

略画による画像検索への MPEG-7 の応用

青木秀一[†] 青木輝勝[†] 安田浩[†]

[†] 東京大学大学院工学系研究科

〒153-8904 東京都目黒区駒場 4-6-1

E-mail: s-aoki@mpeg.rcast.u-tokyo.ac.jp

あらまし 近年、マルチメディアコンテンツの普及により画像検索の重要性が高まっている。筆者らは、特に目標とする人物を固有名詞で検索し、その後さらに絞り込む場合には自動でタグ付け可能なメタデータを用いた略画による検索が適当であると考えている。しかし、従来の略画による画像検索では精度の高い検索のためには略画を丁寧に描く必要があるなどの問題があった。そこで人物シーンの検索クエリーとして容易に描ける人物を表す略画を画像検索のクエリーとして用いることを検討した。本稿では、メタデータとして MPEG-7 で規定されている記述子のうち自動タグ付け可能な記述子を採用することを前提に、クエリー略画とデータベース中の画像とを適切にマッチングするための略画処理アルゴリズムを提案し、評価実験によりその有効性を示す。キーワード 画像検索、略画、MPEG-7、輪郭線、動き情報

An application of MPEG-7 to picture retrieval with rough sketch

Shuichi AOKI[†], Terumasa AOKI[†], and Hiroshi YASUDA[†]

[†] Graduate School of Engineering, The University of Tokyo

4-6-1 Komaba Meguro-ku, Tokyo, 153-8904 Japan

E-mail: s-aoki@mpeg.rcast.u-tokyo.ac.jp

Abstract The importance of picture retrieval is increasing by reason of spread of multimedia content, these days. Picture retrieval with rough sketch that take automatically tagged metadata, is one of efficient methods, when we retrieve target picture of some person furthermore after retrieving with proper name. Picture retrieval with rough sketch, though, had such problems as its efficiency of retrieval dependent on quality of user's sketch. We propose we take symbolic rough sketch as query for picture retrieval. In this paper, on the assumption that we use some automatically tagged descriptors of MPEG-7, we propose transforming algorithm for matching query rough sketch appropriately. Additionally we give proof of its effectiveness.

Key words picture retrieval, rough sketch, MPEG-7, contour shape, motion information

1. まえがき

近年、マルチメディアコンテンツの普及により動画画像中から所望の画像を検索したいという要求は強くある。筆者らは特に動画画像から所望の人物シーンを検索することを想定した検索システムを検討している。画像検索は検索用のタグをあらかじめ画像に付与しておくことで実現することができる。しかし、テキストによるタグが付けられた画像では、次のような場合には不都合が生じる可能性がある。

- ① 語句による検索をした後のより詳細な絞り込み
- ② 不適切な語句によりタグ付けされた画像
- ③ 一意の語句により表現しにくい画像

①は名詞句で検索可能であるが候補画像が多く提示されてしまい目標画像を目視で確認するのが困難になってしまう場合である。②は地方独特の名称など一般的ではない語句でタグ付けされてしまい、テキスト入力では検索不可能になってしまう場合である。また③は固有名詞や適当な語句では表現しにくい画像の場合である。このような場合には、画像から自動抽出可能なタグを用いた略画クエリーによる検索が有効であると考えられる。ところが、ユーザが画像検索におけるクエリーを入力する際に、丁寧に細部までスケッチを描くことは期待できない。そこで、筆者らは動画画像における人物シーン検索の際に、シンボリックに使われている略画を用いることを検討した。検索対象を人物に限定することで対象の特徴を最大限に利用した実用的な画像検索インターフェースを実現することが可能になる。ただし画像解析、画像認識により抽出される特徴量において独自のものを用いる必要があるシステムではインターオペラビリティの確保という点で好ましくない。そのため独自ではない汎用的なメタデータのみを用いつつ、同時にユーザにとって直感的で容易に描くことのできる略画をクエリーとして、動画画像中の所望の人物シーンを検索することを目的とする。

本稿では、2.において略画による検索の従来研究の問題点を指摘するとともに本稿で採用するメタデータを示す。3.において本研究の目的を達成するための略画処理アルゴリズムを提案する。提案アルゴリズムを実際の略画に適用した結果を4.に示し、本提案の有効性を示す。最後に5.でまとめを行う。

2. 画像検索の関連研究

2.1 略画による画像検索

略画を検索クエリーに用いた研究例として、例え

ば[1]がある。[1]ではシルエット画像を対象とした画像検索プロトタイプシステムを構築している。ここでは、シルエット画像を凸分解し、それらの部分領域の属性及び領域間の関係を対象画像の特徴量とすることで、類似画像検索を実現している。一方、[2]ではウェーブレット変換による多重解像度解析を用いることで、民族資料の画像データベースに対して類似画像検索を行ったことを報告している。しかし、これらはいずれも独自の特徴量を用いるという点で、汎用性に乏しい。また[1]ではユーザの描いた折れ線に対し、属性を指示することができるが、これは人間の直感的な情報ではないと考えられる。[3]において筆者らは日常的に描かれている人物を表すシンボル略画を画像検索のクエリーに用いることを提案した。本稿でもシンボル略画をクエリーに用いることで、ユーザにとって直感的な検索インターフェースを実現することを目的とする。ただし[3]では検索目標とする人物の動き情報を用いていない。目標オブジェクトが動いているか静止しているかの情報を積極的に用いることでより精度の高い検索が可能になると推測される。本稿ではシンボル略画の変換処理の際に、検索目標とする人物の動きの情報を参照する。

2.2 略画検索に用いるメタデータ

従来、動画画像中から対象オブジェクトのみを切り出し、時間方向でその動きを追跡することは困難であった。しかし、任意形状オブジェクトの抽出と動物体領域の追跡はMPEG-4 Core Profileで必要となる技術でもあり、[4]～[6]など動物体の追跡を行う研究が進められている。このような研究が進められることで、注目オブジェクトの色や形、またその重心座標の時間的な変化といった要素は自動で抽出可能となる。本稿ではこのように自動抽出可能な情報のみをメタデータとして利用する。

一方、ISO 標準として勧告されるMPEG-7[7]ではPart3において画像から抽出される特徴量をメタデータとすることを規定している。この中には、オブジェクトの色や動き、また輪郭線形状といった記述子が規定されている。筆者らは、図1に示すような手書き入力によるシンボル略画を画像検索のクエリーとして用いることを検討しているが、このようなシンボル略画から比較的容易にマッチング可能なメタデータとしてオブジェクトの輪郭線の曲率をローパスフィルタによる多重解像度解析を行いメタデータとするContourShapeDSを採用する。

3. シンボル略画の処理アルゴリズム

本稿では動画像からの人物シーンを検索する際にクエリーとして図1のようなシンボル略画を想定している。図1では、目標人物の概形とその動きを示している。つまり図1の場合、人物が画面左側から右側へこれから移動しようとする画像を意味している。

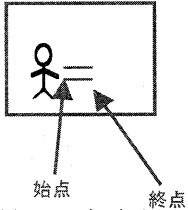


図1 シンボル略画

このように描かれたシンボル略画をユーザの意図する目標画像に対しては類似度という尺度で距離を小さくし、ユーザの意図していない画像に対しては距離を大きくするような変換を行う。変換は

3.1~3.3に示す手順で行う。

3.1 シンボル略画からの座標の取得

ユーザの描いたシンボル略画からは図2に示す7点の座標を取得する。ただし、シンボル略画が描かれた位置により7点の座標が取得できない場合が存在する。そのため、入力クエリーは図3~図7にあるような5パターンに分類する。

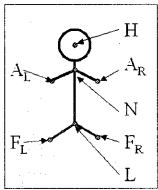


図2 取得する座標

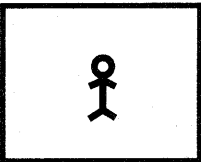


図3 "7点完全" (7点)

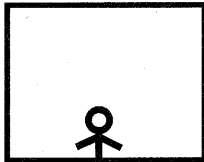


図4 "下欠け" (4点)

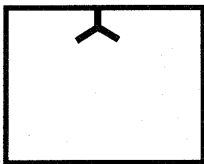


図5 "上欠け" (3点)



図6 "左欠け" (5点)

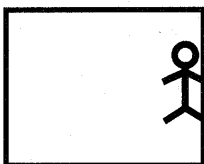


図7 "右欠け" (5点)

まず描画領域の外周がシンボル略画と交わっているかどうかを判定する。交わりが全くない場合は図3の7点完全略画として7点の座標を取得する。下部境界と1点で交わる場合は図4の下欠け略画として4点の座標を取得する。上部境界と1点で交わる場合は図5の上欠け略画として3点の座標を取得し、左境界に1点以上の交わりがある場合には図6に示す左欠け略画として5点の座標を取得する。逆に右境界に1点以上の交わりがある場合には図7に示す右欠け略画のシンボル略画として5点の座標を取得する。また上記のいずれにも該当しないクエリー略画は無効とする。

描かれたクエリー略画からは次のような手順で各点の座標を取得する。

H:描かれた略画の閉曲線部分の重心を H とする。ただし、上欠け略画の場合には H は存在しない。

N:閉曲線から外部へ向かう線分をたどり、最初に左方向に分岐が発生した点を N とする。ただし、上欠け略画の場合には存在しないこととする。

L:閉曲線から外部へ向かう線分をたどり、二度目の左側への分岐または二度目の右側への分岐が発生する点のいずれか先に生じた点を L とする。ただし、下欠け略画の場合には存在しない。また、上欠け略画の場合には、描かれたシンボル略画と上部境界との交点から線分をたどり、最初の左側または右側への分岐点とする。

A_R:点 N から左への分岐をたどり端点を A_R とする。ただし、上欠け略画の場合には存在せず、右欠け略画の場合には、右側境界との交点を A_R' とする。

F_R:点 L から左への分岐をたどり端点を F_R とする。ただし、下欠け略画の場合には存在せず、右欠け略画の場合には、右側境界との交点を F_R' とする。

A_L:閉曲線から外部へ向かう直線をたどり最初に左側への分岐が発生した点をさらに進み端点を A_L とする。上欠け略画の場合には存在せず、左欠け略画の場合には左側境界との交点を A_L' とする。

F_L:閉曲線から外部へ向かう直線をたどり二度目に左側への分岐が発生した点をさらに進み端点を F_L とする。下欠け略画の場合には

存在せず、左欠け略画の場合には左側境界との交点を F_L とする。

手書きで略画がラフに描かれた場合、右腕の付け根と左腕の付け根、また右足の付け根と左足の付け根とが一致しない場合を考慮し、上述のように点 N と点 L との座標を得ることとする。

以上のように、ユーザの描いたシンボル略画から座標を取得する。

3.2 動きの向き取得

シンボル略画の動きは図 1 に示したように、いくつかの動きを表す線分で示される。(以下、動き線と称する。) 本稿では動きによるマッチングについては特に述べない。しかし、動き線の存在の有無、また存在するとしてらいつれかの方向を向いているかは略画処理に必要となる。まず動き線が複数描かれている場合は、始点同士の幾何平均、終点同士の幾何平均で表される座標を結ぶ線分を最終的な動き線とする。

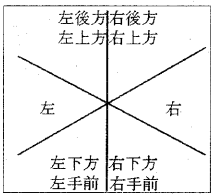


図 8 動きの向き

この動き線が図 8 に示される 6 方向のいずれの方向に属するかを判定する。

動き線が存在する場合には、その方向を 3.3 に示す基準人型モデルの変形の際に用いる。

3.3 基準人型モデルの変形

2.2 で述べたように MPEG-7 で採用されている ContourShapeDS は多重解像度解析によりオブジェクトの支配的な輪郭を優先的に記述する記述子

である。オブジェクトの輪郭の周囲長は 1.0 に正規化されるため、オブジェクトの輪郭の細かい凹凸よりも、大局的な凹凸を輪郭外周上で同じような割合の位置に配置することで類似度が小さい(よく似た)画像を作ることができる。

そこで人体の可動部分、可動範囲は限られていることを考慮し、あらかじめ適当な人体輪郭モデルを用意しておき、そのモデルを入力されたクエリー画像の座標により変形させることとする。データベース中の画像が自然画像であるか、人間の形が極端にデフォルメされたアニメーション画像であるかによって変形させるモデルを交換することで、最適



図 9 自然画用基準人型の例



図 10 アニメーション用基準人型の例

な類似度を与えることが可能である。図 9, 図 10 に基準人型モデルの例を示す。最初から基準モデルをユーザに変形させるクエリー入力方法では、例えば人体の可動部分をマウスでドラッグして変形するなどの方法が考えられるが、クエリーとして直接シンボル略画を描くことで、スピーディーにまた直感的にクエリーを入力することができる。

基準人型モデルは、胴の回転→胴の配置→右腕の配置→左腕の配置→両足の配置、という手順で変形される。

3.3.1 胴の回転と配置

図 1 のようなユーザの描くクエリー画像では、人物が横向きか正面向きかは不定である。しかし、

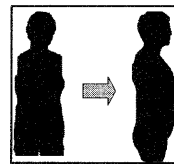


図 11 胴の回転

動き線が左右いずれかの方向へ向いている場合には横向きの可能性が高いと推測される。そこで動き線が左右のいずれかに応じて横向きの基準人型を用いることとする。この場合、右腕と左腕の結合位置、また右足と左足の結合位置は一致させる。図 11 に横向きの基準人型モデルの例を示す。

動き線を参照し向きを決定した基準人型モデルの胴は、入力されたクエリー画像から取得された座標 H および L に基準モデルの相当部分が合致するように拡大縮小し配置する。ただし、上欠け略画の場合には座標 H が存在しないため、座標 L の位置に基準モデルの相当位置を合わせ、胴体部分の半分が境界の外にはみ出すように拡大縮小する。また下欠け略画の場合には座標 L が存在しないが、上欠け略画の場合と同様に座標 H の位置を合わせた上で、胴体部分の半分が下部境界からはみ出すように拡大縮小をする。

3.3.2 腕の向きと配置

上欠け略画の場合には両腕とも配置をしないこととする。それ以外の場合には腕の配置を行う。

クエリー略画より取得した座標から線分 HN と NA_R とのなす角を θ_{AR} とする。 AR が存在しない場合には AR' の座標を用いて θ_{AR} を求める。(図 12)

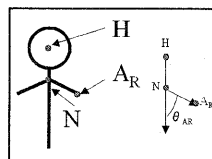


図 12 θ_{AR}

次に θ_{AR} に対し、あらかじめ定められた閾値 θ_1, θ_2 との大小により次のように右腕を回転させて配置する。

右腕の拡大縮小率は胴の拡大縮小率と同じである。

$\theta_{AR} \leq \theta_1$ の時は胴に付ける(図 13 左側)

$\theta_1 < \theta_{AR} \leq \theta_2$ の時は θ_{AR} を保つ(図 13 中央)

$\theta_2 < \theta_{AR}$ の時は垂直にあげる(図 13 右側)

シンボル略画を描く際には、腕をあるていど開くあるいは、両腕を一本の線分で描く傾向があるが、実際には腕が胴体についており、輪郭が凸部とならないことがあるため、上記のように閾角 θ_1 を定めて θ_1 以下の場合には胴体と腕とを接続してしまう。左腕についても同様に配置を行う。

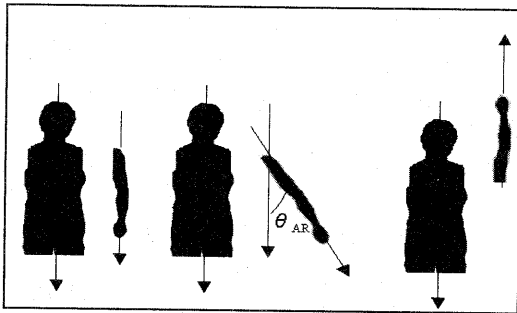


図 13 θ_{AR} に応じた右腕の配置

3.3.3 両足の配置

両足の配置は、左右の区別を行わず両足のなす角度にだけ着目し配置する。ただし、下欠け略画の場合には両足とも配置を行わない。まずクエリー略画より取得した座標から線分 LF_R と LF_L とのなす角を θ_F とする。(図 14)次に、動き線がない

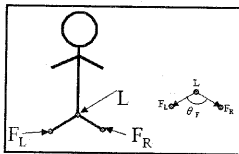


図 14 θ_F

場合には目標画像中のオブジェクトは静止していると推測される。ところが、通常シンボル略画を描くときには、実際の人間の形とは異なり両足を胴体と垂直に描くことは少ない。そこである閾角 θ_3 を定め次のような場合分けを行い両足の開き具合を決定する。また両足の拡大縮小率は胴の拡大縮小率と同じである。

$\theta_F \leq \theta_3$ の時は両足とも垂直位置(図 15 左側)
 $\theta_F > \theta_3$ の時は θ_F を保つ(図 15 中央と右側)

一方、動き線が描かれ方向が左右いずれかの場合には、クエリー画像の θ_F に関わらず、両足は一定角度 θ_4 だけ開いて配置する。ここで θ_4 は例えば平均歩幅、平均足長から導き出せる角度の

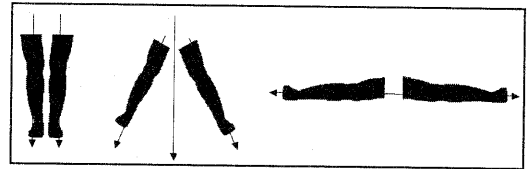


図 15 θ_F に応じた両足の配置

1/2 などを選択する。また右欠け略画、左欠け略画の場合には、両足のうちいずれか一方が領域からはみ出す場合があるが、その場合は左右境界をオブジェクトの輪郭とする。

4. 類似度計算による実験

提案の有効性を確認するため、実際の略画を本アルゴリズムにより変換した。変換前のクエリー画像と変換後のクエリー画像から MPEG-7 VisualPart FDIS[7]に従って、ContourShapeDS の記述子を生成した。また、類似度計算には MPEG-7 Visual part of eXperimentation Model Version 9.0[8]で提示されている類似度計算式を用いた。この類似度計算式ではクエリー画像と比較画像が完全一致の場合に 0 になり、値が大きいほど形状が似ていないことを意味している。

また実際に略画を描く際には、人物を表すシンボル略画と動き線との区別はあらかじめユーザが明示するとした。本アルゴリズム適用に必要な閾角は $\theta_1=45^\circ$, $\theta_2=135^\circ$, $\theta_3=60^\circ$, $\theta_4=55^\circ$ とし、基準人型として図 11 に示したのものを用いた。

目標画像を図 16、図 17 として、それぞれに対する入力クエリーを図 18、図 19 に示す。また本アルゴリズムにより変形された画像を図 20、図 21 に示す。図 18 には動き線が描かれているため、図 20 の変換後のクエリー 1 は横向きとなっている。また目標画像以外の画像として図 22~図 26 の画像に対しても類似度計算を行った。なおこれらの目標画像および目標画像以外の画像は MPEG-7ContentSet の動画像(v19 及び v28)からフレームをキャプチャを行い利用した。

各画像に対して類似度計算を行った結果を表 1、表 2 に示す。



図 16 目標画像 1



図 17 目標画像 2

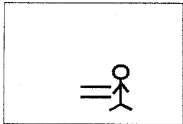


図 18 入力クエリー 1

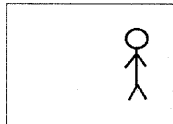


図 19 入力クエリー 2



図 20 変換されたクエリー 1



図 21 変換されたクエリー 2



図 22 目標外画像 1



図 23 目標外画像 2



図 24 目標外画像 3



図 25 目標外画像 4



図 26 目標外画像 5

	入力クエリー 1	変形クエリー 1
目標 1	1.011	0.744◎
目標外 1	1.414	1.004
目標外 2	1.393	0.862○
目標外 3	0.721○	1.483
目標外 4	3.401	1.121
目標外 5	0.508◎	1.986

表 1 入力クエリー 1 に対する類似度

	入力クエリー 2	変形クエリー 2
目標 2	2.361	0.409○
目標外 1	1.808	0.917
目標外 2	1.231◎	3.381
目標外 3	1.853	0.404◎
目標外 4	1.419○	1.574
目標外 5	2.284	2.922

表 2 入力クエリー 2 に対する類似度

表中の◎は類似度が最も小さいため、輪郭線形状が最も似ている画像であり検索の際に第一候補となる画像、○は第二候補となる画像であることを示す。表中の数字から読みとれるように、本変換アルゴリズムが目標画像に対して類似度を小さくし、他の目標外画像に対してはおおむね小さくしないことが分かる。従って、シンボル略画をクエリーに用いた場合に本アルゴリズムを適用することで、精度の高い検索を実現することが可能であることが分かる。

5. むすび

本稿では人物シーンの略画検索に用いるシンボル略画の処理アルゴリズムについて提案を行い、類似度計算によりその有効性を確認した。入力クエリーは手書きを前提としているため、今後は手書きによる描き方のぶれを吸収するアルゴリズムまた、より弁別能力の高い変換アルゴリズムを検討する予定である。略画に描かれた動き線により動き情報によるマッチングを同時に行うことも検討する予定である。

文 献

- [1] 金原史和, 佐藤真一, 濱田喬, "形状分解によるユーザーの視点に基づいたシルエット画像検索," 情報処理学会論文誌 Vol36 No.12 pp.2800-2810(1995)
- [2] 小早川倫広, 星守, 大森匡, 照井武彦, "ウェーブレット変換を用いた対話的類似画像検索と民族資料データベースへの適用," 情報処理学会論文誌 Vol40 No3 pp.899-911(1999)
- [3] 青木秀一, 青木輝勝, 安田浩, "MPEG-7 記述子と画像検索インターフェースの検討," コンピュータビジョンとイメージメディア研究会(CVIM) 2001-CVIM-129-2, Sep., 2001
- [4] 福井和広, 久野義徳, "マルチスネークによる動物体の輪郭追跡," 情報処理学会研究報告コンピュータビジョン No.080-024
- [5] 米山暁夫, 中島康之, 柳原広昌, 菅野勝, "MPEG ビデオストリームからの移動物体の検出," 電子情報通信学会論文誌 D-II Vol.J81-D-II No.8 pp.1776-1786, Aug. 1998
- [6] M.Naemura, A.Fukuda, Y.Mizutani, Y.Izumi, Y.Tanaka, and K.Enami, "Morphological Segmentation of Sport Scenes using Color Information," IEEE Transactions on broadcasting, VOL. 46, NO.3, Sep. 2000
- [7] ISO/IEC JTC1/SC29/WG11/N4358 "Text of ISO/IEC 15938-3/FDIS Information technology Part 3 Visual"
- [8] ISO/IEC JTC1/SC29/WG11/N3914 "MPEG-7 Visual part of eXperimentation Model Version 9.0"