

AVIRG

会報

Vol.33 No.2 (1999.9)

発行：視聴覚情報研究会(AVIRG)

代表幹事：伊藤 崇之

〒157-8510 世田谷区砧 1-10-11

日本放送協会放送技術研究所

TEL 03-5494-2361

FAX 03-5494-2371

. 7 月例会報告

「画像中の図形が持つ階層構造の記述」

講演：本谷 秀堅 氏（東京大学工学部）

報告：影広 達彦（日立中央研究所）

《概要と感想》

画像認識における現状は、ある特定分野（文字認識、工業検査、車のナンバー認識等）において実用化されつつあるが、対象物を限定しない汎用的な手法においては、まだ決定打はない。本谷氏の研究は、画像中の特徴の階層構造に着目し、画像認識に有用な図形形状の記述を、対象物の先験的知識なしに行うものである。

今回の本谷氏の講演では、3つの話題があり、

- (1) 地図データからの文字行領域の抽出
- (2) ぼかしを使った階層的特徴抽出[2]
- (3) 輪郭線からの階層的形状抽出[1]

であった。

まず(1)は、対象を地図画像データに限定した実用に近い話であった。この話は本題への導入的な位置付けであったが、画像認識の実用的な応用としては興味深かった。地図画像データには、等高線や地図記号など、文字以外のパターンが多数記載されており、この中から正確に文

字行領域を抽出するのは困難な課題である。詳細な手法はここでは割愛するが、本谷氏は文字行領域抽出の過程を木構造で表現し、領域生成の安定した段階を記録しておき、この記録をフィードバック処理の有用な情報として用いている。ここで言う処理過程の安定した段階が、本谷氏が本題で述べている画像中の特徴を階層化して記述する概念と繋がっていたと思われる。

次に本題である(2)の講演に入るが、ここでは入力画像に対し、意味をもった形状を抽出するには離散化された視野があるという主張であった。ここで言う視野とは、画像認識を行う際の対象物のスケールに対応するであろう。画像データ中に対象物体のスケールがマイクロからマクロまで連続して存在する訳ではなく、ある離散的な視野においてのみ意味のある形状が抽出できる。そして、この離散的な視野のサイズは、対象物に対する先験的な知識からではなく、画像データそのものから得る事が可能であると述

べていた．その方法としては，入力された濃淡画像をぼかしていく過程を時間軸にとり，画像の変化が大きく変化する瞬間に意味のある形状を示してしており，この瞬間を決定するには画像データ中の主曲率を観測している．本谷氏の手法では，ぼかしていく過程における画像の変化に着目しており，対象物のスケールに関する知識は必要ない．これは，画像データからの汎用的な特徴の記述が可能であることを意味する．実験として文字を含んだ実画像からの文字領域抽出を行っていたが，この場合でも文字のスケールを意識せず処理を行うことが可能である．本研究は基礎的な位置付けであるため今後の課題となるであろうが，入力画像に対するロバスト性や，本手法で記述された形状特徴の使い方の検討や提案を期待したい．

次に(3)の講演に入ったが，これは本谷氏が述べるには現在進行中のテーマであった．物体の輪郭線を，荒い形状の部分と細かい形状の部分に分け，それぞれに対し輪郭線を平滑化していく．この輪郭線を平滑化していく過程を観測し，変動の大きい瞬間の輪郭線をプリミティブという部分的な形状として登録する．ここでも(2)と同様に対象物体に対する先験的知識を一切必要としない．本手法により，物体の輪郭線から大局的な形状と詳細な形状の情報を階層的に抽出できれば，画像中における物体認識には非常に有用であると思われる．本谷氏もサンプルとして「鍵」の輪郭線を示していたが，確かに対象物が「鍵」の場合，大局的な形状情報から輪郭線が「鍵」であることを認識してから，詳細な形状から何の「鍵」であるかを認識するニーズがあると思われる．従来の物体認識の発想なら，「鍵」であることの物体認識アルゴリズムと，何の「鍵」であるかを判別するアルゴリズムは分けて考えられることが多く，一貫した手法に

よる形状情報の獲得は難しい．この本谷氏の手法では，どの程度詳細な情報が必要かどうかを意識せず形状情報の抽出が可能であり，パラメータ設定に悩むことは無くなると思われる．ただし，実画像から物体のエッジ抽出から生成した輪郭線には，多数のノイズ，撮影方向による物体の歪み，光源のムラなどの外乱が大きく影響すると思われ，そのような入力における本手法の挙動に興味がある．

本谷氏の研究発表は，昨今のマルチメディアブームのような華やかさはないが，画像認識の根幹に関わる基盤技術として非常に興味深かった．今後，コンピュータが提供する様々なメディアにおいて双方向性を求めていくと，画像認識の底力を要求されるようになり，このような汎用的な認識基盤技術へ大きな期待がかかると思われる．本谷氏は，まだまだ計算時間が莫大であると述べていたが，年々倍増するプロセッサースピードはそれほど先ではない時期にリアルタイム処理を可能にするかもしれない．また，質疑応答でも出ていたが，本谷氏の述べている概念と人間の認知機構との関係を調べることに より，より興味深い現象が説明できるかもしれない．この人間の認知機構との対応付けは，もしかしたら画像認識における大きなブレイクスルーを産むかもしれない．

《参考文献》

- [1] 本谷秀堅，出口光一郎，“スケールスペース解析に基づく局所ぼけ変換を用いた輪郭線図形のマルチスケール近似”，情報処理学会論文誌，vol.35，No.9，pp.1722-1731，1994
- [2] 本谷秀堅，出口光一郎，“濃淡画像の多重解像度解析に基づく図形に固有な形状および大きさの抽出法”，情報処理学会論文誌，vol.39，No.11，pp.3018-3026，1998

「動画像からの3次元情報推定」

講演： 遠藤 利生 氏（富士通研究所）

報告： 木下 敬介（ATR人間情報通信研究所）

《概要と感想》

画像から、3次元情報を抽出するというのは、コンピュータビジョンの基本的な、しかし、最も困難な問題の一つである。そこには、(1)画像間の対応点を求めなくてはならない、(2)容易には精度の高い3次元情報が得られない、といった問題が存在する。本講演は、これらを解決した画期的な研究である。

画像から、3次元情報を復元する方法は様々あるが、広く知られているのは、動きからの3次元情報復元(Structure from Motion)や、ステレオ視である。カメラの前を、剛体の対象物体が運動した場合、画像上で、オプティカル・フローと呼ばれるみかけの動きが生じる。本講演の最初のテーマは、「オプティカル・フローから、いかに精度良く3次元情報を獲得するか」である。

画像から抽出した情報には、必ず誤差が含まれているが、そのような場合でも、3次元情報を精度良く推定する方法が必要とされる。従来は、最適な推定方法として、最尤推定法が用いられてきた。それは、最尤推定法は、不偏であり、かつ、分散が最も小さくなると信じられてきたからである。しかし、本講演では、コンピュータビジョンのいくつかの問題では、最尤推定法よりも分散が小さくなるような推定法があることが示唆された。たとえば、オプティカル・フローからの3次元形状推定などの問題では、使用する点の数が増えると、それと比例して変数の数が多くなり、また、オプティカル・フローに含まれる誤差は、オプティカル・フロー自身と比較して、とても大きいので、統計的な推定問題に素直に帰着させることができないので

ある。

最尤推定を含め、多くの推定問題では、データに含まれる誤差の2乗項までしか考慮しないが、先に挙げたような問題では、誤差が相対的に大きいので、誤差の4乗項を無視することができなくなる。しかし、別の見方をすると、4乗項をうまく活用することで、分散をより小さくできる可能性がある。そこで、2乗項と、4乗項の重みづけを適応的に変化させることにより、不偏なことはもちろん、ほぼ、クラメル・ラオの下界を達成するような推定法が考案された。この重みを計算するためには、対象物体の表面がなめらかであるという拘束条件を利用する必要があるが、普通の対象物体なら満たしている、ごく一般的なものである。最尤推定法と、提案手法とが、シミュレーション実験によって比較され、提案手法の方が、分散が小さいことが明らかにされた。

コンピュータビジョンにおける、もう一つの困難な問題は対応点探索である。複数の画像間で、一方の画像上の点が、他方のどの点に対応しているのかを決定しなくてはならない。これは、ステレオ視による3次元復元方法の基本となるものである。本講演の次のテーマは「対応点決定と3次元情報復元を、同時に推定する」である。

各画像に、ガウシアン・ラプラシアン・フィルタを施し、極大、あるいは、極小となる点を、特徴点とする。各画像の特徴点のペアをランダムに選び(仮説)、3次元空間から画像への「投影」と、画像から3次元空間への「逆投影」が矛盾しないかどうかを調べる(検定)。矛盾のない組み合わせを見つけることで、カメラパラメー

たと、3次元形状が同時に推定されるのである。

こうやって獲得した3次元形状と、対応点情報を組み合わせて、対象物体の任意視点からの映像を生成する方法が紹介された。3次元形状がそれほど正確でなくても、近傍の画像情報から補間することで、高品質の映像を生成できる。対象物体をターンテーブルの上ののせて撮影した全周囲画像から、任意視点映像を生成するシステムがノートパソコンで実演された。近年の計算機の発達が目覚ましいが、身近にある計算機で、数十秒待つだけで処理が終了してしまうのは驚きである。この技術は、インターネット上の電子ショッピングのカタログ閲覧や電子図書館などにも応用できるであろう。

会場からは、前半の講演に対して、誤差の4乗項を考慮しなくてはいけない場合と、そうでない場合についての質問が出た。多くのコンピュータビジョンの問題では、残念ながら、4乗項まで考慮しなくてはいけないようである。また、後半の3次元復元と映像生成のシステムでは、対象物体の形状を、相対精度0.1%で計算できているそうである。また、カメラキャリブレーションは必要だが、それほど厳密にする必要はない。

理論の導出から、画像処理、プログラミングまで、すべてを一人でこなしているのには敬服

した。今後、このような、アプリケーションも視野にいれた研究がコンピュータビジョンの研究に必要になってくるであろう。しかし、行過ぎると、アプリケーションに特化した汎用性のない研究になってしまう。今回の講演は、理論と応用のバランスが非常にうまくとれていたように感じた。

《参考文献》

- [1] Toshio Endoh, Takashi Toriu, and Norio Tagawa, "A Superior Estimator to the Maximum Likelihood Estimator on 3-D Motion Estimation from Noisy Optical Flow", IEE Transactions on Information and Systems, Vol.E77-D, No.11, pp.1240-1246, 1994.
- [2] Takashi Toriu, and Toshio Endoh, "Maximum Likelihood Estimator for Structure and Motion from Multiple Perspective Views", Proceedings of Second Asian Conference on Computer Vision, Vol. 2 of 3, pp.707-711, Singapore, Dec., 1995.
- [3] 鳥生 隆, 遠藤利生, "ランダムな仮説検証に基づく多視点画像からの3次元復元", 電子情報通信学会論文誌, Vol.J82-D11, No.5, pp.909-918, May, 1999.

. 10月例会予定

AVIRG 10月例会は、

日時：10月7日(木)14時~17時

場所：東京大学工学部6号館2F61号講義室
で開催いたします。

テーマは、『音声認識・言語処理』です。講演者およびタイトルは以下の2件を予定しております。奮ってご参加ください。

「MDL基準を用いた音声認識単位の自動生成」

講演者：篠田 浩一 氏

(NEC C&Cメディア研究所)

現在、音声認識においては隠れマルコフモデル(Hidden Markov Model; HMM)を用いた認識手法が主流であり、そこでは、通常、triphone(前後の音素コンテキストを考慮した音素)が基本認識単位として用いられている。しかしながら、一般の言語に出現可能な triphone の総数はたいへん多く、学習データの不足による認識性能の劣化を招く(data-insufficiency problem)。この問題に対処するためには、モデルパラメータのクラスタリングが有効であることが知られており、様々な方法が提案されている。しかしながら、従来、クラスタ数の最適化のための有効な手法はなく、専ら、多くの計算量を必要とする発見的な手法が用いられてきた。この講演では、HMMの状態のクラスタリングにおいて、情報量基準の一つであるMDL基準をクラスタ数の最適化に用いる方法を開発したので紹介する。従来手法と同等以上の認識性能を少ない計算量で実現している。

視覚研究者の方にも音声認識の問題をご理解頂けるよう、背景説明にも十分時間を割く予定です。

《参考文献》

- [1] 韓 太舜, 小林欣吾, "情報と符合化の数理", 岩波講座 応用数学, 岩波書店, 1994.
- [2] 篠田浩一、渡辺隆夫、"情報量基準を用いた音声認識単位の自動生成", 日本音響学会平成8年度秋季研究発表会, 2-3-11, 1996.
- [3] 篠田浩一、渡辺隆夫、"情報量基準を用いた状態クラスタリングによる音響モデルの生成", 電子情報通信学会信学技報 SP96-79, 1996.

「音声認識のための精密かつ頑健な音響モデル」

講演者：中村 篤 氏

(ATR 音声翻訳通信研究所第一研究室)

音声認識において、音響モデルの性能向上は認識性能を大きく左右する重要な問題である。音響モデルとしては、現在、ほとんどの音声認識システムにおいて、隠れマルコフモデル(HMM)により音声スペクトルパターンの挙動を確率的に表現する手法が用いられている。HMMによる音響モデルは、実際の音声データ(学習データ)に基づいて、モデルパラメータの推定(学習)を行なうことによって構築される。より正確な音声認識のためには、精密なモデルが必要であり、そのため、モデルパラメータ数を増やしてモデルの表現力を高める必要がある。しかしながら、パラメータ数に対して不十分な量の学習データで学習を行えば、学習データに特化しすぎた頑健性の低いモデルとなり、逆に認識性能の低下を招いてしまう。すなわち、音響モデルの構築においては、その精密性と頑健性のバランスに留意してパラメータ構造、及びパラメー

タ数を決定する必要がある。

ATR ではこの問題に対して、尤度最大基準に基づくパラメータ・クラスタリング・アルゴリズムを中心とした手法を提案してきた。本発表では、これらの手法について、これまでの経緯や近年の成果を交えながら紹介する。

《参考文献》

- [1] Rabiner et. al., "Fundamentals of Speech Recognition", Prentice-Hall, Englewood, Cliffs, NJ, 1993.
- [2] Sagayama, "Hidden Markov network for precise and robust acoustic modeling," In C.Lee et al., editor, "Automatic Speech and Speaker Recognition: Advanced Topics," pp. 159-184, Kulwer Academic Publishers, 1996.
- [3] Takami et. al., "A successive state splitting algorithm for efficient allophone modeling," Proc. of ICASSP 92, pp. 573-576, March 1992.
- [4] Nagai et al., "ATREUS: A comparative study of continuous speech recognition systems at ATR," Proc. of ICASSP 93, pp. 139-142, April 1993.
- [5] Ostendorf et. al., "HMM topology design using maximum likelihood successive state splitting," Computer Speech and Language, 11, pp. 17-41, 1997.
- [6] Chou, "Optimal partitioning for classification and regression trees," IEEE Trans. PAMI, 13(4), pp. 340-354, 1991.
- [7] Nakamura, "A restructuring of Gaussian mixture density functions in speaker independent acoustic models," Proc. of ICASSP 98, pp. 649-652, May 1998.

～ 会員登録情報の変更のお願い～

AVIRG会員の御所属、会報送付先など登録情報に変更がありましたら、お手数ですが以下のいずれかにご連絡ください。

(財)日本学会事務センター 会員業務係

電子メール(1999年度中) avirg-member@vision.STRL.nhk.or.jp (AVIRG幹事宛)

(注) 会員の確認のために、御氏名とともに、必ず会員番号を明記して下さい。

会員番号および学会事務センターの連絡先は会報郵送時の封筒に印刷されています。