

領域構造の解析に基づく動画像からの セマンティックオブジェクトの抽出

目 黒 光 彦 電気通信大学大学院電気通信学研究科電子工学専攻助手

あらまし

昨今、デジタル画像データの取得、伝送、加工等の画像処理全般に関するニーズが、産業界のみならず一般家庭においても高まっている。それを強く支えている技術の一部として、1) CCD等の画像を取得するためのセンサ技術、2) インターネット、移動体通信、デジタル放送等で重要となる伝送技術、3) 画像データの圧縮や強調、復元等の画像処理アルゴリズム技術、等がある。

画像処理技術の中で画像データの圧縮技術である画像符号化は、その国際標準として、静止画像符号化規格である JPEG、動画画像符号化規格である MPEG-1,2,4 が制定されている。MPEG で規定されている画像符号化手法により、莫大な動画像データを劇的に削減することが可能となっている。

画像符号化等で見られるように実用に供せられている多くの画像処理技術は、画像を単に2次元ないし3次元の数値データの集合として取り扱う処理である。これら、数値レベルで取り扱われる画像処理技術は成熟した感がある。しかしながら、これから解決が望まれる研究として、画像データを数値として機械的に処理するのではなく、画像から得られる知的な情報を元にした情報処理である。つまり、画像、動画像データから意味のある情報を抽出し判断を下す“画像認識・理解”に関する研究に重点を置く必要がある。画像認識・理解の研究を実用レベルまで推し進めるためには、まず、画像を意味のある領域に分割し、それぞれの領域間の関係を表す画像構造を把握することが不可欠である。

本研究では、画像信号に含まれるエッジや、色、領域の動きを求め、それらの情報を統合することで画像構造を記述し、最終的に、動画像を意味的にまとまりをもった対象（セマンティックオブジェクトと名付ける）と背景とに正確に分離を行うことを目的とする。画像構造の把握のために、階層画像に対する Watershed に基づく領域分割、および、均等色空間における領域統合を進めることによって、過分割を抑えた領域分割結果が得られることを示す。さらに、フレームごとに求めた領域分割結果を、隣接フレームにおいて時空間に連続した同一領域として統合させ、それをチューブとして生成する。各チューブにおける動きの有無を調べることで、動きのあるセマンティックオブジェクトの抽出が可能となる。加えて、ユーザによる大まかなインタラクティブ操作によって、静止画からオブジェクトを抽出する方法を検討する。これは、ユーザによるインタラクティブ指示によって得られるオブジェクトの輪郭線と、Watershed による領域分割結果とを重ね合わせ、あらかじめ設定している判定規則を用いてオブジェクトを抽出する手法である。これらの手法により静止画像や動画像からセマンティックオブジェクトの抽出が実現され、コンピュータビジョン等の知的画像処理、画像圧縮規格 MPEG-4 や画像検索規格 MPEG-7 の実用化を高める意味においても重要な成果が得られたことを報告する。

1. はじめに

昨今のインターネットの爆発的な普及によって、現代人は、文字、音声、音楽、写真、映像等の種々の情報を簡単に得られる環境にある。しかしながら、インターネット上に存在する莫大な情報の中から、ユーザが欲するコンテンツを簡単に検索できるシステムは、「Yahoo!®」や「Google™」に代表される文字情報に関する検索データベースのみであるといえる。なぜなら、文字情報そのものが意味を表す情報としてすでに記号化されている構造化された情報であるため、ユーザの要求にかなう検索結果が得られやすい。代わって、写真や映像などの情報は、データそのものから意味的な情報を抽出しながら、ユーザの要求に沿ったコンテンツを検索しなければならず、情報の抽出、情報の例示を含めて、ユーザの要求を満たすほどの成果をあげているとは言えない。写真や映像の数値データは、それ自体意味を持たず、ユーザら人間が解釈して初めて、コンテンツの情報が理解される種

類のデータである。この場合、どの情報をどのように抽出・整理したうえで映像メディアを記述し、その後の処理に結び付けばいいかを考慮に入れる必要がある。ユーザが映像メディアを操作するとき、ユーザの介在を最小限に留めた手続きでなされることが好ましい。つまり、ヒューマンインターフェースの観点からユーザにとって使いやすい、映像メディア処理システムの構築が急務である。

そのためには、画像を意味のある領域に分割し、それぞれの領域間の関係を表す画像構造を把握する「構造化」が不可欠である。あらかじめ「構造化」を行っておくことによって、映像コンテンツをオブジェクトに基づく情報として整理した形になり、コンテンツ情報の自動抽出や意味内容の自動認識への一歩として重要な処理であるといえる。映像データ構造化の一つの方法として、画像処理において盛んに研究されてきた領域分割を用いる方法が考えられる。静止画像に対する領域分割によって、色やテクスチャ等が均一と思われる領域ごとに分割することが可能である。動画像に対しては、それぞれのフレームごとに領域分割を行い、隣接フレームにおける領域分割結果と似通った領域を統合させる。これにより、時空間において色やテクスチャの性質が均一な領域に分割することが可能となる。この時空間領域をチューブと呼ぶ^{[2],[5],[6]}。動画像からのチューブの生成は、まさしく、動画像に対する時空間的な「構造化」であるといえる。特定の動きのある領域とその周囲にある隣接領域に属するチューブを追跡することにより、動いた軌跡を時空間的に表現したチューブを作成することができる。これにより、意味的にひとかたまりと考えられるセマンティックなオブジェクト領域の抽出が可能となる。

セマンティックオブジェクトの抽出を成功させるための領域分割は、分割された領域の境界線とエッジとが一致していることが必要条件となる。それを満たすものが、画像の輝度勾配に基づいた Watershed による領域分割である^[1]。これは、輝度勾配によって求められる尾根をたどることによってできる閉領域を一つの領域とした領域分割手法である。しかしながら、実際の画像に対して Watershed に基づく領域分割を行った場合、雑音や陰影等の影響による不必要なエッジ成分によって過分割が生じやすい。そこで本研究では、平滑化処理、および、解像度変換を施した階層画像に対する Watershed 処理による領域分割手法の検討を行った。平滑化処理では、Watershed 処理の前提となる輝度勾配画像を作成する前に、あらかじめ、画像から不要な雑音成分を取り除く。これにより、ある程度の雑音の影響を抑えた輝度勾配画像が得られる。平滑化と平行して、解像度変換を施すことで階層画像を作成し、それぞれの階層画像に対してしきい値により選択を行って生成した輝度勾配画像に対して Watershed 処理を行う方法を検討した。低解像度の画像では、過分割を抑えた領域分割結果が得られている。高解像度の画像では過分割の傾向となるが、オブジェクトの境界線を正確に得ている。よって、複数の解像度を有する階層画像に対する Watershed に基づく領域分割結果を統合することによって、過分割を抑え、かつ、正確な境界線を有する領域分割が実現される。さらに、フレームごとに求めた領域分割結果を、隣接フレームにおいて時空間的に連続した同一領域として統合させ、それをチューブとして生成する。生成されたチューブの動きの有無や隣接するチューブの位置関係を利用して、同一オブジェクトと思われるチューブの組み合わせを選択することにより、セマンティックオブジェクトの抽出を行う。加えて、複雑な背景を有する静止画像から、セマンティックオブジェクトを抽出する手法の検討も行った。これは、マウスやタブレットなどのポインティングデバイスを用いたユーザによる大まかな操作によって、特定のオブジェクトを抽出するものである。処理例として、複雑な背景を有した上半身の人物画像に対して、大まかなユーザのインタラクティブ指示によって顔の領域を楕円で囲むことで人物の輪郭を推定し、想定している状況に沿った判別規則に基づいて人物の抽出が効果的に実現される。

2. 領域分割の基本的技術

2.1 Watershed に基づく領域分割

オブジェクト抽出を念頭においた領域分割の場合、分割結果の輪郭線がエッジと一致することが必要である。そこで、領域の輪郭線がエッジと一致する Watershed に基づく領域分割を用いる^[1]。Watershed による領域分割は、差分オペレータによって処理対象の輝度画像から輝度勾配画像を求め、そこで求めた輝度勾配の尾根によって囲まれる閉領域を一つの領域として分割する手法である。 図1に一次元の輝度勾配に対する Watershed による領域分割を示す。差分オペレータは画像内のエッジのように信号が変化する部分を求めるため、Watershed に基づく領域分割はエッジと一致した領域を求めることが出来る。本研究ではカラー画像を対象としているが、R, G, B それぞれ独立にソーベル演算子により輝度勾配を求め、その中の最大値をしきい値と比較し

輝度勾配画像を求めることにする。Watershed は大変有効な領域分割手法である。しかしながら、雑音や照明条件等で生じる陰影の影響により、**図2(b)** のように不必要に分割をしてしまう過分割を抑える対策も必要である。次に、均等色空間における領域統合について述べる。

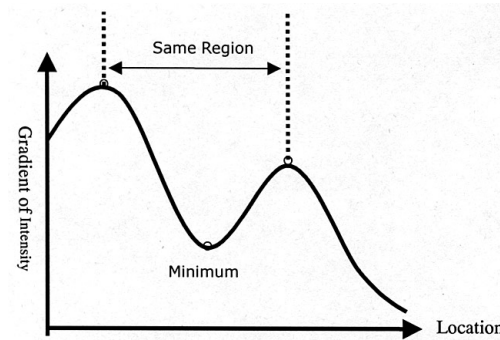


図1 Watershed に基づく領域分割

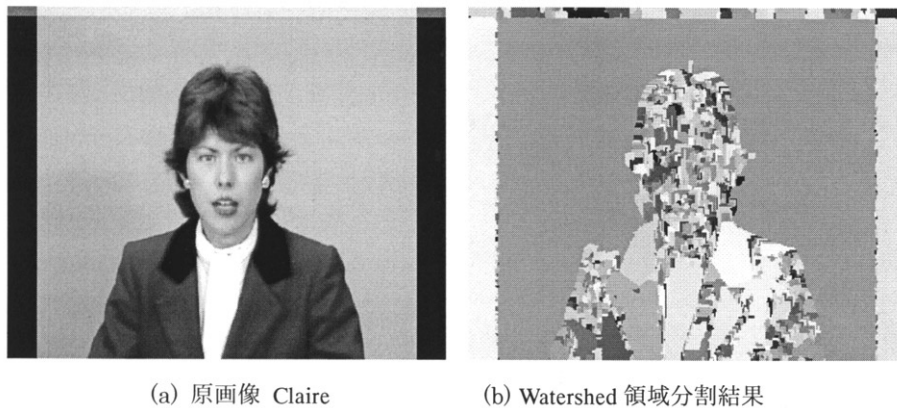


図2 Watershedによる領域分割結果

2.2 均等色空間に基づく領域統合

Watershed により求めた領域分割結果における過分割に対処するため、同一領域と思われる隣接領域を統合するための処理を行う。人間の視覚に近い均等色空間のうち $L^*a^*b^*$ 空間において、隣接した領域各々の平均色の差があるしきい値より小さいとき、隣接領域同士を統合する^[7]。本研究では RGB により表現されているカラー画像を処理対象としている。そこで、 RGB から $L^*a^*b^*$ への変換法について述べる。

$L^*a^*b^*$ 表色系への変換を行うために、まず RGB 表色系を XYZ 表色系に変換する必要がある。 XYZ 表色系は、数学的取扱いを容易にするため便宜的に RGB 表色系から数学的変換により導き出された表色系である。 XYZ 表色系への変換式を式(1)に示す。

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 2.7689 & 1.7517 & 1.1302 \\ 1 & 4.5907 & 0.0601 \\ 0 & 0.0565 & 5.5943 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1)$$

次に、 XYZ 表色系から $L^*a^*b^*$ 表色系へ変換する。 $L^*a^*b^*$ 表色系への変換を式(2)に示す。

$$\begin{aligned}
L^* &= 116 \left(\frac{Y}{Y_w} \right)^{\frac{1}{3}} - 16 \\
a^* &= 500 \left\{ \left(\frac{X}{X_w} \right)^{\frac{1}{3}} - \left(\frac{Y}{Y_w} \right)^{\frac{1}{3}} \right\} \\
b^* &= 500 \left\{ \left(\frac{X}{X_w} \right)^{\frac{1}{3}} - \left(\frac{Z}{Z_w} \right)^{\frac{1}{3}} \right\}
\end{aligned} \tag{2}$$

$[X_w, Y_w, Z_w]$ は完全拡散面の標準光の光による色度であり，NTSC 方式では $[98.072, 100.000, 118.225]$ である。

3. 階層画像における Watershed 領域分割と均等色空間における領域統合

3.1 画像平滑化と階層画像の生成

領域分割を行うための前処理として，処理対象画像に重畳している雑音やテクスチャ成分のパワーを抑えておくことが有効である。そこで，動画の各フレームに対しあらかじめ平滑化処理を行った後，ガウシアンピラミッドに基づいた階層画像を生成する。平滑化処理と階層画像化により，上位階層（粗い）の画像においては，雑音やテクスチャのパワーが抑えられているために，大まかな領域分割結果を導く画像が得られる。一方，下位階層（細かい）の画像においては，正確な境界を有した領域分割結果が得られることが期待される。以降，**3.2** においては各階層それぞれの画像に対して同様の処理を行うこととし，**3.3**において階層ごとの結果を統合したフレーム単位の領域分割結果を得る。

3.2 しきい値選択による複数の輝度勾配画像による Watershed 領域分割

正確にオブジェクトを抽出するためには，あらかじめ領域の境界線がエッジと一致した領域分割が必要となる。この条件を満たす手法としては，Watershed を用いた領域分割がよく知られている^[4]。しかしながら，Watershed による領域分割では画像のエッジ等の輝度勾配データを基に処理を行うため，多くのエッジ成分を有した画像に対しては過分割が発生しやすい。本研究では，平滑化，および，階層画像に対する Watershed 処理により過分割を抑えている。ただし，オブジェクト抽出のための領域分割として，十分に領域の統合がなされていない場合もある。さらに領域分割数を抑制するため，あるしきい値以下の輝度勾配値を不要なエッジ成分として取り除いた輝度勾配画像を新たに作成し，新しい輝度勾配画像を用いた Watershed による領域分割を行う。ここでは，異なるしきい値によって複数の輝度勾配画像を求めておき，それぞれに対して Watershed 領域分割結果を算出しておく。**図3**に，しきい値を 1, 5, 10としたときの Watershed に基づく領域分割結果を示す。しきい値の大小により領域分割結果が大きく異なっていることがわかる。選択されるしきい値が大きいくほど，輝度勾配画像内の詳細な変化成分が取り除かれるため，Watershed による最終的な領域分割は過分割を抑えた結果を得ている。代わってしきい値が小さいほど，より細かい領域まで分割された結果を得ている。複数のしきい値にて作成した輝度勾配画像に対して，それぞれ Watershed 処理による領域分割結果を比較した結果，オブジェクトの境界線等の重要な領域を保存しながら領域統合されていることがわかる。



図3 輝度勾配しきい値選択による Watershed 領域分割

3.3 複数の輝度勾配画像により求めた領域分割結果および複数の階層画像による領域分割結果の統合

各階層において、複数のしきい値選択により求めた輝度勾配画像を用いて得られた Watershed による領域分割結果、および、階層ごとに得られた領域分割結果の統合を行う。まず、選択したしきい値が大きい場合の領域分割結果の境界線を膨張させ、小さい値の場合の領域分割結果の境界線との論理積を求め、途切れた部分の境界線をはさんだ隣接境界の統合を行う。これにより、どの階層においても境界線としている部分の保存がなされ、適切な領域統合が実現される。領域統合の後においても残存した、ある面積以下の領域に関しては、その隣接領域のうち、 $L^*a^*b^*$ 空間において最も平均色の差が小さい領域との統合を行う。次に、階層ごとに得られた領域分割結果に対しても、上記と同様な処理による領域統合を行う。上位（粗い）階層により求めた画像を下位画像と同一の大きさにアップサンプリングした後、境界線を膨張させ、下位（細かい）階層画像の境界線との論理積を求め、途切れた境界線をはさんだ隣接領域の統合を行う。これにより、各フレームにおいて適切な領域統合が実現される。図4(a)に下位階層、図4(b)に上位階層における領域統合結果を示す。Claire に対しては、下位階層、上位階層それぞれに対し、しきい値を10から1まで1刻みで変化させながら計10枚の輝度勾配画像を用いて、領域分割と統合を行った。図5に、階層ごとの領域分割結果をフレーム内で統合した結果を示す。領域が適切に統合されていることがわかる。

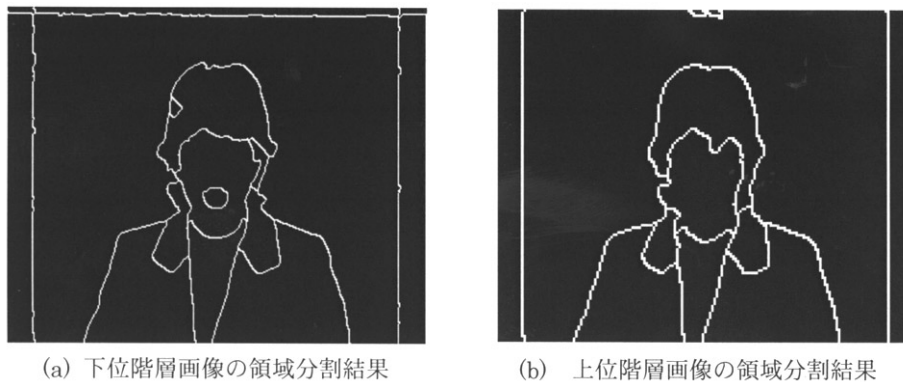


図4 しきい値の異なる分割結果を統合させた領域分割結果



図5 異なる階層画像から得た結果の統合結果

4. チューブの生成とオブジェクト抽出への応用

4.1 領域分割結果からのチューブの生成とチューブに基づく動画の時空間領域分割

フレームごとに独立に行う領域分割は、ノイズや照明の影響により隣接フレームには存在しない孤立した領域を生成しがちである。しかしながら、実際に存在する輪郭線に囲まれたオブジェクト領域に対しては、変形が見られるものの、オクルージョン等が生じない限り隣接フレームにも同一と思われる領域が存在する。そこで、フレームごとに求められた領域の境界線と隣接フレームの境界線とを比較して、隣接フレームにわたって存在する領域に、各フレームで独立して生ずる不要な領域を統合する修正処理を行いながらオブジェクト輪郭の抽出を試みる。あらかじめ、フレームごとに Watershed に基づく一連の領域分割処理を行っておく。次に、前フレームにおける境界線を膨張処理した線と現フレームにおける境界線との論理積をとり、境界線が途切れた隣接領域同士を統合させる。この一連の修正処理をフレームごとに行いながら領域統合を進める^[3]。次に、領域統合を施した領域分割結果を使って、動画からのオブジェクト抽出を行う。分割した各領域内において求めたフレーム間差分値が、雑音の影響により生じていると考えられる差分値よりも大きい領域を動領域として求め、最終的にオブジェクトとして抽出する。チューブを作成し、チューブに基づいたオブジェクトを抽出した例を4.2にて示す。

4.2 領域追跡による動画からのチューブ生成

4.1にて、フレームごとの領域分割結果を、各フレームにて対応する領域の追跡を行いながら、時空間における領域分割結果であるチューブを生成する方法について述べた。ここでは、隣接フレームにおいて対応する領域を追跡するための一手法として、Hausdorff 距離に基づくマッチング手法について述べる^[6]。チューブの生成結果を観測すると、フレームごとに空間的に分割された領域については、オクルージョンや消滅等が生じない限り

隣接フレームにも必ず対応領域が存在する。そこで、求めた領域の境界線と隣接フレームの境界線との間の変形に対して、Hausdorff 距離に基づくロバストなマッチングにより領域を追跡する^[6]。Hausdorff 距離は、2つの集合 $A = \{a_1, \dots, a_p\}$ と $B = \{b_1, \dots, b_p\}$ により構成されているパターン間の距離として、以下の式により定義される。

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (3)$$

なお $h(A, B)$ は、 $h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$ として定義される。式(3)は、注目領域を A 、比較対象領域を B として式(3)の値を規範として、領域 A に対応する領域を探すことを意味する。領域追跡が不可能な、各フレームにおいて独立して存在する不要な領域を検出した場合には、隣接領域と統合させる。時空間的に追跡可能な領域はチューブとして生成する。

4.3 チューブに基づく動画像からのオブジェクト抽出

4.1で説明した方法により動画像の各フレームにおいて Watershed による領域分割を行った後、4.1または4.2の方法により隣接フレームにおいて関連する領域を統合させることによって、時空間に渡って均一な領域であるチューブの生成を行った。動画像 Clair および Table Tennis に対するチューブ生成処理の結果を図6、7に示す。色やテクスチャが均一な領域が一つのチューブとして生成されていることが分かる。



図6 チューブ生成結果 (Clair)

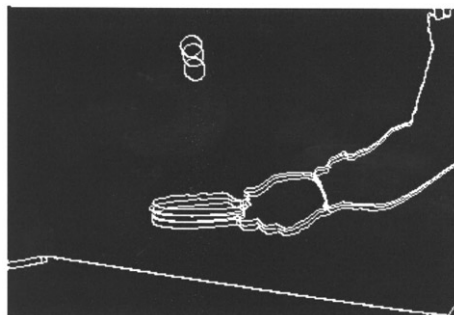
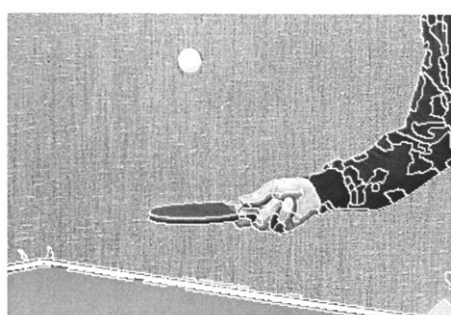


図7 チューブ生成結果 (Table Tennis)

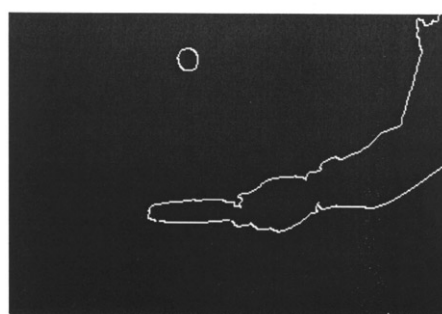
チューブによる時空間領域分割表現では、あるフレームにおいて領域の動きを検出した場合、それ以前、それ以降のフレームでの動きの有無によらずに、チューブに該当する領域が、動オブジェクトの候補として考えられる。さらに、常に動領域同士が隣接している場合、それらが同一のオブジェクトであると判定することも可能である。チューブ表現からオブジェクトを抽出した結果を示すこととする。図8、図9において、動画像 Clair および Table Tennis に対して、生成されたチューブに基づいてオブジェクト抽出手法を適用した結果を示す。Table Tennis では、しきい値を4から1まで変化させた計4枚の輝度勾配画像を用いた。図9(a)は、しきい値3での Watershed 領域分割結果、図9(b)は最終的なオブジェクト抽出結果である。図8、図9(b)から、提案手法によって、動領域が適切に一つのオブジェクトとして抽出されていることがわかる。



図8 オブジェクト抽出結果 (Claire)



(a) 領域分割結果



(b) オブジェクト抽出結果

図9 オブジェクト抽出結果 (Table Tennis)

5. インタラクティブな操作による静止画像からのオブジェクトの抽出

第4章において、動画画像からオブジェクトを抽出するための手法について紹介してきた。その手法は、最初に画像の領域分割を行い、次に隣接フレームと対応する領域を追跡しながら時空間の領域分割結果であるチューブを生成し、最後に、チューブの動きを観測することでオブジェクト抽出を実現するものである。図8、図9(b)の結果より、適切にオブジェクトを抽出する手法が実現された。しかしながら、第4章で提案した手法は、チューブ内に動きが存在していなければセマンティックなオブジェクトとして抽出することが出来ない、複雑な背景を有する動画画像に対してオブジェクト抽出に適切な領域分割やチューブの生成が困難、等の問題がある。そこで本章では、ユーザの最小限のインタラクティブな操作によって、静止画像からセマンティックオブジェクトを抽出する手法について提案する。これにより、複雑な背景を有する画像や、色による領域統合が難しい背景とオブジェクトの色が似通った画像に対しても、オブジェクトの抽出を実現することを目標とする。

インタラクティブな操作とは、ユーザによってマウスやタブレット等の位置を指し示すポインティングデバイスによって、静止画像のオブジェクトの位置を指定することである。たとえば、ユーザの手入力操作によってオブジェクトを抽出する場合、1画素単位の正確さで処理をすることも可能である。しかしながら、静止画像からオブジェクトを抽出する場合においてもユーザの手間は莫大であり、動画画像の場合では非現実的ともいえる。そこで、最小限のユーザ操作によってオブジェクトを抽出する方法を提案する。提案する手法は、Watershed に基づく領域分割を行うと同時に、ある特定のオブジェクトの特徴的な部分の輪郭を、マウス等によって太い線で大まかに指定する。そして、指定した輪郭からオブジェクトの輪郭部分であると推測される領域をマスクとして生成し、用意している判別規則に基づいて、分割された各領域に対してオブジェクト領域と背景領域とに分類しながらオブジェクト領域を抽出するものである。特に、本研究においては、多くの画像で用いられる機会の多い、上半身の人物を含んだ静止画像を対象とした判別規則における提案手法の有効性を検討する。上半身人物の顔領域を大まかに指定することによって、上半身人物の抽出を目指す。

5.1 インタラクティブな操作によるオブジェクトマスクの生成

上半身の人物を含んだ静止画像に対して、マウス等のポインティングデバイスを用いたユーザのインタラクティブな操作によって顔領域の輪郭を指定する方法、および、オブジェクトか背景かを判断するために作成されるオブジェクトマスクの作成法について提案する。顔領域の輪郭線を指定する場合、大まかな操作によっても適切な処理が行えるように、ある程度の幅を持たせた太い線によって囲むことにする。ここで用いる適切な線の太さは、画像の大きさやオブジェクトの大きさに依存する。例えば、本研究ではインタラクティブ操作に用いる線の太さ L を

$$L = \sqrt{S}/50 \quad (4)$$

として求めることとする。但し、 S は画像の面積とする。

次に、顔領域の輪郭は、設定された幅 L の線により、**図10**に示すようにマウス等のデバイスを用いたユーザのインタラクティブ操作により指定する。入力した顔領域の輪郭線から、顔の横幅 m_x 、顔の長さ m_y 、および、顔領域の重心 $G(c_x, c_y)$ を求める。重心 $G(c_x, c_y)$ は顔領域内の全ての画素数を N とすると、座標 $(x_i, y_i) (i = 1, \dots, N)$ の平均値として求めることができる。つまり、

$$(c_x, c_y) = \left(\frac{1}{N} \sum_{i=1}^N x_i, \frac{1}{N} \sum_{i=1}^N y_i \right) \quad (5)$$

と算出される。

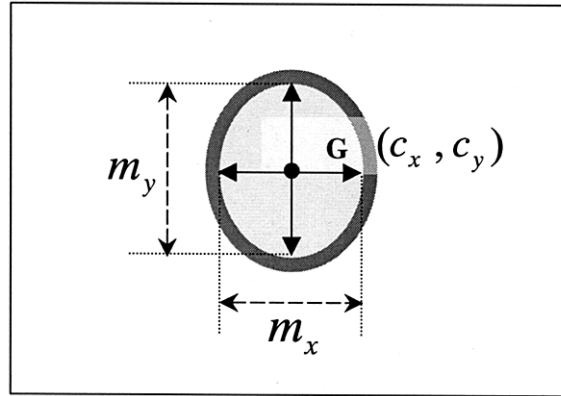


図10 顔領域の指定と重心位置の決定

次に、求めた顔の横幅 m_x 、顔の長さ m_y 、および、重心 $G(c_x, c_y)$ を元にして、肩の位置 $S(s_x, s_y)$ を

$$(s_x, s_y) = (c_x \pm am_x, c_y - bm_y) \quad (6)$$

と定めた。ここで、 $a=1.4$, $b=0.7$ 程度に設定した。肩の位置の一例を**図11**に示す。そして、決定された肩の位置から、傾き c (または $-c$) の直線を顔輪郭線と交差するまで伸ばす。同様に、肩の位置から傾き d (または $-d$) の直線を下方向に伸ばす。ここでは、 $c=0.5$, $d=4$ と設定した。その結果得られた幅 L の線を人物抽出のためのマスクと呼び、**図12**に示す。マスクで仕切られた領域のうち、重心を含んだ側をオブジェクト領域、外側を背景領域として判断する。画像 Claire に対して、顔領域における輪郭線のユーザによる指定の様子を**図13**に、**図13**にて設定した顔領域から人物の領域と考えられるマスクの設定の様子を**図14**に示す。おおまかに顔領域を指定することでも、上半身の人物の形に近いマスクの作成が図られた。

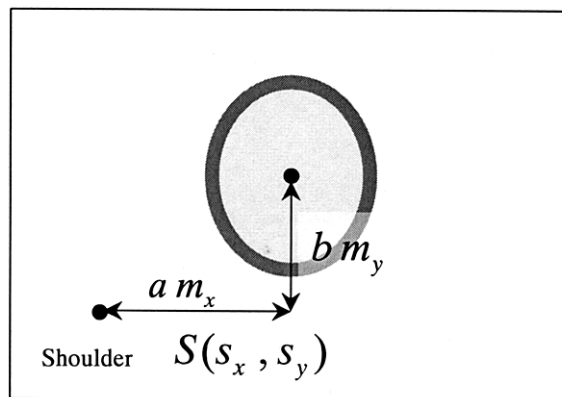


図11 肩の位置の決定

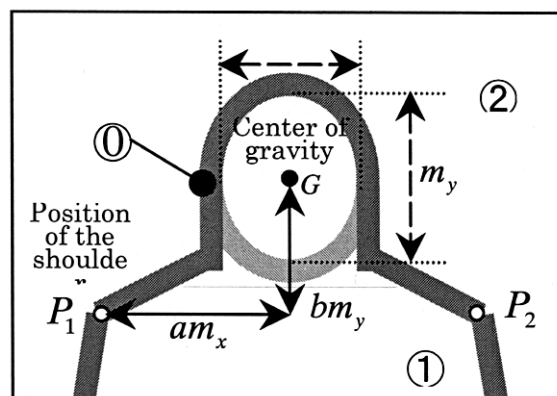


図12 マスクの決定



図13 頭的位置指定例

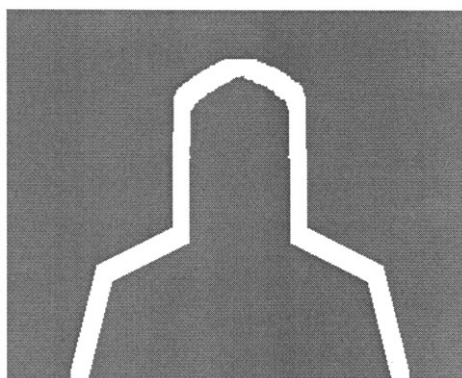


図14 マスク生成例

5.2 作成されたマスクと Watershed 領域分割結果を用いたオブジェクトの抽出

ここでは、**5.1**で求めたマスクを用いて、オブジェクトを抽出する手法について述べる。第3章で説明を行った Watershed に基づく領域分割手法を行うと同時に、**5.1**で求めたオブジェクトに対するマスクを求める。次に、Watershedに基づく領域分割結果とマスクとを重ね合わせる。その結果、分割領域はラベル 1, 2, 0 のいずれか、または、複数のラベルと重なる。人物と背景のどちらかの領域に統合するかの判断は、**表1**に従った判別規則により行うものとする。領域のラベルがラベル0のみの場合は隣接領域との色差を調べ、値が小さいほうの領域と統合する。全てのラベルが含まれる場合はラベル1と2の面積を比較し、多くの面積を占めているラベルがラベル1である場合は人物領域として、ラベル2である場合は背景領域として統合する。

Claire に対する Watershed 領域分割結果を**図15**に示す。この結果と**図14**のマスクの結果とを重ね合わせてオブジェクトを抽出した結果を**図16**に示す。適切にオブジェクトが抽出されているのが確認される。さらに、別の種類の動画像に対する処理例として、Akiyo および Mother and Daughter に対する処理結果を、**図17, 18**に示す。顔領域の輪郭線を大まかに指定するのみのユーザの手間にもかかわらず、効果的にオブジェクトを抽出していることが分かる。ただし、オブジェクトと背景の領域の色が Akiyo の髪の毛と背景のように似通っている場合では、抽出された領域の境界線の精度は良くない。人間の視覚特性に従った工夫、または、統合における判別規則の工夫が必要と考える。

表1 マスクに基づく領域のクラスタリング

Label	Merged Region
Label 1	Object
Label 2	Background
Label 0	Compare color of adjacent regions
Label 1,0	Object
Label 2,0	Background
Label 1,2,0	Compare area of each label

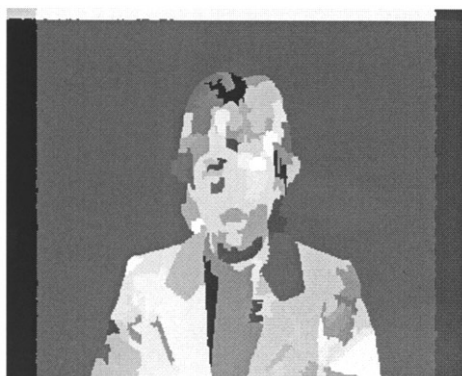


図15 領域分割画像



図16 統合画像 (Claire)



(a) ユーザインタラクション例



(b) 抽出処理結果

図17 オブジェクト抽出処理例1 (Akiyo)



図18 オブジェクト抽出処理例2 (Mother and Daughter)

6. むすび

本研究では、①動画を一フレームごとに領域分割を行い、隣接フレームの領域分割結果と対応する領域を追跡しながら領域統合をすすめることで、時空間にわたって色やテクスチャが均一な性質を有する時空間領域をチューブとして生成する手法の検討、②求められたチューブの動きから意味的にひとかたまりであるセマンティックなオブジェクトを抽出する手法の検討、③おおまかなユーザのインタラクティブな操作によって複雑な背景からも上半身の人物を取り出す手法の検討、等を行った。チューブの生成、セマンティックなオブジェクトの抽出のためには、分割された領域の境界線が正確で、かつ、過分割を抑えた領域分割が必要である。そこで、処理画像から階層画像を作成し、各階層ごとに Watershed に基づく領域分割を行い、さらに、分割された隣接領域との色の近似度を $L^*a^*b^*$ 均等色空間によって判定し、領域統合を行った。これにより、過分割を抑えながら人間の視覚特性にあう領域分割結果を得ることができた。各フレームにおける領域分割結果を、隣接フレームにおいて関連のある領域同士を時空間に渡って領域統合させた。その結果、チューブが生成され動画の構造化を実現した。チューブの動きの有無や隣接チューブとの位置配置を観測することによって、意味的にひとかたまりであるセマンティックなオブジェクトの抽出が可能となった。最後に、ユーザの大まかなインタラクティブな操作によっても、複雑な背景を有した静止画像から上半身の人物の抽出を実現した。最小限のインタラクティブな操作によるオブジェクト抽出の実現は、種々の動画への適用、顔領域の自動検出手法を利用したオブジェクトの全自動抽出への道を開いたといえる。

謝 辞

本研究に対して研究費の助成をいただきました、財団法人 電気通信普及財団殿に深く御礼申し上げます。また、本報告の研究の一部に対して、電気通信大学大学院電気通信学研究科修士課程1年生の服部智広氏に多大な貢献をいただきました。ここに謝意を表します。

参考文献

- [1] L. Vincent and P. Soille: "Watersheds in digital spaces: An efficient algorithm based on immersion simulations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.13, no.6, pp.583-598, June 1991.
- [2] 内堀, 苗村, 金子, 原島: "3次元ベジエ曲面チューブを用いた時空間領域画像の記述," 映像学誌, vol.51, no.10, pp.1688-1695, Oct. 1997.
- [3] 目黒, 金子, 樽松: "フレームごとの領域修正に基づく動画からのオブジェクト抽出," IMPS2001, I-1.02, pp.21-22, Nov. 2001.
- [4] 服部, 目黒, 金子: "インタラクティブな操作による緩やかな拘束を利用したカラー画像からのオブジェクトの抽出," 2002年映像情報メディア学会年次大会講演論文集, 2002.08. 発表予定
- [5] 目黒, 金子, 樽松: "チューブ生成に基づく動画時空間領域分割とオブジェクト抽出への応用," 2001年映像情報メディア学会冬季大会講演予稿集, 12-4, p.131. 2001. 12.

- [6] 目黒, 金子, 樽松: “領域追跡により生成されたチューブに基づく動画像の時空間領域分割,” 2002年電子情報通信学会総合大会講演論文集 (情報・システム部門D-11-84, p.84). 2002.03.
- [7] Kostas Haris, Serafim N. Efstratiadis, Nicos Maglaveras, Aggelos K. Katsaggelos: “Hybrid image segmentation using watersheds and fast region merging,” *IEEE Transactions on Image Processing*, vol.7, no.12, pp.1684-1699, Dec. 1998

〈発 表 資 料〉

題 名	掲 載 誌 ・ 学 会 名 等	発 表 年 月
フレームごとの領域修正に基づく動画像からのオブジェクト抽出	2001年映像メディア処理シンポジウム (IMPS 2001), I-1.02, pp.21-22.	2001.11
チューブ生成に基づく動画像時空間領域分割とオブジェクト抽出への応用	2001年映像情報メディア学会冬季大会講演予稿集, 12-4, pp.131.	2001.12
領域追跡により生成されたチューブに基づく動画像の時空間領域分割	2002年電子情報通信学会総合大会講演論文集 (情報・システム部門 D-11-84, pp.84)	2002. 3
インタラクティブな操作による緩やかな拘束を利用したカラー画像からのオブジェクトの抽出	2002年映像情報メディア学会年次大会講演論文集, 発表予定	2002.08予定