している。Snake は線やエッジのような特徴に向かうイメージ力と、外部の拘束力によって導かれるエネルギーを最小化するスプライン曲線を与え

† Lip Contour Extraction, Complement, and Tracing by Using Energy Function and Optical Flow by HIROSHI MITSUMOTO (Faculty of Engineering Science, Osaka University), KOZO OKAZAKI (Faculty of Engineering, Tottori University), NAOKI KAJIMI (Chugoku Electric Power Company), SHIN-ICHI TAMURA (Osaka University Medical School), HIDEO KAWAI (Osaka Electro-Communication Junior-College) and YUTAKA FUKUI (Faculty of Engineering, Tottori University).

†† 大阪大学大学院基礎工学研究科

††† 鳥取大学工学部

†††† 中国電力(株)

‡ 大阪大学医学部

‡‡ 大阪電気通信大学短期大学部

る。すなわち、このシステムはいくつかの拘束条件をエネルギー関数の形で定義し、そのエネルギーを最小化することによって変形可能なモデルを抽出画像の輪郭にマッチングさせている。これは、疎らなデータからの“視覚再構築”⁸⁾の一種とみなすことができる。Kassらはステレオマッチングや口唇輪郭抽出とその動きの追跡についてもこのシステムの有効性を示している。彼らの口唇輪郭追跡における問題点としては、やはり口紅を使用していることである。われわれは、先験的知識として口唇自体の持つ拘束条件：(a)滑らかな輪郭、(b)左右上下の対称性、(c)画像特徴への適合度、を評価関数とする。それらをエネルギー関数で表現し、エネルギーの最小化を8近傍探索法⁹⁾で解くことにより、輪郭線の途切れや雑音の影響をなくする口唇輪郭抽出手法を提案する。唇の動き追跡に対してはオプティカルフローを用いる。

2. エネルギー関数

本手法は口唇スプラインモデルをいくつかのエネルギー関数によって拘束し、そのエネルギーが最小になるようにこのモデルを変形する。エネルギー関数を定義する際の対象に関する知識としては、

- (a) 口唇は滑らかな輪郭閉曲線を持つ。
 - (b) 左右形状に対称性を持つ。上下形状に弱い対称性を持つ。
 - (c) 口形画像から得られる特徴とスプライン曲線は適合する。
- などである。

エネルギー関数を(1)式で定義する。

$$E = E_{int} + E_{sym} + E_{image} \quad (1)$$

ここで、 E_{int} は(a)を実現したものである。スプライン補間法で節点間を補間する際の滑らかさを評価するものである。 E_{sym} は(b)を実現したもので、口唇の対称性を評価したものである。 E_{image} は(c)を実現したものでエッジ画素(以下混乱のない限り単に画素と呼ぶ)およびエッジ線素(以下単に線素と呼ぶ)に対しての適合度について評価したものである。(1)式の最小化により口唇輪郭を抽出する。以下ではこれら各項について順に説明する。

2.1 内部スプラインエネルギー

口唇の特徴の1つとして、その輪郭線は滑らかな曲線であることがわかる。この滑らかさを実現するために口唇輪郭モデルとして周期端を持つ3次スプライン補間曲線を用いた。

滑らかな拘束の内部スプラインエネルギー E_{int} はスプライン補間曲線が大きく振動するのを抑えるための項として(2)式のように定義する。

$$E_{int} = \frac{\sum (\alpha |f_s|^2 + \beta |f_{ss}|^2)}{2\sum 1} \quad (2)$$

ここに α , β は定数である。 f_s , f_{ss} はパラメータ表現された2次元平面上の曲線 f の基点からの距離 s に関するそれぞれ1次微分と2次微分を表す。 \sum は曲線上の点について取るものとする。これは高次の項を抑えるように働きがあり、スプライン関数の振動を解消する働きをしている。

2.2 対称性の拘束

口唇はグローバルに見ると図1のように左右ほぼ対称であると考えられる。この特徴を利用して左右方向の対称性エネルギー関数 E_{symh} を(3)式のように定義する。ここでは曲線 f を横軸 x の関数とする。

$$E_{symh} = \frac{\gamma \sum \{f(x) - f(-x)\}^2}{\sum 1} \quad (3)$$

また、同様に上下方向の弱対称性のエネルギー関数 E_{symv} を(3)'とする。

$$E_{symv} = \frac{\delta \sum \{f_u(x) - f_l(x)\}^2}{\sum 1} \quad (3)'$$

ここで、 f_u は上唇外側輪郭、 f_l は下唇外側輪郭であり、 γ , δ は定数 ($\gamma > \delta$) である。 \sum は x 座標について取るものとする。このエネルギー関数は左右方向、上下方向の対称性を保つ働きがある。

自然口形画像をエッジ抽出処理すると、口の周りの髭や歯、唇のしわ、影などによって不要なエッジが出てきたり、上唇外側輪郭のように本質的にエッジの出にくい部分が存在することがよくある。そのようなときにこの拘束条件は、一部の強力なエッジによって、本来存在すべき位置に輪郭線を置く動きをする。

2.3 画素の誤差エネルギー

口唇スプラインモデルを画像中の口唇輪郭に引き付ける最も強い力は画像中の特徴そのものである。本論

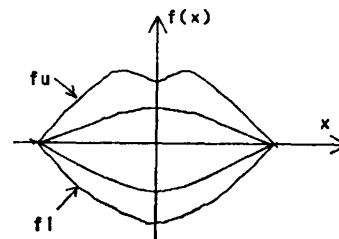


図1 口唇の対称性
Fig. 1 Symmetry of lip contour.

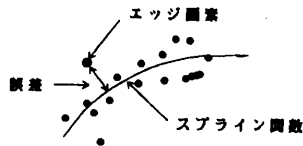


図2 スプライン曲線が画素に引き付けられ適合する様子

Fig. 2 Spline curve being pulled and fitted to pixels.

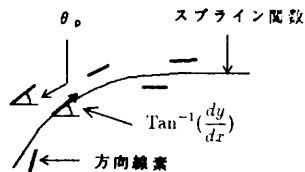


図3 スプライン関数と方向線素の関係

Fig. 3 Relation between spline function and directional line segments.

文では、画像の特徴としてエッジを用いている。しかし、口唇はその画像強度 $I(x, y)$ や輪郭が不安定、不明確なため、輪郭の抽出が困難なことがしばしばある。口形原画像をエッジ画像にしたときにエッジ画素が孤立的で輪郭を形成するまでに至らない場合、それらの不十分な特徴から輪郭成分を取り出すため、画像特徴からの支持を表すエネルギー関数 E_{image} を次のように分けて定義する。

画素とスプライン関数 (図2) との誤差エネルギー (4)式、および方向線素特徴からの支持を表す (図3) ための方向線素とスプライン関数との誤差エネルギーを (5)式のように定義する。これらの和が画素特徴への適合エネルギーとなる。

$$E_{\text{image}1} = \frac{\eta}{\{(fx_i - x_i)^2 + (fy_i - y_i)^2\}^{1/2}} / \sum 1 \quad (4)$$

$$E_{\text{image}2} = \frac{\xi \sum \left\{ \tan^{-1} \left(\frac{dy}{dx} \right) - \theta_p \right\}^2}{\sum 1} \quad (5)$$

ここで、 η 、 ξ は重み係数、 (x_i, y_i) は画素 i の座標値、 (fx_i, fy_i) はそれに最も近いスプライン曲線上の点、 θ_p は線素の方向を示す。 \sum は曲線近傍の画素・線素について取る。なお、(4)式の分子が大きな値になるときは一定値で置き換えるものとする。(4)式によって口唇スプラインモデルは図2のように周囲の上下左右10数ドットを探し、画素に引き付けられるように動く。これは画素が口唇スプラインモデルの近傍により多く存在するような場所を輪郭線とみなす拘束である。(4)式

は画素との距離の負号でエネルギーが定義されている。曲線はその近傍のある幅を持った帯状領域内の画素から支持を受ける。その際、距離の逆数を用いることにより、曲線から離れた点が近傍帯状領域にはいるか否かによる影響が小さくなるようになっている。また、上限値に設けることにより、滑らかさを優先させた画素ずれには寛容な曲線が得られる。しかしながら、(4)式の拘束だけであると、歯、人中、縦しわなどの縦線に対しても曲線が引き付けられてしまう。そこで、できるだけ曲線に平行な方向線素が多く存在するところに曲線が引き付けられるよう(5)式の拘束を導入した。一方、(5)式だけであるとは非負であるから、画素ないしは線素のない場所を曲線が選ぶ可能性がある。(4)式は逆にそれを防いでいる。

3. 口唇輪郭の抽出

3.1 口唇内側輪郭抽出の前実験

音声認識や合成に関しては口唇内側輪郭が比較的重要である。しかしながら、これらは判別分析法によるしきい値処理および簡単な領域処理等で抽出されることが多い^{5), 10)}。図4には「天満 (Temma)」と発声したときの 1/30 秒ごとの動画像第4フレームおよび第5フレームを零交差法で抽出した例を原画面上に重ねて示す。零交差法では口唇内側輪郭および歯が強く抽出されている。細かくみると、本実験のような一般的な天井照明下では下唇内側輪郭は正確に抽出されるが、上唇内側輪郭は影の影響で実際よりやや上側に抽出される。零交差輪郭抽出のしきい値をもっと下げると口唇抽出には不要な妨害成分が多く抽出され、下唇外側輪郭はほとんど抽出されない。本論文の目的は必ずしも口形からの音声認識に限定しているわけではなく、「知識」を埋め込んだエネルギー関数の輪郭抽出に対する有効性を調べることにある。そこで、以下では抽出が困難な自然環境下における口唇外側輪郭を主として抽出の対象とすることにする。



(a) 第4フレーム (b) 第7フレーム
(a) Fourth frame. (b) Seventh frame.

図4 「天満」発声時の零交差法による輪郭抽出
Fig. 4 Contour extraction by zero-crossing method when uttering "Temma".

3.2 処理の概要

図5に処理の概要を示す。まず、入力口唇画像(256階調)から画像処理装置(AVIO EXCEL TVIP-4100)を使ってエッジ画像を抽出する。このエッジ画像の周りに原型となる口唇スプラインモデルの節点を置く。節点数はスタート時は4とする。徐々に節点数を増やして精細なフィッティングが行えるようにする。エネルギー最小値は、8近傍探索法により求めた。すなわち、節点を結ぶスプライン補間曲線を求めたのち、各節点においてその近傍8方向に節点を変化してみたときの、節点の両側のスプライン区間での内部スプラインエネルギー、対称性のエネルギー、画素の誤差エネルギー、線素方向の誤差エネルギーを計算して、その和を全エネルギーとする。この全エネルギーが一番小さくなる方向に節点を移動していく。この処理を繰り返し、エネルギーがそれ以上どの方向に動かしても変化しない場合その節点の両側の区間をそれぞれ等分するように節点数を2個増やして、もう一度繰り返し処理を行う。節点の数が8になったところで高解像度のエッジ画像に切り換えて処理し、最後に節点数が10

になったのち、全エネルギーが最小値(極小値)になったところで処理を終る。

3.3 初期値の設定

図6(a)は180×125画素、256階調で、/a/と発音している状態を入力した画像である。これは髭がなく比較的肌のきれいな人を写したものである。最小化すべき全エネルギー関数Eを(6)式とする。

$$E = E_{int} + E_{sym} + E_{image1} + E_{image2} \quad (6)$$

(1) 係数の初期値

実験の際、最も問題になるのは各エネルギー関数間の重み係数の設定である。これらは数回の実験から経験的に求めた。今回用いた係数は $[\alpha=0.0005, \beta=0.0005, \gamma=5, \eta=600$ または 800 (η は逆数のため大きな数値になる), $\xi=4]$ である。各エネルギーの評価の割合は、画像適合のエネルギーが最も大きく、対称性、内部エネルギーの順に設定した。対称性のエネルギーはあらかじめゼロになるように初期口唇スプラインモデルを設定しておき、そのあと対称性を保つ働きをさせている。

(2) 節点の初期値

あらかじめ口唇の近くに節点(実験では4個とした)を設定し、ここでおおまかなフィッティングを行うことからスタートする。節点は全エネルギーEが小

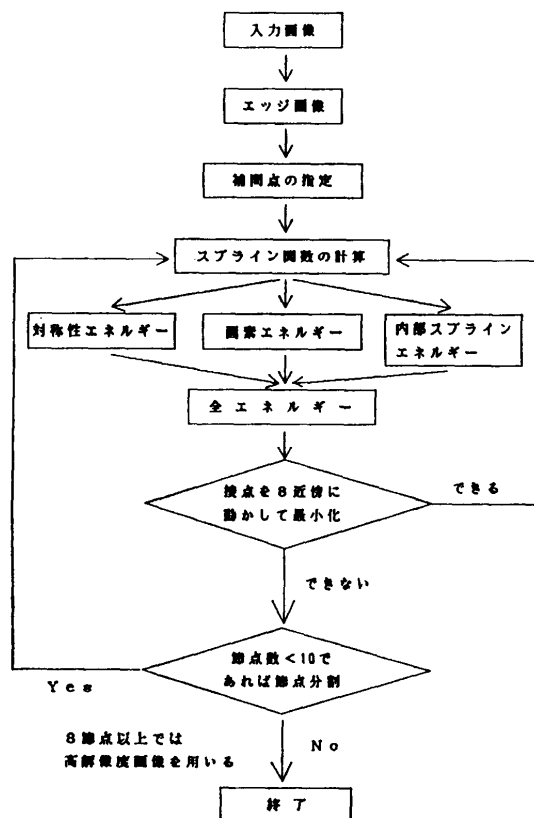


図5 処理の概要

Fig. 5 Outline of processing model.



発音 /a/

(a) 入力画像
(a) Input image.



(b) 低解像度画像
(b) Low-resolution.



(c) 高解像度画像
(c) High-resolution.



(d) (b)のエッジ画像
(d) Edge image of (b).



(e) (c)のエッジ画像
(e) Edge image of (c).

図6 処理に使用した低解像度画像と高解像度画像およびそれらのエッジ画像

Fig. 6 Low-resolution image, high-resolution image, and their edge images used for processing.

さくなる方向へ移動し、最小になる点で止める。口唇輪郭を抽出するために、まず図6(b)の低解像度の画像を用いて処理を行う。

3.4 2重解像度画像の利用

図6、図7にそれぞれ2種類の解像度の画像を用いた処理の概要を示す。図7において口全体のおおまかな形を見るために 15×15 の大きなマスクのフィルタで移動平均を取ったものが低解像度の画像(図6(b))である。また、口の周りの髭や皮膚の凸凹などは、エッジ抽出には非常に悪い影響を与えるため、それを消す程度の 5×5 の小さなマスクで移動平均を取ったものが高解像度の画像(図6(c))である。そして、各々の画像をロバーツのオペレータを使ってエッジ抽出し、細線化した(図6(d),(e))。図6においては照明光が左上方向より当たっているため、上唇外側輪郭は左側のみ抽出されており、右側は出ていない。また下唇外側輪郭は右側のみ抽出されており、左側は出ていない。右側下唇外側輪郭部も下唇の影の影響で上向きにへこんだ形になっている。

2つの異なる解像度の画像は、処理速度を上げるために口唇スプラインモデルの口形エッジ画像に対するフィッティングの割合によって使い分ける。図7に示すように画像は低解像度画像と高解像度画像で別々に処理する。まず低解像度画像に対して口唇スプラインモデルのフィッティングを行い、ある程度のフィッティングができた状態でより精細な口唇輪郭を得るために高解像度画像に対してのフィッティングを行う。この

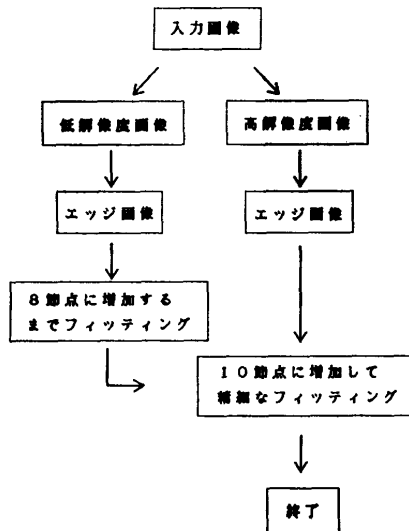


図7 2種類の解像度を利用した処理の概要
Fig. 7 Outline of processing using two kinds of resolution images.

処理によって、全体的な形状をつかみやすくし、なおかつ精細なフィッティングをさせることを狙っている。

3.5 節点の逐次分割

図8は/a/の口唇輪郭を抽出する処理過程で、節点が逐次分割されていく様子を示している。図8(a)は最初の4つの節点でフィッティングをスタートさせたときの状態である。画素および線素の誤差エネルギーの探索範囲は上下左右14ドットとしている。このとき4つの接点は、随時口唇輪郭線にフィッティングしていく。そして、上下中央の2つの接点はエネルギーが最小となりフィッティングが完了した時点で左右均等な位置に接点を増加させ、より細かいフィッティングを行えるようにした。これを示したのが図8(b),(c)である。図8(b)において上中央の節点の両側区間における全エネルギー量は最小になり、そのため移動が止まっている。そこで、上中央節点の両側にそれぞれの区間を等分するような節点を計2個置く。これにより上部輪郭が4つに分割される。次に同様に、下中央節点のエネルギーが最小になって、移動が止まった時点で下部輪郭を各2分割し、同様に収束させる。次にエッジ画像を高解像度のものに変えて、より精密

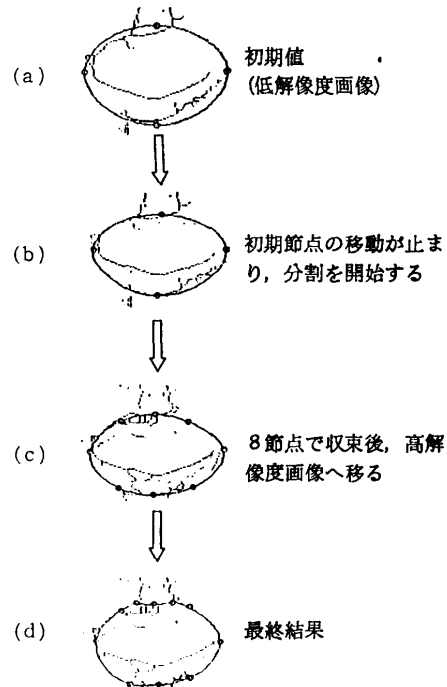


図8 逐次節点分割による口唇輪郭のフィッティング
○印は節点
Fig. 8 Fitting of lip contour by successive node division.
Mark ○ is node point.

なフィッティングをはかる (図 8 (c)). このとき誤差エネルギーの探索範囲は, ある程度のフィッティングができていたため, 必ずしも広範囲でなくてよいので, 上下左右 7 ドットと狭くした. 上唇上部のへこんだ部分 (人中) へのフィッティングを行うために, 最後の処理として上部中央付近をもう一度各 2 分割した. すべての区間でエネルギーが最小になり, どの点も動かなくなったところで終了とした. こうして抽出した口唇輪郭の結果を図 8 (d) に示す.

3.6 エネルギー関数の振舞

図 8 での全エネルギー量の変化を図示したものが図 9 である. 縦軸がエネルギー量, 横軸が繰り返し回数を示している. この図で (a), (b), (c), (d) は図 8 のそれとほぼ対応している. まず (a) の時点が初期値であり, それからエネルギーの減少がみられる. (b) の時点では節点が増えるため一時的にエネルギー量が増える. しかしその後エネルギー量は徐々に減っ

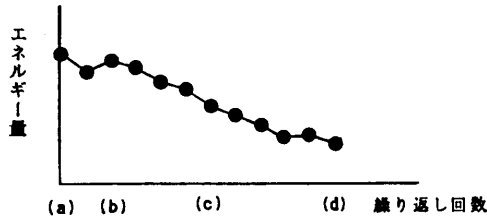
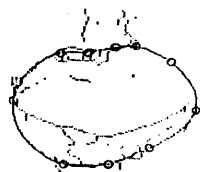
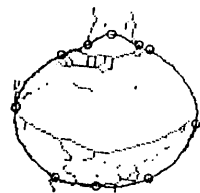


図 9 全エネルギー量の変化 (a)~(d) は図 8 に対応

Fig. 9 Change of all energy. (a)~(d) correspond to Fig. 8.



(a) 対称性のエネルギーを使用しないときの例 (a) Example of case of symmetry energy unused.



(b) 線素エネルギーを使用しないときの例 (b) Example of case of line-element energy unused.

図 10 各エネルギーを使用しないときの輪郭抽出結果 Fig. 10 Results of contour extraction for each energy factor unused.

ていき収束する. 処理の最初の部分ではすべての点が移動するため大きなエネルギーの減少がみられる. その後, 節点が増えて, 小さな区間の変動になるためエネルギーの変動は小さくなり収束する.

また対称性のエネルギーを使用しない例と線素エネルギーを使用しない例をそれぞれ図 10 (a), (b) に示す. 図 10 (a) については画素にフィッティングはしているが対称性がなくなっている. 図 10 (b) は, 上方の鼻の下 (人中) の縦線エッジに引き付けられて形が崩れている.

3.7 その他の実験例

(1) 他の母音の例

図 11 に他の母音 /i/, /u/, /e/, /o/ の場合についての実験結果を示す. いずれも下唇下部のエッジはロバーツオペレータによる処理では一部しか出ていない. 本処理ではうまく補完されている. ただし, 今のところエネルギー関数の中に組み込んだ知識のうち, 滑らか拘束はどこでも均一に評価するため, 口の両側口角部分の尖りにうまく適合できず内側をたどる鈍った形状となってしまう. 左右両端部は滑らか拘束の重みを小さくするか, または零にする, 尖りを評価する, 口角は暗いことなどを利用して口角部を求めた後, 上下輪郭を別々に求める, 等の改善が必要である.

(2) 初期値の設定を変えた例

初期値をずらして行った実験を図 12 に示す. この

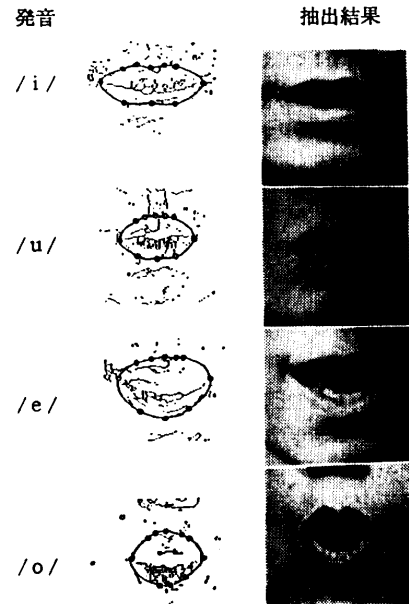


図 11 他の母音の口唇輪郭抽出結果 Fig. 11 Results of lip contour extraction for other vowel.

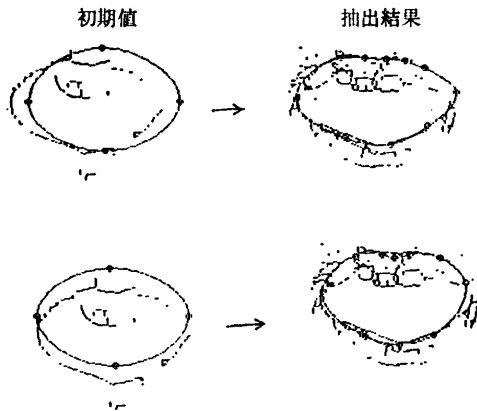


図 12 初期値を変更したときの実験結果
Fig. 12 Experimental results for initial values changed.

実験では、上側は上唇上部輪郭を求めており、下側は下唇上部（口唇内部）輪郭を求めている。図 12 のように節点の初期位置は異なっても、ほぼ正確な輪郭が求まることがわかった。しかし、これは解輪郭線が初期節点からの探索範囲内にあるときに可能であって、その設定値が大きく逸脱している場合には別の輪郭に収束する可能性がある。

4. 口唇輪郭の追跡

オプティカルフローは口唇輪郭の動き追跡の有力な補助手段として考えられる。時系列画像からオプティカルフローを求める方法としては、大きく分けて2つの方法がある。1つは、画像中の特徴点を見出し、画像フレーム間で対応点を探し、速度ベクトルを決定するマッチング法^{11),12)}である。もう1つは、動画中の運動物体の輝度の時間変化と空間変化に一定の関係があることを利用する Horn らの、いわゆるグラディエント法¹³⁾⁻¹⁵⁾である。本論文ではこの後者の方法を用いる。

本章で提案する口唇輪郭の追跡処理の流れ図を図 13 に示す。まず、連続する2枚の画像は各々画像縮小されて、その画像間のオプティカルフローが計算される。次に、いままで節点移動に際して近傍8方向についてエネルギーを計算していたが、オプティカルフローを参照することにより、移動する方向を限定できる。図 14 に示すように、先の処理方法では、図 14 (a) に示した“●”節点の全近傍の画素をサーチしたが、この節点で図 14 (b) のようなオプティカルフローが得られると、口唇の動きがある程度限定できるため、図 14 (c) のようにオプティカルフローが示す

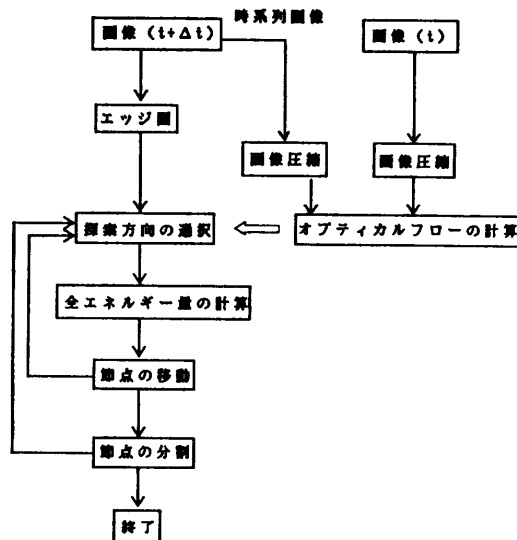
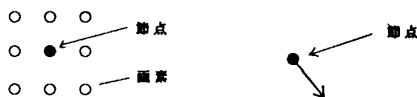
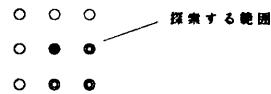


図 13 処理の流れ図
Fig. 13 Flow chart of processing



(a) 節点 (●) と全近傍画素 (○) (b) 節点上のオプティカルフロー



(c) 近傍画素の探索領域 (◎)

図 14 探索範囲の削減

- (a) 節点 (●) とすべての近傍画素 (○)
- (b) 節点でのオプティカルフロー
- (c) 近傍画素での探索領域 (◎)

Fig. 14 Reduction of search space.

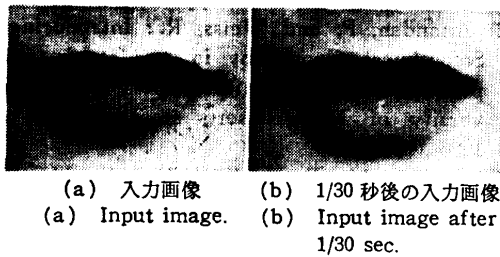
- (a) Node (●) and all neighbouring pixels (○).
- (b) Optical flow on node.
- (c) Searching space (◎) of neighbouring pixels.

“◎”の3近傍および現在位置のみに限ってエネルギー量を計算すればよい。この方法を用いると、パターンサーチの探索範囲を4/9に減少できる。ただし、オプティカルフローを計算する処理が加わるため全体としての計算量はあまり変わらなくなってしまう。オプティカルフローを用いることの利点は計算量よりもむしろ、口が大きく動いたとき、誤った追跡が起りうるので節点の移動方向の指示を与えそれを防止することである。

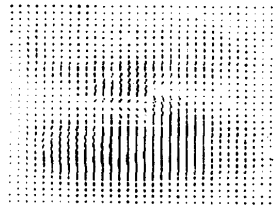
実験した画像は口を閉じた状態から /a/ と発音している時系列画像 (120×90 pixel, 各 1/30 sec ごと) 15

枚である。これらの画像は上から光を当てた状態であり、これをエッジ処理した場合、上唇のエッジは比較的簡単に抽出できるが、下唇下側のエッジは、少ししか出ていない上、紛らわしい影があったりする。また口を広げたときに歯等の不要なエッジが出たりする。/a/ は、口唇の動きが最もはっきり表れる口形の1つである。追跡処理に先立って縮小画像上でオプティカルフローを求めるため、それらを各 32×32 pixel (実データ領域は 30×23) に縮小した。各々の画像をフィールドに分離、圧縮し、画像間でのオプティカルフローの計算を行う。

時系列画像の最初の画像の口唇輪郭を得るために、まずその初期値として4節点を持つ適合口唇スプラインモデルを人が与えた。2枚目以降は1枚前で自動的に求めた口唇輪郭を初期値にした。図15は、5枚目(図15(a))と6枚目(図15(b))の画像を示す。図15(a)と図15(b)の間のオプティカルフローは、図15(c)のように計算される¹⁴⁾。図15(c)の画像中の点は画素を表しており、各画素での速度ベクトルを線分の大きさ、方向で表している。図16は口唇スプラインモデルが口唇輪郭を追跡する様子を示している。これは/a/と発声したときの1/30秒ごとの時系列画像において、閉じた状態から唇が開く瞬間の連続した4コマの追跡状態を示している。唇の開く速度は、かなり早いことがわかる。下唇外側の輪郭は非常に出ていく、また影による紛らわしいエッジが出ている。こ



(a) 入力画像 (b) 1/30秒後の入力画像
(a) Input image. (b) Input image after 1/30 sec.



(c) (a)(b)間のオプティカルフロー
(c) Optical flow between (a) and (b).

図15 /a/発音時のオプティカルフロー
(最も動きの顕著な例)

Fig. 15 Optical flow when pronouncing /a/
(example of major movement).

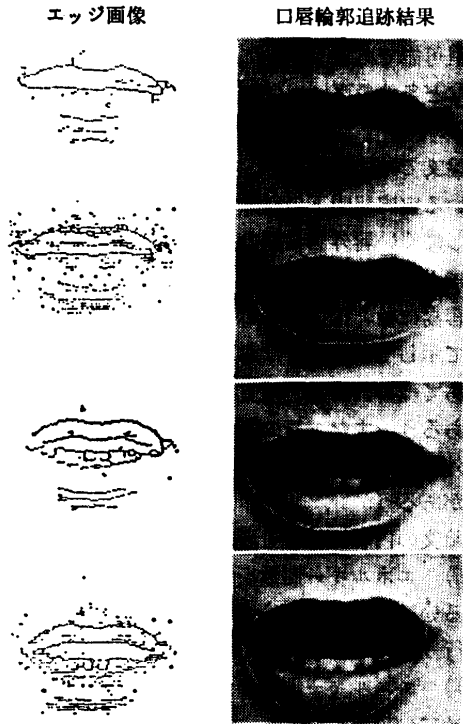


図16 開唇瞬間の口唇輪郭追跡
Fig. 16 Lip contour tracking at the moment of lip opening.

れらに対して本手法はうまく唇外側輪郭を抽出していることがわかる。

オプティカルフローを使わない方法では NEC-PC 9801 の C 言語で2枚の画像間の追跡に1分強かかっていた処理時間がオプティカルフローを用いることによりそれが追跡時間に限って言えば30秒となり、約2倍のスピードアップになっている。ただし、オプティカルフロー計算には現在25秒かかっているのも、それがICチップ化¹⁶⁾されるなどの要素がなければ処理の高速化とはならない。

5. むすび

口形輪郭は不十分な輪郭情報しか画像内に存在しないため抽出しにくいことが多い。このようなとき、先験的知識をエネルギー関数に埋め込むことにより、輪郭復元を行うプロセスについて検討した。比較的髭の少ない人について口唇輪郭抽出を行ったところ、母音を識別するには十分であろうと思われる抽出が行われた。さらに時系列画像についての動きの計算(オプティカルフロー)を利用することにより、口唇の動きをより確実に追跡する手法を提案した。口形の抽出は、高雑音下などにおける音声認識装置の精度向上、聴覚

障害者のコミュニケーション補助, 口唇運動解析などに利用可能である。さらに, エネルギー関数法自体は角膜など条件の悪い画像からの輪郭抽出¹⁷⁾に適用可能である。

本論文では, 母音 /a/, /i/, /u/, /e/, /o/ を発声している室内照明下の自然な状態の口形画像から特に抽出しにくい唇外側輪郭の抽出を行った。唇内側輪郭は歯による妨害要因はあるものの, 照明等の工夫で比較的容易に輪郭抽出することができる¹⁰⁾ので, 一部を除いて示していない。なお, 唇を対象とした場合には面のあてはめによる“視覚再構築”も検討してみる必要がある。また, 唇の厚さを考慮して, 口唇内側輪郭と外側輪郭を連動させた形で抽出することも検討の価値がある。

本論文の内容・特徴を要約すると次のようになる。

(1) エネルギー関数を定義することにより, 不連続あるいは疎らなど不完全な特徴から口唇輪郭抽出復元を行った。

(2) エネルギー関数の最小値を8近傍探索法により求めている。

(3) 初期値の設定にある範囲の任意性を与えても, 解が一意的に求まる。

(4) 抽出精度向上のため, 2重解像度画像を用いている。

(5) オプティカルフローの適用により, 画像中の比較的速い動きに対しても確実な追跡を可能とした。

残された課題は

(1) 口唇の左右両端口角部のフィッティング。

(2) 唇内側輪郭および音響の情報との組合せによる高精度音声認識。

(3) 面のあてはめによる唇領域抽出の検討。などである。

近年, ニューラルネットの研究が盛んに行われている。本論文の手法は, エネルギー関数を用いており, ニューラルネット上で処理を行うことが可能である。またオプティカルフローについて, Kochら¹⁶⁾やPoggio¹⁸⁾は抵抗ネットワークを用い計算を行っている。近い将来, これらのハードウェア化ができれば実行時間に関しても満足できるものとなる。

なお, 本研究の一部は文部省科研重点領域研究(63608505, 63633516)による。

参 考 文 献

1) 西田: メディアテクノロジーの現状と将来, システムと制御, Vol. 30, No. 7, pp. 393-400

(1986).

- 2) Petajan, E.D.: Automatic Lipreading to Enhance Speech Recognition, *Proc. of IEEE Global Telecommunication Conf.*, pp. 265-272 (1984).
- 3) 齊田, 清水, 山下, 荒井: 口唇情報を利用した発声発話訓練器, 第19回感覚代行シンポジウム, pp. 55-58 (1983).
- 4) 松岡, 古谷, 黒須: 画像処理による読唇の試み, 計測自動制御学会論文集, Vol. 22, No. 2, pp. 191-198 (1986).
- 5) 内村, 道田, 都甲, 相田: 画像解析による日本語母音の識別, 信学論, Vol. J 71-D, No. 12, pp. 2700-2702 (1988).
- 6) 山田, 山本, 松浦: 2次元 DP 整合法による超音波腎臓像の認識, 信学論, Vol. J 68-D, No. 9, pp. 1649-1656 (1985).
- 7) Kass, A., Witkin, A. and Terzopoulos, D.: Snakes: Active Contour Models, *Int. J. Comput. Vision*, Vol. 1, No. 3, pp. 321-331 (1988).
- 8) Blake, A. and Zisserman, A.: *Visual Reconstruction*, MIT Press, Cambridge, Mass. (1987).
- 9) Okazaki, K. and Tamura, S.: Spherical Shading Correction of Eye Fundus Image, *Proc. of IEEE International Conf. on Systems, Man and Cybernetics*, pp. 1084-1087 (1983).
- 10) 田村, 小川, 黒須: 口話を併用する音声認識, 文部省科研重点領域研究「音声言語によるマン・マシン・インターフェースの高度化」昭和62年度研究報告書, pp. 93-98 (1988).
- 11) Prager, J.M. and Arbib, M.A.: Computing the Optical Flow: The Match Algorithm and Prediction, *Comput. Vision Graphics Image Process.*, Vol. 24, No. 3, pp. 271-304 (1983).
- 12) Anandan, P. and Weiss, R.: Introducing a Smoothness Constraint in a Matching Approach, COINS Technical Report, 85-38 (University of Massachusetts at Amherst), pp. 1-26 (1985).
- 13) Yachida, M.: Determining Velocity Maps by Spatio-temporal Neighborhoods from Image Sequences, *Comput. Vision Graphics Image Process.*, Vol. 21, No. 2, pp. 262-279 (1983).
- 14) Horn, B.K.B. and Schunck, B.G.: Determining Optical Flow, *Artif. Intell.*, Vol. 17, pp. 185-203 (1981).
- 15) Nagel, H.H. and Enkelmann, W.: An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences, *IEEE Trans. Pattern Anal. Machin. Intell.*, Vol. PAMI-8, No. 5, pp. 565-593 (1986).
- 16) Hutchinson, J., Koch, C., Luo, J. and Mead, C.: 画像中の動きを抵抗ネットワークで検出する, 日経エレクトロニクス, No. 453, pp. 171-

185 (1988).

- 17) 光本, 田村, 河合, 可児: エネルギー関数によるヒトの角膜内皮細胞の形状抽出・補完, 第28回日本 ME 学会大会, 3-A-5, p. 353 (1989).
- 18) Poggio, T.: Early Vision: From Computational Structure to Algorithms and Parallel Hardware, *Comput. Vision Graphics Image Process.*, Vol. 31, No. 2, pp. 139-155 (1985).

(平成元年4月16日受付)

(平成2年1月16日採録)



光本 浩士 (正会員)

昭和 61 年鳥取大学工学部電気卒業。昭和 63 年同大学院修士課程修了。現在大阪大学大学院博士課程 1 年在学中。コンピュータビジョンとニューラルネットの研究に従事。電子情報通信学会, 日本 ME 学会など各会員。



岡崎 耕三

昭和 44 年広島大学大学院修士課程修了。同年宮崎大学工学部助手, 昭和 45 年鳥取大学工学部助手。以来, 擬似ランダム系列の応用, 画像情報処理などの研究に従事。1983 年 Pattern Recognition Society 論文賞, 工学博士。計測自動制御学会, 日本音響学会, 電気学会, 電子情報通信学会など各会員。



梶見 直樹

昭和 62 年鳥取大学工学部電気卒業。平成 1 年同大学院修士課程修了。同年, (株)中国電力に入社。在学中, コンピュータビジョンの研究に従事。電子情報通信学会会員。



田村 進一 (正会員)

昭和 41 年大阪大学基礎工学部電気卒業。昭和 46 年同大学院博士課程修了。同年同大学基礎工学部助手, 現在同大医学部教授。工学博士。この間, 通信方式, パターン認識, 画像処理, 人工知能などの研究に従事。昭和 58 年 10 月 10 日刊工業新聞十大新製品賞, 1983 年および 1988 年 Pattern Recognition Society 論文賞。昭和 63 年人工知能学会大会優秀論文賞受賞。著書「データ伝送と計算機ネットワーク」など。IEEE, 電子情報通信学会, 日本認知科学会, 人工知能学会, 日本 ME 学会, 医学放射線学会, Cognitive Science Society, Pattern Recognition Society など各会員。



河合 秀夫 (正会員)

昭和 46 年立命館大学理工学部卒業。同年大阪大学基礎工学部教務職員。現在大阪電気通信大学短期大学部助教授。パターン認識と学習, 画像処理の研究に従事。最近は画像理解, 人工知能の研究に興味を持っている。工学博士。日本 ME 学会, 電子情報通信学会, 計測自動制御学会など各会員。



副井 裕

昭和 39 年大阪府立大学工学部電気卒業。昭和 41 年同大学院修士課程修了。同年鳥取大学工学部助手, 以後同大講師, 助教授を経て, 現在, 同大教授。工学博士。回路素子および機能のシミュレーション, 能動フィルタ, デジタルフィルタ, カレントコンベアの研究に従事。IEEE, 日本音響学会, 電気学会, 電子情報通信学会など各会員。