

映像データベースのための被写体形状の取得

林 路彦, 富井 尚志, 有澤 博

横浜国立大学 工学部 電子情報工学科
〒240 横浜市保土ヶ谷区常盤台 79-5

E-mail: {osaru,tommy,arisawa}@arislab.dnj.ynu.ac.jp

あらまし

多様な情報を持つ映像データを取り扱う際に、映像の中に映っている被写体に着目することにより、被写体に関する検索／加工を可能とした映像データベースを実現することが可能となる。映像データベースの実現のための重要な技術として、映像中の被写体形状の取得があげられる。

本研究で取り上げる被写体形状半自動取得システム CHASER は、映像に映っている一つの被写体に着目し、人間が最初におおまかに指定した被写体の外郭線を求め、運動を分析して時間軸方向の前後に被写体を追跡していくシステムである。本稿では、CHASER の構成について述べ、評価を行なう。

A Form Acquisition of Subjects in Image Databases

Michihiko HAYASHI, Takashi TOMII, Hiroshi ARISAWA

Division of Electrical and Computer Engineering
Faculty of Engineering
Yokohama National University
79-5, Tokiwadai, Hodogaya-ku, Yokohama-shi, Japan

E-mail: {osaru,tommy,arisawa}@arislab.dnj.ynu.ac.jp

Abstract

A video image contains various subjects. Capturing such subjects enables us to retrieve and emphasize them on the result movie in Video Image Databases. One of the important techniques to realize Image Databases is a form acquisition of subjects in video frames.

In this study, A semiautomatic form acquisition system 'CHASER' is proposed. It extracts edges of the subject pointed by a user, analyzes the movement of it, and chase it forward/backward. This paper describes an architecture and algorithm of CHASER and its evaluation.

1 はじめに

近年、コンピュータ技術の発達により、プロダクション・エンジニアリング、医療情報処理、人間工学など、多くのコンピュータ応用分野において、映像やグラフィック情報などのマルチメディア情報を取り扱う要望が高まっている [1][2]。それに伴ってこれらのマルチメディアデータを一つの枠組の中で、保存・伝達・蓄積・表現を行なう必要性が生じている。以上のような背景から、マルチメディアデータベースの研究・開発が盛んに行なわれている。

データベースとは、現実世界をモデル化し、そのモデル化されたデータを蓄積するものである。しかし、そのモデル化の過程において、多くの情報が失われてしまう。

そこで、著者らは、現実世界(リアルワールド)に生じた様々なデータをできる限り忠実に取り込み、そのデータに対して、さらに解析・モデル化を行なったデータをも同一の枠組の中に蓄積し、検索操作をすることのできるデータベースシステムとして、リアルワールドデータベース(Real World DataBase: RWDB)をすでに提案している [3]。また、現実世界に生じた事象をできる限り忠実に取り込む手段として映像に着目し、RWDBに基づく映像データベースシステムを構築した。

映像データのモデル化の一つの手法として、映像に映り込んだ被写体にインデックスを付加する方法が考えられる。そこで、本稿では、映像データベースを構成する要素の一つである、映像の解析・モデル化を行なうリアルワールドモデラに着目し、映像の中から意味情報として被写体の形状を取得する CHASER について述べ、その評価を行なう。

2 映像データベースシステム

RWDB の目的は、対象とする世界をできる限り忠実にデータベースに取り込み、記録し、さらに、データベースに記録されたデータに対する種々の解析やモデル化を行なった結果をも同一の枠組の中に蓄積することである。本章では、RWDB に基づいて構成した映像データベースシステムについて述べる。また、データベース検索/加工のインデックスとして被写体を利用するために用いる物理カッ

ト、論理カットの概念を述べる。

2.1 リアルワールドデータベースに基づく映像データベースシステムの構成

RWDB に基づく映像データベースの概要を図 1 に示す。

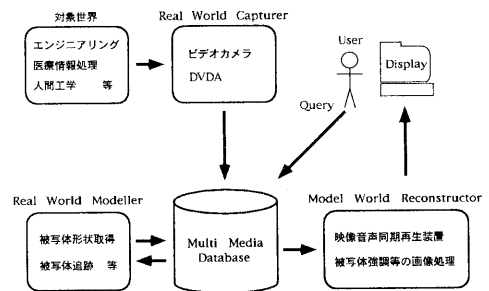


図 1: RWDB に基づく映像データベースの構成図

従来、特に映像などのように、その膨大な情報量をリアルタイムにとらえることは不可能であると考えられてきた。そのため、モーションキャプチャリングなどのようにアプリケーションからの要求に応じて、入力時に非常に少ないデータに変換して蓄積することがよく行なわれてきた。しかし、最近では、フルカラー精密画像を非圧縮でリアルタイムに取り込むことが可能となり、その結果、情報の獲得と、解析・モデル化のフェーズを完全に切り離すことが可能となった。この解析・モデル化システムをリアルワールド・モデラ (Real World Modeller) と呼び、情報獲得システムをリアルワールド・キャプチャ (Real World Capturer) と呼ぶ。このリアルワールド・キャプチャは、対象世界の映像を撮影し、その映像を非圧縮で蓄積しておくシステムで、ビデオカメラ、DVDA 等のハードウェアで構成される。また、リアルワールド・モデラは MMDB に蓄積されている映像データに対して、被写体形状の獲得、時間軸方向の被写体追跡等、目的に応じた解析が行なわれ、そのため、目的別のソフトウェア群で構成されている。

マルチメディアデータベース (Multi Media DataBase: MMDB)[4] にすでに蓄積されているデータ

に対して、リアルワールド・モデラによる解析が繰り返し行なわれ、その解析による新しい結果がスキーマと共に MMDB に蓄積されていく。このため、MMDB には、多様なデータを効率良く表現し、蓄積し、操作を行なうことができる能力が要求される。MMDB には、いろいろな見方やつながりを自由に表現できるという観点から単位となる情報が小さく、かつ平坦な構造を持つモデルが望ましいため、データモデルとして、関数型データモデルである‘AIS モデル’を用いた。

また、リアルワールド・モデラによって解析された被写体領域は、MMDB に対する問い合わせ [5] の際に、検索結果に強調をかけるなどの操作にも用いられる。この過程をモデル再構築 (Model World Reconstructor) と呼ぶ。

このように、被写体の形状を元の映像から抽出し、取得するということは、RWDB の中でも重要な技術であるといえる。

さらに、リアルワールド・モデラで付加されるインデックスで、抽出した被写体領域に基づく情報として、次節では論理カットと物理カットを考える。

2.2 物理カットと論理カット

映像の検索/加工を行なうためには、映像にインデックス付けを行なう必要がある。そこで、物理カット (Physical Cut) と論理カット (Logical Cut) という概念を用いることとする。

物理カットとは、ある一定の意味を持つ区切りで、カメラ割りのように物理的に区切られた範囲を指す。物理カットは連続したフレーム列で、物理カットどうしが重なることはない。

論理カットとは、物理カット内で1つの被写体を表す区切りである。例えば、作業者が車にタイヤを取り付けるという作業があるとすれば、作業者が映っているフレームからフレームアウトする直前のフレームまでを作業者についての論理カットという。もちろん、車やタイヤに関しても同じことがいえる。一つのフレーム内に複数の被写体が映っている場合は当然考えられるので、論理カットどうしが重なり合うことは許される。また、物理カットを越えた論理カットは存在しない。これは、物理カットを越えた時点で同じ被写体でも意味が変化してい

るためである。

物理カット、論理カットのそれぞれの概念を図2に示す。

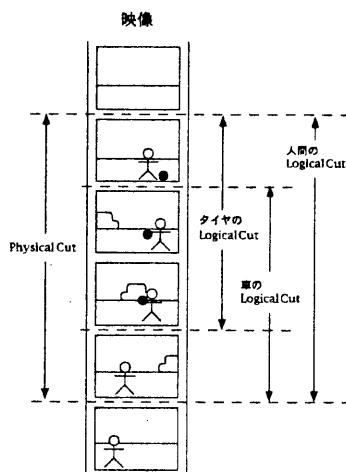


図 2: 物理カットと論理カット

3 CHASER とその評価

RWDB において、映像の解析・モデル化を行なうのが、リアルワールドモデラである。前章で述べた論理カットを映像から抽出するためには、映像から被写体の形状を切り出さなければならない。そのために開発されたツールが CHASER である。

CHASER とは、映像に映っている一つの被写体に着目し、人間が最初におおまかに指定した被写体から、その外郭線を求め、運動を分析して時間軸方向の前後に被写体を追跡していくものである。

CHASER の解析の流れを図3に示す。本手法は、ある1枚のフレームにおいて大まかな被写体の形状の指定を受け、その被写体の形状を求める「静止画における被写体抽出法」と、被写体の形状が与えられている時に、隣接フレームでの同じ被写体の存在領域を求める「映像における被写体抽出法」とに分けられる。

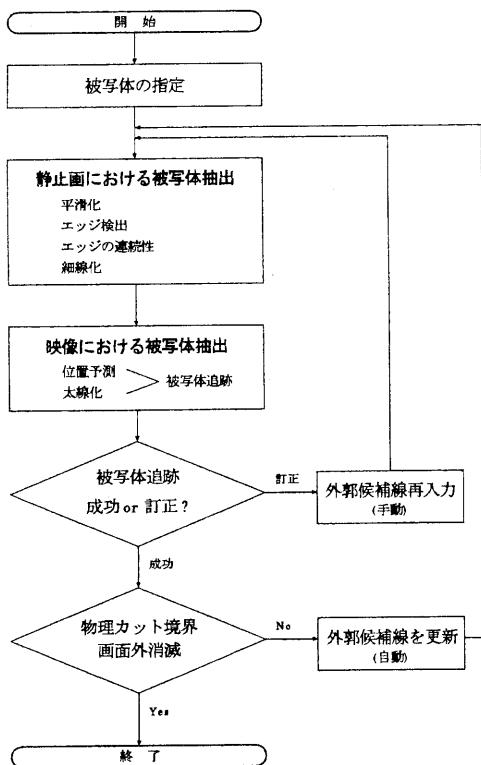


図 3: CHASER の流れ図

3.1 静止画における被写体抽出法

静止画における被写体抽出法は、平滑化、エッジ検出、エッジの連続性、および細線化の4工程に分けられ、以下にそれぞれの手法について述べる。

平滑化

映像には、一般的にさまざまな雑音が含まれている。この雑音を除去するための処理が平滑化である。本研究では、濃度変化の大きい外郭線に着目した。そのため、外郭線の重要な情報である濃度変化の情報を損なわない平滑化手法として、エッジ保存平滑化[6]を用いた。

エッジ保存平滑化とは、ある画素の周辺の局所領域を複数のサブ局所領域に分割し、中心画素と同一の領域を占めるサブ局所領域を選び、そのサブ領域

で平滑化を行なう手法である。

5×5局所領域の中に図4に示すような9種類のサブ局所領域を設定する。1から8は、7画素からなる45度ずつ回転させた領域で、9は中心画素付近の9画素に対応している。1から9の各サブ領域について濃淡レベルの分散を求め、その値が最小となったサブ領域での平均値を中心画素における出力値とする。分散の小さいサブ領域は、中心画素と同じ均一領域に含まれる可能性が高いことを意味する。

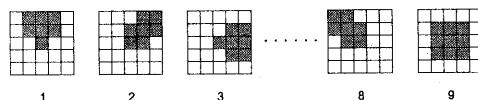


図 4: エッジ保存平滑化の9種類のサブ局所領域

エッジ検出

画像の中で明るさが急激に変化している部分をエッジとみなし、検出する。本研究では、局所オペレータを用いる手法のうち、テンプレート型オペレータの一つである Nevatia and Babu オペレータ[7]を用いた。

テンプレート型オペレータとは、局所領域内で画素の明るさをとる空間微分のための差分型オペレータをあらかじめ多方向について用意し、最大出力を得るマスクから各画素におけるエッジらしさの程度

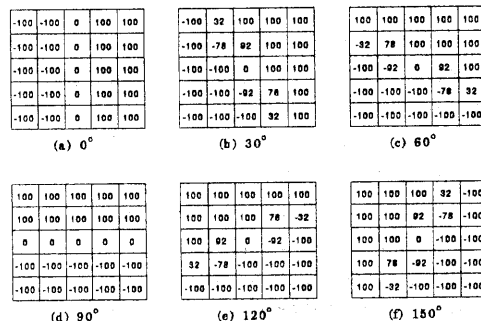


図 5: Nevatia and Babu オペレータ

を表すエッジ強度と、明るさの変化が最も大きい向きを表すグラジエント方向を定めるものである。図5に Nevatia and Babu のオペレータを示す。これは、5×5 のサイズで12 方向を持つマスクである。これらを外郭候補点の各画素を中心とする5×5 の領域で積和計算を行ない、最大出力を与えるマスクからエッジ強度とグラジエント方向を求める。

x_4	x_3	x_2
x_5	x_0	x_1
x_6	x_7	x_8

連結数	画素の分類
0	孤立点または内部点
1	端点
2	連結点
3	分岐点
4	交差点

図 6: 8 近傍 と 画素の分類

エッジの連続性

画像上の明るさの変化は連続であるため、各エッジのグラジエント方向と隣接するエッジのグラジエント方向は同じか、もしくはわずかな角度差しか持たない可能性が非常に高いといえる。このため、検出したエッジ要素のうち隣接するエッジのグラジエント方向との角度差が大きい場合、これを不要な雑音としてとり除く。本研究では、各外郭候補点について求められたグラジエント方向の角度を θ としたとき、グラジエント方向に直角の方向に隣接する2画素のうち、片方だけ θ と 60° 以内の方向を持てばその点におけるエッジ強度を半減し、両方も 60° を超える角度差の場合はエッジ強度を0としている。このような処理によってエッジの品質を高めることができる。

細線化

ある点が削除可能かどうかを決定するために連結数 [8] というものを用いた。

注目画素 (x_0) の8 近傍画素 ($x_1 \sim x_8$) のパターンは $2^8 = 256$ 通りある (図 6)。式 (1) に示す連結数により注目画素を分類することができる。

$$\text{連結数} = \sum_{k=1,3,5,7} \bar{x}_k (1 - \bar{x}_{k+1} \cdot \bar{x}_{k+2}) \quad (1)$$

$$(\bar{x}_k = 1 - x_k, \quad x_9 = x_1)$$

「端点の画素を追加/削除しても図形の連結性は変化しない」[8] という2 値図形のトポロジカルな性質より、閉ループをなす外郭候補線という2 値図形の細線化は、全ての枝を削除していくこととなり、最終的に連結点 (連結数 = 2) のみの集まりである閉ループとなる。よって画素削除の条件、

$$\text{端点 (連結数} = 1) \implies \text{削除可能点}$$

を満たすような外郭候補点なくなるまで、エッジ強度の弱い順に外郭候補点の削除を繰り返すことにより、細線化が行なわれる。

3.2 映像における被写体抽出法

映像における被写体抽出法は、位置予測と太線化の2 工程に分けられ、以下にそれぞれの手法について述べる。

位置予測

次フレームでの被写体の存在位置を予測するためには、まずそれ以前のフレームにおける位置の履歴を知る必要がある。被写体の過去の運動の様子より、現フレームから次フレームへの運動を推定することができる。そこで、図7のように第n フレーム (#n) での被写体 X の座標、速度、加速度を以下のように定める。

$$\text{外接矩形} \quad (L_n, U_n) - (R_n, D_n) \quad (2)$$

$$\text{中心点座標} \quad (C_{xn}, C_{yn}) = \left(\frac{L_n + R_n}{2}, \frac{U_n + D_n}{2} \right) \quad (3)$$

$$\text{速度} \quad v_{L_n} = L_n - L_{n-1} \quad (4)$$

$$\text{加速度} \quad a_{L_n} = v_{L_n} - v_{L_{n-1}} \quad (5)$$

(R,U,D の速度, 加速度はLと同様)

映像上での物体の運動は一般に不規則な加速度運動をしているが、2次元である画像のx,y成分を見ると微小時間では加速度の変化はあまり大きくない。

この方法では、外接矩形の4つの辺がそれぞれ等加速度運動をしていると仮定して追跡を行なう。いま、現フレーム番号がn であるとする。この時、次フレームにおける被写体 X の予想される各数値は等加速度運動という仮定から次のようになる。

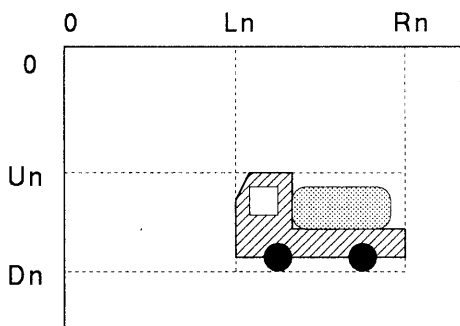


図 7: 被写体の座標

$$a_{L_{n+1}} = a_{L_n} \quad (6)$$

$$v_{L_{n+1}} = v_{L_n} + a_{L_n} \quad (7)$$

$$L_{n+1} = L_n + v_{L_{n+1}} \quad (8)$$

(以上 R, U, Dとも同様)

$$(C_{x_{n+1}}, C_{y_{n+1}}) = \left(\frac{L_{n+1} + R_{n+1}}{2}, \frac{U_{n+1} + D_{n+1}}{2} \right) \quad (9)$$

これにより、#n+1での被写体 X の存在位置が予測される。ただし、これはあくまで予測であり、実際にこの位置に正確に存在するとは限らない。そのため、補正のための太線化の処理を行なう。

太線化

通常、映像中の運動する被写体は、微小時間である隣接フレーム間では位置や形状が極端に変化することはあまりない。しかしあらゆる映像に対し、被写体の位置、あるいは外郭線の変化をある程度知ることができても完全に予測することはまず不可能である。

そこで、被写体が存在すると予測される位置に前のフレームにおける被写体の形状情報を当てはめ、その近辺から真の形状情報を得るために次のような処理を行なう。

位置予測により、#n+1での被写体の位置が予測されている。その位置に#nでの外郭線を移動させ、その周辺に#n+1での外郭線が存在するものとする。ここで移動された線幅 1 の外郭線を、被

写体の運動による適当な量だけ外郭線を太線化して、太くなった外郭線の中に本来の外郭線が含まれるようにする。これが太線化である。求められた太外郭線を#n+1における外郭候補線とすることで、静止画における被写体抽出法によって外郭線を求めることができる(図8)。

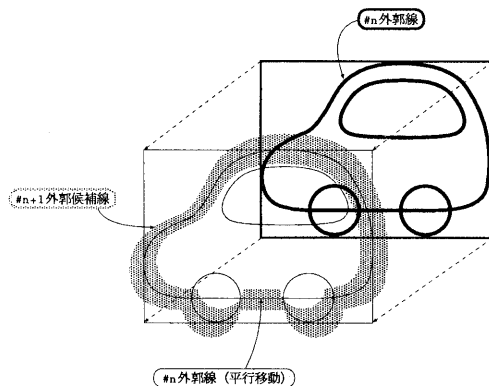


図 8: 太線化の様子

太線化においても細線化と同じように図形の連結性は保持されなければならない。よって前に述べたように「端点の画素を追加/削除しても図形の連結性は変化しない」という 2 値図形のトポロジカルな性質より、膨張処理を行なうために点を付加する際に、その点における連結数を求め、付加する画素が端点であれば図形の連結性は保持されることになる。このことから、太線化のための画素付加条件は

$$\text{端点 (連結数} = 1) \implies \text{付加可能点}$$

となる。

さらに、ここで問題となるのが、どれだけの太さを持った外郭候補線を生成すれば本来の外郭線をその中にとり込めるか、ということである。太線化の量を T で表すと T は被写体の移動量により決まってくる。移動量が多い時は、位置予測の誤差が大きくなってしまいが、それを吸収できるように太線化量も多くする必要がある。また、移動量がごくわずかの時は、太線化量も微小となって外郭線を取り込めない可能性があるため、太線化量は最低でもある

一定値 α を持たせ、 $T \geq \alpha > 0$ とする。

物体の運動は、平行移動、回転、拡大/縮小など、さまざまなものがある。外郭線上の各点は、全体でひとつの動きをするのではなく、個々が違った動きをするので、被写体の外接矩形の中心点の運動のみを追跡するだけでは不十分である。そこで、外接矩形の上下左右の計4辺の運動を追跡することにした。各辺についての、移動速度、加速度より、次フレームでの辺の存在位置の予測がすでに行なわれている。太線化量 $\{T_L, T_R, T_U, T_D\}$ としては、 $\#n$ から $\#n+1$ への予測移動量 $\{v_{L_{n+1}}, v_{R_{n+1}}, v_{U_{n+1}}, v_{D_{n+1}}\}$ を用いた。これは、 $\#n$ から $\#n+1$ の移動した区間を太線化するということである。また、前述の一定値 α を含めると、太線化量は次のようになる。

$$T_L = \max(|v_{L_{n+1}}|, \alpha) \quad (R, U, D \text{も同様})$$

この太線化量に基づいて図9のように、移動した $\#n$ の外郭線の各点の上下左右に外接矩形の上下左右の辺の移動量の膨張処理を行なう。

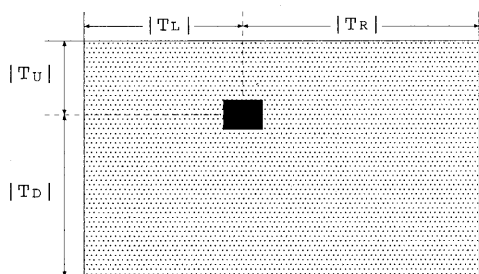


図 9: 外郭線のある点における太線化量

このような手法で追跡するために、過去の運動を知ることができない1枚目での被写体の指定は、できるだけ変化の少ないフレームで指定しなければならない。

また、時間軸の逆方向に追跡することも同様の処理によって可能である。

3.3 評価

CHASER の実行画面を図 10に示す。図 10中の左上のウィンドウが元映像、右下のウィンドウが元映像の中から切り出された被写体のマスクである。

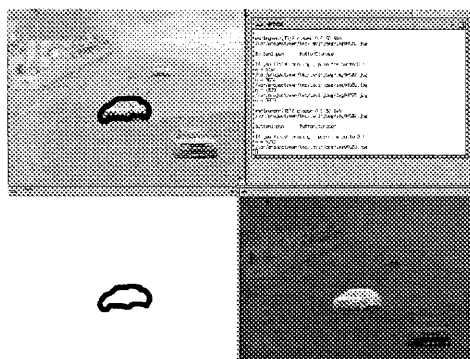


図 10: CHASER の実行画面

CHASER の計算量について考察する。外郭候補点、求められた外郭線を構成する点の数をそれぞれ n, m とすると、静止画、映像の被写体抽出における各処理 (図 3参考) の計算量を図 11に示す。

平滑化	$O(n)$
エッジ検出	$O(n)$
エッジの連続性	$O(n)$
細線化	最良時 $O(n - m)$ 最悪時 $O((n - m)!)$
位置予測	$O(1)$
太線化	$O(m)$

図 11: 各処理の計算量

細線化のアルゴリズムは「端点である外郭候補点が無くなるまで、エッジ強度の弱い順に外郭候補点の削除を繰り返す」ために、計算量のオーダーが一定でない。しかし、ここでの処理の重さは他に比べて格段に軽い処理である。いま、外郭候補線の中心に外郭線があったとすると、外郭候補点のうち、外郭線から離れるほどエッジ強度は小さく、さらに、外郭線から離れた点ほど端点である可能性が高い。すなわち、外郭候補線と背景の境界に近い点ほどエッジ強度が小さい時が最良時で、逆に外郭候補線の中心に近い点ほどエッジ強度が小さい時が最悪時となる。本手法では、外郭線はエッジが強い点で構成されているとし、それを含むように外郭候補線を

生成しているので、最悪時のようになることはまずありえない。仮に最悪時に近付くようなことがある時は、それは外郭候補線の追跡に失敗している時と考えられる。このような時は人間による修正が加わることから、オーダーが大きくなる時は、考慮に入れる必要がないといえる。実際には細線化のための計算回数が、図 12 のような分布を示す。

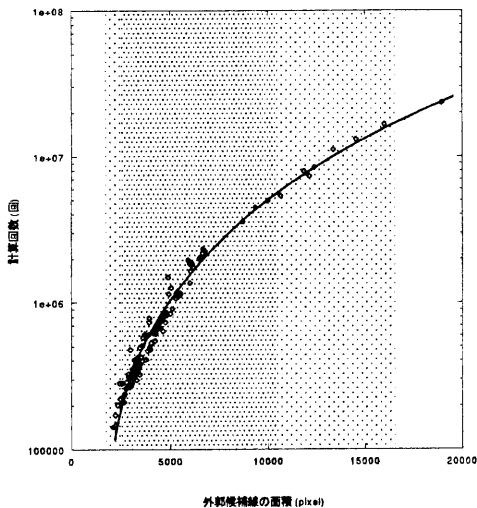


図 12: 細線化処理の計算量

図 12 より、細線化処理の計算量の分布は、ある特定の範囲の中に含まれることが分かる。また、計算量のオーダーは指数関数以下であることがわかる。追跡すべき被写体の面積は、ある程度限られた大きさをもつ範囲に多く含まれるので、外郭候補点の数 n が非常に大きくなる場合についても、考える必要がない。細線化の計算量はシステム全体の計算量にはそれほど影響を及ぼさないといえる。以上のことから、作成したシステムは、非現実的な計算量となることはない。

4 まとめ

被写体に関する検索/加工を可能としたデータベースを実現するための中核技術の一つとして、映像から意味情報として被写体領域を抽出するシステ

ム‘CHASER’について述べ、評価を行なった。

謝辞

CHASER を開発するにあたり、メインプログラマとして上原寛氏 (現日立ソフト) の御協力をいただきました。ここに感謝の意を表します。

参考文献

- [1] 仲町 英治: “ヴァーチャルファクトリー 未来工場への挑戦”, 工業調査会, 1994
- [2] 林 善男, 小木 和孝, 中村 昇太郎, 矢頭 介, 行待 武生: “人間工学”, 日本規格協会, 1990
- [3] 有澤 博: “リアルワールド・モデリングについての考察”, 電子情報通信学会, May 1996
- [4] H.Arisawa, T.Tomii, H.Yui, H.Ishikawa: “Data Model and Architecture of Multimedia Database for Engineering Applications,” IEICE Trans. Inf. & Syst, Vol.E78-D No.11, November, 1995
- [5] 富井尚志, 有澤博: “マルチメディアデータベースにおける映像モデリングと操作言語”, 電子情報通信学会論文誌, Vol.J79-D-II, No.4, 1996
- [6] 高木幹男, 下田陽久監修: “画像解析ハンドブック”, 東京大学出版会, 1991
- [7] 画像処理ハンドブック編集委員会編: “画像処理ハンドブック”, 昭晃堂, 1987
- [8] 横井茂樹, 鳥脇純一郎, 福村晃夫: “標本化された二値図形のトポロジカルな性質について”, 信学論 (D), 56-D11, pp.662-669, November, 1973