

1993年コンピュータビジョンとパターン認識会議参加報告

浅田 稔⁺, 丸 典明^{*},
田中 弘美⁺⁺, 大谷 淳⁺⁺, 志沢 雅彦⁺⁺,
荒川 賢一^{**}, 堀越 力^{**}, 森 偉明^{**}

⁺ 大阪大学工学部, ^{*} 大阪大学基礎工学部
⁺⁺ ATR 通信システム研究所
^{**} ATR 人間情報通信研究所, ^{**} NTT HI 研

1993年6月, 米国NY市で開催されたコンピュータビジョンとパターン認識会議に出席したので, その概要を報告する.

A Report on CVPR'93

Minoru Asada⁺, Noriaki Maru⁺,
Hiromi Tanaka⁺⁺, Jun Ohya⁺⁺, Masahiko Shizawa^{*},
Kenichi Arakawa⁺⁺, Tsutomu Horikoshi⁺⁺, Takeaki Mori⁺⁺

⁺ Osaka University, ⁺⁺ ATR CSR Lab.,
^{*} ATR HIPR Lab., ⁺⁺ NTT HI Lab.

This report describes an overview of the 1993 Computer Vision and Pattern Recognition Conference, which was held at New York city, USA during June 14th and 17th.

1 はじめに

1993年6月14日から17日にかけてNYマンハッタンで開催されたIEEE Computer Society Conference on Computer Vision & Pattern Recognition(CVPR'93)に参加したので、その概要を報告する。今年の会議では、表1に示す様に20の本セッション・2つのポスターセッションに分かれ、431件の投稿論文のうち、89件の本セッション発表・97件のポスターセッション発表が行われた他、各1件の基調講演、招待講演、パネル討論、2つのワークショップが催され、合計約450名程度の参加があった。また、会議前に日曜と月曜にそれぞれ3つずつのチュートリアルが開催された。

今回のプログラム委員長はMaryland大のAloimonos教授、IBMのBolle博士、Columbia大のBoult教授らで、3氏とも30代半ばまでの若手である。又、プログラム委員会にも、若手(助教授で、3~4年目位)が多く起用され、座長にも女性研究者が多く起用された。(全20セッションのうち6セッションの座長が女性。これは、主催者側から女性研究者への激励であろう)。この分野の世代の交替が徐々に進められていることがうかがえた。

2 Workshop on Qualitative Vision

Qualitative Visionのワークショップは、1990年に開催されて以来、毎年開かれているようであるが、毎回、その定義や限界などが論議され、必ずしも明確にされてこなかった経緯があり、今回もその印象を拭えなかった。

最初にIBMワトソン研究所のJoskowiczが、定性推論の概要を述べ、定性視覚との関係を明らかにしようとした。共通点は高レベルの記述生成であるが、違いは、異なる数学を用いること、定性視覚にとって、幾何学が本質的であり、認識に重きがあること、また動的システムはあまり重要でないということを発言したが、筆者は特に最後のコメントに異論を唱えたい。動的システムの方が、より本質的と考えられるし、認識ばかりが目的ではなからう。

Utah大のThompsonは、定性、定量が必ずしも対峙する概念ではなく、連続的なスケールを持ち得ること、それが目的やコンテキストによって変化することなどを述べ、これからまだ多くの仕事が残っていると主張した。

これらのあと、午前中は8件の「推論と認識」に関する論文発表があった。午後は、FaugerasとSandiniの招待講演があった。Faugerasは、射影幾何に関する論議で、あまり定性推論に関係あるとも思えなかった。Sandiniは、ペンの挿入作業を例に、画像上の定性的な情報のみのフィードバックで、作業が遂行できることを示したが、昨年の本会議(CVPR'92)の目的性視覚(Purposive Vision)のワークショップとの差が、あまり感じられなかった。これらのあと「表現と航行」に関する7件の論文発表があった。特に印象に残ったものとして、物体間の因果関係を推論して、シーンを記述する手法の提案があったが、それら以外は、定性視覚の定義の曖昧性を象徴するかのようになり、必ずしも定性視覚と密接な関係を持たないものも多かった。

Table 1: Contents of Conference

Motion-Based Recognition	4
Discontinuity preserving visual recovery	4
Capabilities of Active Vision Systems	3
Modeling of Shape & Form	3
Model-Based Recognition	5
Real-Time Systems & Hardware	5
Approaches to Recognition	4
Stability of Visual Reconstruction	4
Recognition: Template & Image-Model Matching	6
Scene Recovery	6
Statistical Approaches to Recovery	5
Recognition & Classification of Patterns	5
Segmentation	4
Characteristics of Recognition Strategies	4
Navigation	5
Inspection, Mapping, Coding & Monitoring	5
Determining Parameters of Active Visual Systems	5
Low-Level Vision	5
Capabilities of Partially Calibrated Systems	3
Modeling of Reflectance & Texture	4
Applications	12
Low-Level Vision	17
Recognition	12
Active Vision	8
Motion	13
Grouping & Segmentation	10
その他	27
招待講演	1
パネル	1
Workshop on Qualitative Vision	15
Color Vision Symposium	7

3 Model-Based Recognition

一般円筒モデルとしてSHGCが対象となっていたが、Zerroug & Nevatiaは、より一般的な対象としてのGCの半不変的な投影幾何の性質を解析し、2次元画像から3次元形状パラメータ復元法を示した。断面は円形に限定されるが、軸は直線である必要はない。Huang & Stockmanは一般チューブと称する、長細い物体の認識(断面は円と仮定)について、濃淡情報と輪郭情報を用いた手法を提案し、種々の実画像に適用した。Sullivan, Sandford & Ponceは、2ないし3次元画像を対象として、パラメトリックモデルとのマッチングで、制約つき最適化問題と最小自乗推定の組合せによる幾何距離の最小化問題を扱った。Nakamura, Asada, & Shiraiは、昨年のCVPRで発表した弱一様乱反射面仮定による形状復元をSHGCの再構成に適用し、実画像から形状パラメータを復元した。Mundy, Forsyth, Zisserman, & Rothwellは、回転対象の曲面がその輪郭線から再構成できることを利用して、SHGCの効率的に認識する手法を提案し、実際の画像に適用した結果を示した。

4 Navigation

5件の発表のうち、本来の航行に関すると思われるものが2件しかなかった。Fermüllerは、ノーマルフローのみを用いた3次元運動の定性的再構成法を発表した。フローベクトルを正確に求める必要がなく、誤差に強い手法として注目された。Victor, Samdini, Curotto, & Garibaldiは、蜂の複眼ステレオに習ったセンタリング航行を移動ロボットに適用したデモを紹介した。Cooms & Robertsも、フローベクトルを用いたセンタリング航行法を発表した。いずれも、動物の行動系を模擬するもので、Behavior-basedロボットの研究のがビジョンでも流行する兆しが窺えた。Tomashi & Shiは、従来からのフローのよる解析ではなく、特徴点の視線間の角度の変化から、観測者の方向を推定する手法を提案した。特徴点の視線間の角度が回転に依存しないことから、その影響を取り除くことができる。Soatto, Perona, Frezza, & Picciは、拡張カルマンフィルタを利用した運動と構造推定法を示したが、計算機シミュレーションの結果の呈示のみが行なわれた。(以上、浅田、一部田中)

5 Statistical Approaches to Recovery

5件の発表の内、3件が事後確率最大化(MAP)を適用しており、この他にもDiscontinuity preserving visual recoveryのセッションなどでも数件、MAPを適用しているものが発表されたので、今回はかなりMAPアプローチの形状再構成の研究が集ったようである。ちなみに本セッションの残り2件は、拡張カルマンフィルタ(EKF)と確率的緩和法を用いたものであった。

MIT Media LabのAzarbayejaniらは、カメラの動きが既知でない場合のrelative orientation問題を、対応点の中心投影下での見えの拘束が次々と伝播するようにEKFでモデル化し解く手法を提案した。この方法からTomasi & Kanadeの問題のように、カメラの動きと対象物の3次元構造が得られるが、今回提案のものはTomasiらの方法に比べ対応点の消去/出現にロバストであるということである。もちろん、中心投影を仮定している点でもより一般的になっている。実験結果として、Media Labの建物を撮影したものから3次元構造を復元し、その結果を様々な方向から表示したビデオが示され、再構成結果が優れたものであることがよくわかり非常に印象的であった。

フラクタル表面の再構成法として、Florida大のVemuriらからとNTTのArakawaらからの発表があった。前者はdeformable superquadricをウェーブレット基底により記述し、それが連続的なフラクタル面を構成することを示した。また、その面をMRIによる脳のデータにMAPにより当てはめることにより、構造分割を行なえることを示した。後者では、Szelińskiの手法では滑らかな(すなわち、高周波成分を保存しない)フラクタル面しか再構成できないこと示し、ランダムフラクタルの生成過程の考えに従いGibbs Samplerでの温度を拡張して粗さまで保存するフラクタル面再構成法を提案した。これにより、距離画像データから地形形状をより写実的に再構成できることを示した。

Maryland大のAdbel-Mottalebらから、動画像とステレオの考えを融合してカメラの動きと対象物の3次元構造を得る手法について発表があった。動画像列中の画像ペアについて、FOEを求め相対的な視差マップ作成し、それから事前確率を構成してステレオ問題をMAPで解く方法を示した。モーションステレオにおいて、動画像解析の結果を手がかりにステレオの結果を改良する手法として興味深いものであるが、実験結果の評価が少々不十分である点が残念である。

6 Determining Parameters of Active Visual Systems

今回はアクティブビジョンシステムのためのセルフキャリブレーションのセッションが行なわれ、5件の発表があった。全ての発表で、なんらかのカメラの動きが既知であり、それにより撮られた画像の対応点などの追従によりキャリブレーションを行なうことを想定していた。その場合、目標の3次元座標は知らなくてもいい点、また、動画像を撮影していくにつれカメラパラメータを動的に更新していく点が通常のキャリブレーションと異なる。このセルフキャリブレーションの問題は、ECCV'92でFaugeras, Hartleyらによって解法が発表されて以来、注目されているトピックのようである。

Oxford大のDuらは、水平軸及び鉛直軸周りに回転して撮影した画像から3点もしくは2点のオプティカルフローの差異から、実効焦点距離、画像中心、画素のアスペクト比を推定する手法を示した。しかし、これはレンズ歪みに弱いので、その対応策として、特徴点の動画像間での追従を行ない、その対象軸から画像中心などとともに、放射方向わい曲収差によるレンズ歪み(radial distortion)も推定する方法を示した。Alberta大のBasuは、カメラを水平軸及び鉛直軸の周りに小さく回転させて3つのエッジ上の点を追従していき、実効焦点距離(とアスペクト比)、画像中心を修正しつつ求める方法を提案した。これは、Kanataniによるカメラの回転による方法と異なる不変量特徴の考えに基づいている。MIT AI LabのDronは既知の方向にカメラを並進し対応点の追従を行ない、光学軸のずれまで考慮した一般的な線形カメラモデル(投影行列)を求める方法を示した。また、線形最小自乗法を適用して、画像の入力により順次、推定値を更新していく方法も提案した。

アクティブビジョンシステムの開発が盛んである今日、このようなセルフキャリブレーションはシステム構成の上で避けられない問題である。しかしながら、それぞれの手法の結果評価はなかなか難しい。それぞれ異なるシステムを対象としており、直観的に有効性を示す方法がないので、仕方ないことかもしれないが、例えば、3次元情報の計測を行なうシステムのキャリブレーションであれば、その計測値を既知の真値と比較するなどして有効性を示すべきではないだろうか。(以上、荒川)

7 Motion Based recognition

4件の発表のうち、最初の3件は動作認識、あとの1件は動画を利用した形状認識であった。

ロチェスター大の Polana らは、実際の連続画像から周期性のある動作を検出する手法についての発表であった。動作の認識では、基本動作(歩く、飛ぶ、走るなど)をまず検出する事が重要であるという観点から、時空間画像の濃淡情報をフーリエ解析することで、周期性を持つ動作を検出する手法である。パワースペクトルを基に周期性の判定をしているため、陰影、移動、回転、大きさ、速度に対して不変であり、雑音にロバストな検出が可能である。実画像に対して適用した結果、周期性のある動作の有無の判別は可能であるが、実際の動作は、各基本動作の複合である場合が多いため、この様な動作の識別は、さらに検討が必要であろう。

ハンブルグ大の Rohr は、人の3次元モデルをベースにして、濃淡画像からの歩行動作認識手法を発表した。円筒を部品とした平均的な人物モデルを作成し、キネマティクスによる基本動作を動作抽出の際の制約条件として用いている。実験では、まだ等速で方向転換が無い状態を対象としていたが、フレーム間の姿勢推定にはカルマンフィルターを利用することで、実画像を適用した時の雑音低減を行なっている。デモビデオではモデルと実画像を重ねて動画表示をしており、かなりリアルな動きを見せていた。

東大の Hsu らは、動物体が時空間画像内に形成する曲面を平均曲率やガウス曲率により分類することで、対象物の動きのインデックス化を検討した。手を振る動作の実画像へ適用し、本手法の妥当性を示した。

エール大の Kriegman らは、2Dのビデオ画像から3次元形状を求める手法を提案した。対象物表面の接平面上の直線で複数回対象物と接するような点の集合をHOTカーブと呼んでいる。HOTカーブの種類として、特に bitangent と変曲点を用いている。このカーブは、遮蔽輪郭を有する対象からも抽出できる情報であるため、対象形状の制約はない。ピンホールカメラ仮定の下、カメラの動きが既知という条件で、HOTカーブの3次元空間での位置を求めている。この手法は点と曲線との照合であり、点と点との照合よりは安定に検出が可能である。HOTカーブの軌跡を求めることで、対象物の3次元的な形状を求めている。実画像での実験結果を見る限り雑音に比較的ロバストである。対象の概略的な姿勢推定として利用可能であるが、筆者らが述べているもう一つの目的、認識におけるインデックスとしてはもう少し特徴点が欲しいところである。

8 Approach to recognition

モデルベースビジョンでは、対象範囲が広がるにつれてモデルデータベースの大型化が避けられず、それに伴い対応づけ問題の解決が重要になってくる。このセッションでは、4件の発表があり、その内の3件は、モデルデータベースの巨大化に対する効率的なモデルと対象の照合手法に関する検討であった。傾向として、確率・統計を利用した理論的解釈が目立っていた。そして、3件目の Sengupta らの論文がベストペーパーに選ばれた。

バイズマン工科大(イスラエル)の Basri は対象物の分類(categorization)と同一分類の中での特定化(identification)の2段階からなる物体認識手法を提案した。2D画像から3D物体を認識する手法に“プロトタイプモデル”を

導入し、各カテゴリでの一般的な形状をプロトタイプモデルとして、モデルの特徴点と対象物から得た特徴点との照合を行なっている。実験として、2種類のモデルとの照合をした。プロトタイプと実画像との対応づけが現在マニュアルで行われており、対応づけが今後の課題である。

NTTの Horikoshi らは、AIC(赤池の情報量規準)を部品の統合規準として用いることで、疎らな3次元データからボリュームモデルを求める手法を発表した。ボリュームモデルのプリミティブとして超二次関数を用い、対象を少数の関数の組合せで表現している。輪郭データや距離画像に適用した。物体認識に適用するためには、分割結果の一意性などの検討が必要である。

オハイオ州立大の Sengupta らは、物体認識のためのモデルライブラリの構造として階層的クラスタリング手法を提案した。これは、グラフ理論的なアプローチを取った手法であり、面の形状、穴の数、面と面のなす角などを要素とする相違度を定義し、この相違度に基づいてモデルを階層的(tree構造)に記述している。一般的に、このような階層構造は、ハイレベルの知識処理への適用が可能である反面、計算コストがかかるという欠点があるが、これに対し情報理論に基づいた解析を行い、データベースの効率的な構成方法に関する考察を行なった。実際に100個のCADモデルのデータベースを用いて、モデルの数とスピードアップとの関連性について検討している。

IBMの Rigoutsos らは、Geometric Hashingの手法をベイズ理論を下に解釈し、新たな手法として提唱した。Geometric Hashingとは、NY大の Lamdan らが提案した、対象図形の回転、移動に不偏なマッチングを可能にするモデルベースの認識手法である。筆者らは、この手法が並列に処理可能であることから、複数のワークステーションで分散処理を行う、形状認識システムを作成した。そして、実画像に適用し、提案手法の妥当性を示した。(以上、堀越)

9 Capabilities of Active Vision Systems

本セッションは、アクティブビジョンシステムの能力に関する発表で、バージェンス機能を持ったシステムのためのステレオ視に関する発表1件と、レンズのフォーカスを変化させて入力した画像からのシーンの奥行き情報の復元に関する発表が2件行われた。

バージニア大の Olson は、ロボットと人間におけるステレオ視のためのバージェンス制御とアクティブビジョンの密接な関係について調べた。ステレオ視は視野全体の距離マップを構築するのには向いていなく、シーン中の特定の興味のあるターゲットについて、注視点に関する座標系で情報を集めるのに向いていることに着目し、シンプルで高速なステレオシステムを提案した。ターゲット上に両眼を注視し続けることができると仮定することにより、処理を注視点近傍の狭い奥行き範囲に限定することができる。その結果、必要な計算量を減らし、対応点問題を単純化し、選択的な処理を可能にし、視差フィルタリングによる分割を助けることができる。従来の研究では、ステレオ視のゴールは特徴点或は表面の精密な距離マップを作成することであったが、この研究では異なったステレオ視の役割を仮定

している点が異なる。即ち、ステレオ視は主として一つの注視に対して相対的な距離情報を求める物で、セグメンテーションや認識のための定性的な形状表現の作成を助ける物であり、これは精密なキャリブレーションや注視角度の精密な知識がなくても行えると主張している。

ニューヨーク州立大学のSuryaらは、異なった口径のレンズを用いて撮像された2枚の画像のみを用いて、シーン中の全ての点の距離を決定する方法を提案した(STMAPと呼ぶ)。これは、画像の焦点をずらすことにより物体の距離を決定するための空間領域のコンボリューション/デコンボリューション変換(S変換)のアプリケーションである。この方法は、簡単な局所演算のみを必要とする点で従来の方法よりも優れており、容易にパラレルに実行できる。STMAPはロボットの視覚のためのパッシブな測距法や高速な自動焦点法として役に立つと主張している。

カーネギーメロン大のXiongらは、フォーカスとデフォーカスから距離情報を得る方法において、低い精度のハードウェアや不正確な数学モデルによる問題点を改良する方法を提案した。まず、フォーカスからの距離計算においては、しばしば局所解に陥るFibonacci探索の代わりにFibonacci探索とカーブフィッティングの組合せを提案し、高精度のカメラモデルシステムを用いて正確な結果が得られることを示している。また、デフォーカスからの距離計算(Depth-from-defocus)においては、画像のボケのみならず幾何学的なボケを考慮することにより新しいボケ効果のモデルを提案し、より現実的なキャリブレーションモデルを提案した。

10 Poster: Active Vision

本ポスターセッションでは、アクティブビジョンという枠組で8件の発表があった。しかしながら、全てがアクティブビジョンに関する発表ではなく、種々のモジュールの情報の統合化に関する発表や、ダイナミックなビジョンに関する発表も含まれている。

MITのBobickは、シーンの3次元解釈の安定性の簡単な解析により、低レベルの処理の妥当性を調べる一方法を提案した。この論文では特に、オブティカルフローの計算の妥当性を確認する方法を提案している。計算したオブティカルフローと既知の自己運動が与えられれば、観測者中心の距離画像を計算することができる。以前の距離画像の有限の履歴を、現在の観測者中心の座標系に変換することにより、各点で距離の推定値の安定性を解析する。安定性が低い領域は、オブティカルフローの計算の信頼性が低いことを示している。

セントラルフロリダ大のCryerらは、形状復元のための種々のモジュールの統合化、特にステレオ法と陰影からの形状復元方法(Shape from Shading)の統合化について発表した。ステレオの問題は、粗い深さマップしか得られないことであり、Shape from Shading法は、表面のより細かな情報が得られる。粗いエリアと詳細なエリアの両方に対してうまく動作する方法を実現することは重要である。そこで、ステレオからの低周波の情報を保ったまま、それにShape from Shading法により得られた高周波の情報を加える方法を提案した。

トロント大のSafaee-Radらは、ロボットによる組み立て作業のための新しい3次元物体の認識法を提案した。本方法は、2次元のトポロジカルに別の数枚の透視図(標準図)のみを用いて物体をモデル化する方法、標準図を獲得できるようにカメラを移動する方法、獲得した2次元標準図と物体の2次元標準図のライブラリとの間で、マッチングプロセスを実行する方法からなる。本論文では、これらの方法の概要について述べられた。

阪大の丸らは、従来のパッシブなステレオ視の種々の問題点が、ヒューリスティックな拘束条件にあることに着目し、ステレオカメラをアクティブに運動することにより得られる運動視差を用いて、ヒューリスティックな拘束条件を用いずに効率的に対応点問題を解く一方法を提案した。

ウイスコンシン大のKutulakosらは、「オクルーディング境界やエッジパラメータ化を用いて増加的に再構成されるように、任意の滑らかな表面の高密度の画像系列を生成するためには、観測者はどのように視点を変えれば良いのか?」という問題を考え、この問題を解決するために、表面の見え方に基づいて目的を持って視点を制御する定性的な振舞いの集合を提案した。本論文では、Shape from motionパラダイムとPurposive visionパラダイムを組み合わせて、任意の物体のグローバルな3次元表現を求めるための視点制御の振舞いの集合を開発している。

ノースウエスタン大のProkopowiczらは、アクティブビジョンのためのダイナミック網膜を提案した。ダイナミック網膜は、通常望ましくないカメラのジッターを利用して、従来の空間近傍オペレータの代わりにダイナミックな受容野を用いて、画像のコントラストを検出することができる完全に局所的な方法である。脊椎動物の受容野のための光順応モデルに基づいた受容野の反応関数とカメラの運動を用いることにより、空間的な画像のコントラストを計算することができることと、移動物体の運動パラメータも抽出できることを示している。

ATRの大谷らは、温度画像と明るさ画像のステレオのペアを統合化することにより、時系列の距離画像を獲得するための新しい方法を提案した。ステレオの温度画像は同じ温度の領域に分割され、同じ温度の領域に対して外形に基づいたマッチングを行う。外形マッチングから得られた粗い距離データを補間するために、明るさの輪郭、或は明るさのステレオ画像の中のエッジに対してダイナミックプログラミングを行う。マッチングプロセスから得られたピクセルを対応させることにより、その点の3次元座標を計算できる。移動する人間を含んだシーンをを用いた実験結果が示されている。

ジェノア大のGrossoらは、ダイナミックなステレオのための一般的な枠組(モーションとステレオ視を組み合わせた協調システム)を提案した。カメラの注視点をアクティブに制御できる両眼システムは、ステレオと運動解析を有利に利用して相対的な移動方向や環境に衝突するまでの時間を計算することができるかと主張している。本論文では、何枚かのステレオ画像系列に対して計算した両眼視差とオブティカルフローを組み合わせて、衝突するまでの時間の $2\frac{1}{2}$ D表現を得る方法について説明されている。(以上、丸)

11 Qualitative Vision: Capabilities of Partially Calibrated Systems

このセッションはメイン会場で行なわれ、カメラのセルフキャリブレーションと3次元構造の復元について3件の発表があった。最終日の午後のセッションでもあって、聴講席は少々ガラガラで、ちょっぴり淋しい雰囲気だった。

INRIAのMohrらがカメラを動かしながら撮影した画像でカメラのパラメータを求め、さらにシーンの3次元構造を復元する手法を発表した。それはまずPure projection shapeをそのEuclidean representationに置き換える。それからシーン空間中に4点が共面であれば、その4点が2画像への投影点で成す四角形の対角線交点がエビポーラ拘束を満たすという定理から自動的に画像から5個の参照点を探し、Euclidean shape representationを決める。それによってカメラパラメータと物体の3次元構造を求めることができる。この手法の特徴はTomasiらのカメラモデルをOrthogonal Projectionと設定している手法と比べて、カメラモデルをPerspective Projectionと設定した。

GE-Corporate Research & DevelopmentのRichardらはお互いに非常に大きい視点角度を成すステレオカメラで撮影した画像ペアに対して、それぞれエビポーラ変換マトリックスという射影変換マトリックスをかけることで、エビポーラ線が参照座標系の x 軸と並行するような画像に変換する手法を発表した。これは要するに2つ光軸の方向が異なるカメラで撮影した画像を、光軸が並行するカメラで撮影した画像のように変換するという一種のアフィン変換である。それによって、画像間のDisparitiesが参照座標系の x 軸方向のみに存在する。しかしそのエビポーラ変換マトリックスのパラメータを求めるには画像間の特徴点の対応つけが必要である。実際の実験ではそれを人手で行なった。

INRIAのZhangはステレオカメラを一回動かして撮影した2ペアの画像に写る4個の点からカメラの外部パラメータとモーションとその点の3次元の位置を求める手法を発表した。これはまず問題をClosed-Form Solutionで解いて、それからNonlinear Minimization法で解を最適化する。この手法はWengらの手法と比べてキャリブレーションに使う点数が少ないものの、カメラの内部パラメータを既知にしているとレンズの歪みを考慮してないため、信頼度が低い。

12 Poster: Low-Level Vision

今回のポスタはメイン会場の周りに貼ってあった。主催者は極力に参加者たちにブレイク時間にポスタを見回すように進めた。しかし、会場の周りは照明条件が悪く、ポスタの内容を見えようとするため一苦労しなければならない。このセッションにおいて、3次元情報の復元・リプレゼンテーション、カメラのキャリブレーション、画像のクラスタリング・ラベリング・デコンポゼーションなどなどのコンピュータビジョンに不可欠の基本技術について17件のポスタがあった。悪いコンディションにもかかわらず、説明する側と聞く側の間には距離感なく、熱烈なディスカッ

ションが行なわれた。ここでは筆者の個人的な興味でいくつかを紹介する。

NTT基礎研のMokhtarianはスペースカーブイメージをTSS(Torsion Scale Space)画像に変換し、マッチングを行なう手法を発表した。TSS画像の変換はまずスペースカーブを弧長パラメータで正規化し、それからその3つの座標関数にそれぞれ σ の小さいガウシアンフィルタを畳み込む。その処理をTorsion zero-crossingが非常に少なくなるまで行なう。マッチングはTorsion zero-crossing contoursの極点を用いて行なう。実験ではノイズな3次元チェアデータをモデルデータとマッチングすることに成功した。

NTT HI研の森らはカメラ視線(カメラレンズ中心と画像上特徴点を通る線)によるシーン空間へのコース・ツー・ファインポーティングでシーン空間のコース・ツー・ファインの3次元構造シリーズを復元する手法を提案した。またカメラ視線の誤差分布を解析し、そのポーティングへの影響を $\nabla^2 G$ フィルタリングで改善できることを示した。実験では異なる角度から撮影した6枚の画像からシーン空間の複雑な3次元構造を復元できた。

DEC Cambridge Research Lab.のSzeliskiらはOccluding contoursからなめらかなサーフェスの3次元情報を復元する手法を提案した。この手法はよく知られているCipollaらのエッジをB-snakesで近似して、それらの量的な差を求めることでサーフェスを復元する手法に対して、Kalman filterで直接画像フレーム間のContourセットを推定し、最後にContourセットから統計的に最適推定値を求める。(以上、森)

13 Real-time Systems & Hardware

従来CVPRで必ずしも重視されていなかった実時間システムとハードウェアについてのセッションが設けられたのは注目に値すると言えるだろう。5件の論文発表があり、三眼視ステレオが1件、shape from motionに関するものが2件、アクティブ方式のレンジファインダが1件、多値画像に対するラベリング処理が1件であった。

shape from motionに関しては、Simon Fraser UniversityのJ. Ensらが、ベルトコンベヤ上を運搬されてくる物体を、固定カメラにより観測し、エッジ画素に対応する部分の3次元情報の取得を高速に行っている。この高速化には、マルチスケール・ピラミッド・アルゴリズムをハードウェア化したものを用いており、1画像あたり約0.5秒の処理速度を達成している。マサチューセッツ大学のDuttaらは、自動走行車の視覚機能実現のため、Single Instruction Multiple Data (SIMD)型アルゴリズムを、彼らが提案し構築したIUA (Image Understanding Architecture)にインプリメントした。距離推定は、画像系列間の対応付け情報と、カメラのモーションパラメータを用いることにより行われる。

やはり、自動走行車 (NAVLAB, Dante ロボット) 等の視覚機能のため、CMUのB. Rossが三眼視ステレオを用いたシステムを開発している。ステレオマッチングには、SSSD (Sum-of-sum-of-squared difference)を用いている。

Warp などの並列計算機を利用した結果、 256×240 画素の画像で処理時間が 0.15 秒を実現している。

名工大の Sato らは、アクティブ方式のレンジファインダとして、空間コード化法の原理に基づいた、小型かつ軽量のものを開発した。 512×256 画素の距離画像が 0.3 秒で獲得可能である。適用範囲は限られると考えられるが、動きを含むシーンなどへの適用も期待される。

New South Wales 大学の Nicol は、多値画像のラベリングをシストリックに行うアーキテクチャを提案しており、カスタム LSI によるハードウェア化も行っている。

実時間システムやハードウェアについての本セッションは、とかく理論に偏りがちだった CV の分野に投げかけられた問題提起と言っては、言い過ぎだろうか。今後、CV を用いると何ができるのか、というチャレンジングな論文が増えることを期待したい。(以上、大谷)

14 Modeling of Shape & Form

Siddiqi らは、任意物体の、心理物理 (psychophysical) に関連付けた機能の違いによる形状変化に注目した、物体の階層的部分分割法 (Neck-Based & Limb-Based) を提唱し、その部分 (Parts) 算出法について紹介した (今年度の CVPR 賞の候補に挙げられた)。実際の部分割箇所の抽出には、その物体のシルエットが細くなるような箇所での曲率が大きい点対や、部分割箇所周辺の接線等の幾何学的特徴の連続性を用いている。この手法は、「自然物体はその本質として、異なる機能の連結箇所 (例えば、手首指の関節部分) では形状が変化する」との心理物理的知見に根拠を求めているようである。物体を parts に分割するための観点を興味深い。この手法に限らず、諸観点から提案された新手法の、システムとしての性能を評価する基準やその方法を確立することが重要であろう。

Szelski, Tonnesen, & Terzopoulos は、Physics-based Surface Modeling の新しい手法として、分子の運動を模した "Oriented Particles" を提案した。1 データポイントを 1 分子として扱い、各点で観測される法線や 3D 曲率を基に、分子の動きを制御する内力と外力を、曲面の局所的平滑性拘束 (co-planarity, co-normality, co-circularity) を満たすように調節することによって、任意の位相の曲面をモデリングするパワフルな手法である。対象物に応じて内力を制御することにより、剛体や柔軟体、さらには流体のモデリングにも適用できる。提案した手法による曲面の三角分割法、CT スキャンデータ (疎なシルエットデータ) に適用した曲面再構成実験結果を示し、モデリング手法としての汎用性を示した。しかし現在は、再構成精度や処理時間等は初期サンプリングに依存するため、3D 曲率に基づく適応的サンプリング法を検討しているとのこと。任意の位相曲面を扱うために分子運動までを持ち出す Physics-based Surface Modeling の妥当性と正当性はともかく、Terzopoulos が率いる研究の進展は早い。

Tanaka らは、複雑な自由曲面形状を効率的に且つ忠実に可視化するための、CV と CG の統合を目指したグラフィック指向のモデリング手法を提案した。距離画像から抽出した法線や 3D 曲率に基づき自由曲面本来の凹凸形状やその不連続性に適応する格子 (面) を用いて、任意に指定された

絶対近似精度 (単位球の n 多角形近似精度) を達成する階層的三角パッチ表現を生成する並列アルゴリズムを紹介した。時間的空間的に上限のある並列局所計算にて曲面を再帰的に三角分割し、観測データから自動的に未知形状の CG (ワイヤフレーム) モデルを生成する。

15 Scene Recovery

Navab らは、既知の 3D 直線集合とその画像に投影された 2D 直線の対応が既知の場合のカメラの位置姿勢決定問題に対して、観測された直線が臨界線集合 (Critical sets of lines) に含まれる場合の、解の非一意性とその最大数を決定出来ることを示し、カメラ位置姿勢問題の解の正当性が保証される領域を明らかにした。

Wu らは、単眼中心投影画像の 3 つの三面角 (trihedral angle) 対応から、3D 物体の位置姿勢決定する問題 (A3A problem) の閉形式解を求める手法を示し、これが一般の AnA 問題の完全閉形式解であることを示した。従来手法に対する有効性として、各三面角が座標系に独立に記述されていることを強調している。

Malik らは、まず近隣点間のテキスチャを用いてアファイン変換を推定する手法を用いて、任意方向のテキスチャの歪み (テキスチャグラディエント) を任意方向のアファイン変換 (集合) モデル化し、テキスチャに対する特定の仮定をせずに、局所形状を復元する手法を示した。

Yang らは従来手法と逆に、fine-to-coarse 方向のマッチングにより、局所マッチングの結果の蓄積をより安定な大局的マッチングに反映させた多層ステレオマッチング法の紹介した。

Laurentini は、ビジュアルハルと呼ばれる凸多角形 (多方向からのシルエットの交差から算出される凸多角形) を用いて、インクリメンタル精度を上げて未知形状を推定する手法。効率的に形状を推定するために動的にシルエット獲得方向を選択するために有効であるかも知れない。

Hebert らは、多方向からの疎な観測データから、センサーの精度性能考慮した上で充分な量の観測点が得られた領域を分割し、その領域内で視点に依存しない曲面再構成手法を提案した。曲面の不連続性に対しては考慮していない。

(1) グラフィックを指向したモデリング手法の発展と、(2) 実時間アルゴリズムの開発、又 (3) パターン認識手法の再検討等の流れを感じた。(1) と (2) は、ハードウェア技術発展に伴ってしばらくは研究が進行するように思う。今回、注目を集めた MIT の Pentland らの 2 件で、"Recursive Estimation of Structure & Motion Using the Relative Orientation Constraint" は、復元された物体表面に、画像 (テキスチャー) を逆投影することにより、テキスチャーマップ済みの 3 次元グラフィックモデル自動生成に応用した。また、"Space-Time Gestures" は、実時間時系列画像解析への、パターン認識アプローチ応用の有効性を示している。

しかし、私の担当したセッション (Scene Recovery) は、コンピュータビジョンでは、従来本流を成す研究分野であるが、実験結果を示すビデオを披露した発表は一件もなく、実データを用いた実験結果すら報告されていないも

のが多かった。(以上, 田中)

16 Discontinuity Preserving Visual Recovery

本セッションでは不連続を保存した表面の復元に関する発表が4件あった。オクルージョンなどによる不連続の扱いは、近年重要視されている問題である。

(1) Figueiredo & Leitão は、線過程を取り入れた表面復元のエネルギー関数を最適化する従来の mean field annealing や Graduated non-convexity アルゴリズムに比較して、整数演算だけで計算でき、かつ、多少最適値に近い結果が得られる simulated tearing と名付けた方法を提案した。ただし、問題の定式化自体は、従来と同じであり、最適化の方法を工夫しただけである。

(2) Shah は、表面の不連続とオクルージョンをエネルギー関数による定式化だけで扱うことができるステレオ整合の方法を提案した。表面の不連続においては、一方の画像では見えていない領域が他方の画像では、隠されて見ることができないという half-occlusion の拘束を、2組の互いに左右対称なエネルギー関数によって定式化し、それから、非線形の拡散方程式を導いて、差分近似で交互に解くだけの簡単なアルゴリズムで比較的良好な結果が得られる事を示している。ステレオの half-occlusion の問題は、近年数件の発表が見られるが、本発表は、エネルギー関数の最小化だけによる定式化を行なっているところが特徴である。

(3) Yacoob & Davis は、3次元の拡散過程を用いて、レンジデータの不連続と滑らかな領域を識別する方法を提案した。ステップ、屋根型の不連続のところでは、拡散が抑制され、密度が高く保たれる性質を用いて、高密度の部分を検出することにより、不連続の検出を行った。この方法の問題は、スパースなデータを扱うことができないことである。

(4) Geiger & Kogler は、統計物理学のスケール変換の手法である「くりこみ」の手法を線過程を用いた正則化に応用し、さらに、そのプロセスを逆に使って、エッジ特徴を保存した画像解像度変換を実現した。くりこみの手法をスケール変換に用いること自体は、Gidas によって提案されており、本研究はその応用である。

17 Segmentation

セグメンテーションのセッションでは、4件の発表があり、すべてモーションの関連であった。そのうち3件を紹介する。

(1) Wang & Adelson は、多層表現によるオプティカルフロー場のセグメンテーション法を提案した。アフィン運動を運動のモデルとし、画像列からアフィン運動の推定とクラスタリングを反復的に行なって多層表現の画像モデルを生成した。この方法は、シーン全体の層構造を抽出することができる。これによって、非常に効率よく動画像のコンパクトな記述が得られる。デモンストレーションでは、多層表現から手前の層を消去して動画像再生することにより、背景だけの動画像を生成していたのが印象的であった。

(2) 複数運動物体が存在するシーンでは、運動推定とシーンの分割問題は「鶏と卵」問題に似ている。なぜなら、シーンを分割するためには、運動情報が必要であり、運動情報を正確に得るためには、事前にシーンを各物体に分割する必要があるからである。H. Gu は、最小記述長規範(MDL)を用いて動画像のエッジ特徴の集合を、静止、並進、回転のモデルに分類しつつ、独立な物体に分割し、この問題に対する一つの解決法を呈示した。

(3) Herlin & Giraudon は、心臓の超音波動画像から、心室の運動を追跡する MRF モデルを提案した。これらは、同一の動画像列に対して、スネークや動エッジ追跡などいままでもいろいろな方法を試みてきている。本方法が今まで試みた中でもっともよい結果が得られたと報告している。

18 Low-Level Vision

初期視覚のセッションでは、5件の発表があった。

(1) Shizawa & Iso は、近年、Freeman & Adelson の Steerable filter やそれを拡張した方法によって複数の方向検出を行う時に避けることが出来ない複数の方向性信号の「干渉問題」を解決する方法を提案した。基本的には、多重オプティカルフローで用いた演算子形式の重ね合わせの原理を用いて、任意階数の偏微分演算子の線形拘束を導いた。この方法では、2方向の方向検出が、最低2階の偏微分情報から計算できる。Steerable filter では、最低でも4階の偏微分をもちいなければならず、それでも干渉作用が残る。

(2) Knutsson & Westin は、入力信号の各点ごとの確信度 (certainty)、カーネルの適用度 (applicability) を畳み込み積に導入し、データの欠落や信頼性を考慮し、かつ、窓関数を各点ごとに変化させることができる様に拡張した。これをもとに、確信度と適用度で重み付けされた結果が得られる正規化畳み込みを定義し、さらに微分演算を行なう正規化微分畳み込みに発展させた。この枠組みは、スカラー、ベクトルを問わず、信号の信頼性を考慮したフィルタを実現できる。信号の補間や、オプティカルフローの計算に応用し、著しい改善が得られた。

(3) Davatzikos & Prince は、正則化項の重みパラメータを、エネルギー関数の凸性を保ちつつ、時間的あるいは、空間的に適応的に変化させる様に拡張した動的輪郭追跡アルゴリズムを提案した。パラメータの変化のさせかたはヒューリスティックだが、実験では、太さをもつ曲線に適用し、良好な結果を得ている。

(4) Deriche & Blaszkia は、彼らが提案してきたモデルベースのかど、稜線、3線交点の特徴抽出法において、計算量を削減するための近似解法を提案した。モデルベースの特徴抽出は、精度は高いが、非常に計算時間がかかり、かつ、あらかじめ、特徴の大体の位置を検出する必要があるため、方法の優劣は、この部分を含めて総合的に評価すべきなのではないかという感想をもった。

(5) Lee, Haralick & Deguchi は、ノイズを含んだデータから曲線の曲率を求める方法において、推定の偏差を補正する方法を提案した。偏差の補正法に関しては、他にもいろいろ研究がみかけられるようになったが、それらとの関連を述べると共に比較をしてほしかった。(以上, 志沢)