

CVPR 91 会議報告

尺長 健 (NTT) 杉本和英 (三洋電機) 奥富正敏 (キャノン)
山本正信 (電総研) 八木康史 (阪大) 鄭 絳宇 (ATR)
志沢雅彦 (NTT) 谷内田正彦 (阪大)

IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR91) が、6月3日～6日、マウイ島 (ハワイ州) において開催された。その概要を、発表者8人が分担して報告する。

Report on CVPR91

Takeshi SHAKUNAGA (NTT) Kazuhide SUGIMOTO (SANYO Ltd.)
Masatoshi OKUTOMI (CANON Inc.) Masanobu YAMAMOTO (Electrotechnical Lab.)
Yasushi YAGI (Osaka Univ.) Jiang Yu Zheng (ATR)
Masahiko SHIZAWA (NTT) Masahiko YACHIDA (Osaka Univ.)

IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR91) was held in Maui, Hawaii, U.S.A., June 3-6 1991. This is a report on CVPR91 by eight participants who presented their papers in the conference.

1. 会議の構成

CVPRはIEEE Computer Soc. 主催の会議であり、米国内におけるこの分野の公開された会議としては中心的なものである。査読は匿名のfull paperで行われる。投稿数が史上最高の491件と多かったため、当初は、long、short、posterの3段階のランクづけを行う予定であったものをregular、posterの2段階として採録件数を増やし、regular paper 98件、poster paper 45件が採録となった。それでも、採録率は、regular paperが20%、全体で29%という過去に例を見ない厳しさであった。ちなみに、前回のCVPR'89は、投稿数327で採録率29%(poster paperは無かった)であった。regular paperの発表は2つの並列セッションで行われ、各発表者の持ち時間は約20分であった。物体認識などのhigh level visionとモーション、エッジ、テクスチャなどのearly visionのセッションが並列のペアとされていた。以下に会議規模・構成を示す。

主催：IEEE Computer Society

期間：1991年6月3日(月)～6日(木)(4日間)

会場：Maui Marriott, Kaanapali Resort, Maui, Hawaii, U.S.A.

General Chairman: Prof. Shariar Negahdaripour (Univ. of Hawaii)

Program Co-Chairs: Prof. Gerard Medioni (Univ. of Southern California)

Prof. Berthold K.P. Horn (MIT)

参加人数：約350人

構成：招待：3件、口頭発表：98件、ポスター：45件

採録率：20%(口頭発表)、29%(全体)

投稿数：491件

セッション構成：2並列18セッション

国別口頭発表論文数：

米：67 仏：7 日：6 加：5 独：3 英：3 伊：2

インド、中国、台湾、ベルギー、イスラエル：各1

授賞論文

最優秀賞
入賞

M.Turk & A.Pentland(MIT Media lab.)

S.Bing & K.Ikeuchi(CMU)

M.J.Black & P.Anandan(Yale Univ.)

M.Tistarelli, E.Grosso & G.Sandini(Univ. Genoa)

なお、次回の会議は1992年6月14日-19日にかけてイリノイ州ChampaignのChancellor Hotel & Convention Centerで開催される。

2. 会議・セッションの概要および感想

8人の報告者が見た感想を中心に、会議および担当するセッションの概要および印象に残った論文について報告する。なお、これらはあくまでも報告者各人の主観に基づくものである。

[谷内田]

会議全体の印象

CVPRは元々米国の国内会議であるが年々外国からの発表件数が増えて、今回もヨーロッパ、日本からの発表や参加者が多数見受けられた。また、この会議の前に、CAD-based VisionのWorkshopが開催されており、そのworkshopから引き続きこの会議に参加する人も多数見られた。

会議の全体的な動向はセッション名で分かると思われるので、Posterを除くregular sessionの名前とセッション数を次にあげておこう。なお、各セッションはほとんどの場合5件の発表から構成されていた。Object Recognition (3), Stereo and Surface Recognition (1), Motion (2), Calibration (1), Low-Level Vision (3), Optical Flow(1), Non-Rigid Motion(1), Texture (1), Active Vision and Perceptual Grouping (1), Architecture and Algorithms (1), Shape from Shading (1), Pose Estimation and Face Recognition(1), Depth from X (1)。こうしてセッション名を並べてみると、物体認識、立体認識、動画処理の発表が多く、コンピュータビジョンと呼ばれる分野の発表がこの会議の中心となってきた印象が強い。また、招待講演として、Topological Computer Graphics Modeling, Remote Sensing Issues for Intelligent Underwater Systems, Small Autonomous Mobile Robots: Sensing and Actionの3件があった。筆者としては、最後の講演が最近のマイクロロボットの動きとも関連して興味深かった。

会場となったマウイ島のKaanapali Resortは世界的なリゾート地の一つで、ホテルのすぐ前は海で、近辺には観光名所も多くスポーツ施設も充実しているというすばらしいロケーションであった。こういう観光地で開催した場合の問題点として、発表会場への出席率の悪さがあげられるが、そういうこともな

く、充実した会議であった。ただ、発表時間が短かったため、質疑応答に十分な時間が取れなかったのが残念である。余談になるが、数年前までは、国際会議では質疑応答が盛んなのに、日本では、どうして質疑応答が不活発なのだろうということが盛んに言われていたが、コンピュータビジョン研究会に関するかぎり、国際会議よりはるかに質疑応答が盛んだという印象を強く持つに至った。今後とも、質疑応答を活性化して、一体何を発表していたのか分からないというようなことはなくして、参加者全員が発表の本質と問題点が理解できるという状態になるよう努力いたしましょう。

〔杉本〕

Object Recognition I (2-D)

”Model Group Indexing for Recognition”を取り上げて紹介する。モデルベースの物体認識における問題点として、幾何モデルと画像データとを、頂点特徴等を用いて照合する際、モデル駆動のトップダウン定量的処理を行なう場合の時間的爆発が挙げられる。これを抑えるために、オフラインで作成したモデルの特徴テーブルに工夫をし、対応候補の絞り込みを効率的に行なっている。4ないし5つの頂点特徴を用いて、電話器、マウス、本等からなる机の上のシーンデータからホチキスを抽出させ、有効性を示している。

Calibration

従来の手法の改良や、アプリケーションにおいて、如何に効率良く、カメラパラメータを決定するかといったものを中心に6件の発表があった。ALVや、ハンドアイといったシステムにおいて、簡単な構成データを用い、カメラモデルを決定する手法や、シーン中の参照点に対する、空間中の他の点の相対位置を、射影幾何学に基づいて算出する手法が従来の改良として挙げられる。又、アクティブビジョンへの利用として、分解能、焦点距離、視野角等の拘束を考慮して、シーンの再構成に有効な、視点位置の自動算出の手法があった。

Poster (Tuesday)

火曜日のポスターセッションは、発表件数は15件で、11時30分から1時まで行なわれた。会場は、ポスターを張る場所がなくなると言うようなことはなく、広さとしては十分であった。午前のセッションの後で、ランチタイム前と言う時間設定であったためか、かなりの賑わいを見せていた。各ポスターとも、セッションの時間が過ぎても、質疑応答が盛んに行なわれていたようである。

〔奥富〕

会議全体の印象

ハワイという恵まれた場所で開催されたせいか、或いは昨年CVPRが開かれなかったせいか、応募論文数が非常に多く(491件)、結果として興味深い発表が多かったように感じられた。海辺でのレセプションやルアウディナーショーをはじめ、全体としてリラックスしたムードでのディスカッション及び研究者間の交流がなされた。

Stereo and surface reconstruction

Shape from shading

Poster (Wednesday)

各セッション5、6件の発表論文から、そのセッション全体の印象なり特徴を述べるのは困難であるが、興味深い発表は幾つか見受けられた。例を挙げると、メッシュを2次元或いは3次元的に変形させて、濃淡画像のサンプリングや、距離画像からの面の再構成を、画像に対し適応的に行う手法(Sampling and Reconstruction with Adaptive Meshes, by D. Terzopoulos and M. Vasilescu)、物体境界上の濃度変化から、複数の光源の方向を推定する手法(Sources From Shading, by Y. Yang and A. Yuille)等がある。

〔山本〕

Motion 1: General

追跡とステレオ動画像解析が3件ずつ発表された。Fletcherらは、自己回帰モデルを使って軌跡を予測し、物体の交差が起こった場合でも追跡可能であることを示した。Kollerらは、自動車を対象にその移動軌跡を抽出し、さらに軌跡を言葉(例えば、ブレーキをかけた、ゆっくり走った、等)に変換した。

Non-rigid motion

最近盛んになりつつある非剛体運動の動画像解析についてセッションが設けられた。Mishra, Goldgof, Huangらは、等角写像モデルをつかって、心臓の運動を解析した。入力動画像として、心臓の3次元形状の時系列を使っている。等角写像とは、変形前と変形後で、曲面の3つの第1種基本量について、それぞれの変化率が一様であることで表される。ゴム風船のような相似変形もその一種である。この変化率を運動パラメータとして心臓の動きを表すことが出来る。さて、運動している心臓面上の点に対応付けられて

いるとし、それぞれのガウス曲率が計算される。このガウス曲率の変化から心臓の運動パラメータを推定している。

Metaxas と Terzopoulos は超 2 次関数を拡張して一般の非剛体運動に適用した。通常、超 2 次関数は変形、スケーリング、テーパリング、ヤベンディングをパラメータとして様々な 3 次元形状を表すことができる。彼らはそれらのパラメータが時間変化するとし、それらの変化率を運動パラメータとし、グローバルな運動パラメータと呼んだ。さらに、物体面をパッチの集合で表したとき、各ノードの微小変位をローカルな運動パラメータとして付け加えた。この非剛体モデルを距離動画像 (今のところシミュレーション画像) に適用した。

Depth from ...

Weldon と Liu は、時空間勾配法で距離を測定したとき、測定誤差の要因を定量的に分析した。Adelson と Wang は単眼視から距離を得る新しいセンサを発表した。焦点の合っていない物体を観測したとき、その像は広がって見える。逆にその広がりから、物体までの距離を得ることができる。ただし、これだけでは正常な位置からの遠近を判定することはできない。そこで、レンズの口径を変化させたとき、物体像の動く方向から遠近を判定している。

[八木]

Motion II (Navigation)

移動ロボットのナビゲーションに関する研究が合計 5 件発表された。視覚センサから見ると、ステレオ視が 1 件、残り 4 件は単眼視を用いている。ただし、Yagi らは、円錐ミラーを用いることで 360 度パノラマ状の視野領域が一度に撮像できる全方位視覚センサを用いている。また今回の特徴として、静止環境を対象とした研究が 1 件だけなのに対し、移動物体を含む環境を扱ったものが 4 件と多かった。静止環境を対象とした研究は、ロボットが未知静止環境内を環境構造物との衝突を避けながら移動し環境マップを生成する方法で、円錐ミラーを用いた全方位視覚センサを用い効率的にマップを生成したところに特徴がある。動環境を対象とした研究では、Tistarelli らによるステレオ視にモーションの考え方を融合し、衝突までの時間と物体の奥行き情報を獲得することでロボットが移動可能な空間を発見する方法や、Sawhney らによる Shallow structure を持つ環境を対象とした物体の表現とトラッキングをアフィン変換を利用し行なった研究、さらに Francois らによるモーションベースのセグメンテーションの研究などがある。これらの研究全体を通して、どの研究もロボット (観測者) の移動に伴う時系列画像を扱っていることが印象的である。

Object recognition II (3D)

発表は 6 件あり、内 2 件は 3 次元の形状復元に関する研究であった。形状復元では、3 次元距離データからソリッドモデルを効率的に求めるために楕円座標系により物体の大きな形状をあてはめ、有限要素法と弾性モデルにより複雑な形状を表現する方法と、部分距離画像から 3 次元物体モデルの認識を行なう方法とが提案されていた。両者とも計算コストが高い印象を受けた。他には、対象物体の概念 (例えば椅子) に対応する表現により物体の認識を行なう方法が提案されていたが、十分な認識結果を得るには現状では不十分と感じた。物体のカテゴリー表現には、5 つの基本要素 Dimension, Relative orientation, Proximity, learance, Stability を用いていた。

[鄭]

Low-level Vision

エッジ抽出を扱った論文が多かった。これらの論文では、ある条件のもとでエッジの数学モデルを作成し、正則化・確率過程などの手法を用いて、エッジ検出を実現している。従来の方法と比べて、ある程度の改善が行われたように思われるが、実験規模が明記されていないため、有効性を評価することは難しい。また、あらゆる方向性を持つエッジフィルタをいくつかの基本方向のフィルタの線形結合で表すという研究の発表もあった。

Active Vision and Perceptual Grouping

二つのテーマについて各 3 件と 2 件が発表された。1 件目の A Quantitative Approach to Camera Fixation では、移動カメラの能動的注視行動に伴うオプティカルフローを調べ、3 次元空間の点を分類する方法が提案された。これは、各瞬間に同一またはゼロの値のフローを生成した点は、それぞれ 3 次元空間中では円 (Equal Flow Circles 及び Zero Flow Circle) 上に存在することを利用したものである。ここで、Zero Flow Circle の内と外では、オプティカルフローの方向が逆になる。しかし、そのように表現した空間中の円の集合は、注視点を除くと、カメラ移動に伴い変化するため、あまり信頼できる情報は得られないと思われる。

また、筆者らはカメラの能動的機能を利用し、ロボットの物体操作を誘導する研究を発表した。与えられたタスクに対して、簡単な注視行動によりカメラの空間での移動を制御する。そのため、ロボットの

眼と手の協力操作ができ、画像処理に関する特徴検出、3次元位置の計測、オクルージョンの回避、そして、ロボットの移動に関する位置制御、経路計画に役立つ。

一方、グルーピングの論文としては、抽出したエッジを動的システムで表し、反復法によりエッジ連結、交点連結、ノイズ除去を行う試みが発表された。

〔志沢〕

会議全体の印象

初期・中期視覚の理論的研究から物体認識へ研究の重心がシフトした印象を受けた。初期・中期視覚に関しては、いろいろな方向に前提条件の高度化(複数化、一般化)が試みられるようになってきた。しかし、この様に条件を複雑にする場合には、研究相互の関係を明確にできる様な体系化のための共通の数学的枠組みを作る必要があるのではないかという感想をもった。

Optical flow

オプティカルフローに関しては、(1)フロー場の不連続、(2)運動透過性(各画像点に複数のフローを抽出)、(3)long sequenceの解析、(4)フローの確率分布の評価、などが会議における傾向としてあげられる。特に、モーションによる物体のセグメンテーションが重要視されている。フロー場の不連続を扱うことを可能にする方法として、最小記述長規範(Minimum Description Length Criterion)[Dengler]、多重オプティカルフローの基本拘束[Shizawa & Mase]、ロバスト統計[Black & Anandan]などが提案されている。これらの方法はどれも同じ様な結果をもたらすが、今後はこれらの手法の間の相互関係を明らかにすることが必要になるであろう。

Texture

テクスチャに関しては biological vision をモチベーションにした研究が多い[Manjunath & Chellappa][Farrokhnia & Jain]。これらのいずれもが Gabor wavelet を画像表現として用いて、テクスチャ画像のセグメンテーションあるいは境界抽出を行っている。

〔尺長〕

会議全体の印象

純粋なコンピュータビジョンの理論研究から、応用(ロボット、顔・人物認識)を意識した研究へと重心の移動が見られる。特に、今回は直前に物体認識関連のワークショップが開催されたこともあり、物体認識に関するセッション・発表件数・聴講者が多かった点が印象的であった。逆に、colorに関するセッションがなかったこと、motion関連のセッションが一時に比べ少ないことも特徴的であった。

発表時間は1件18分と短すぎた。これは、投稿件数の増大にも拘らず採録率をできるだけ高くし、かつ2並列セッションで運営した結果であり、参加者からの評判は良くなかった。次回からは並列セッション数を増やす予定らしい。

Architecture and algorithms

Baringerらは、VME-busに接続できる実時間Radon変換ボードについて報告した。

Object Recognition III (General)

Beinglas-Wolfsonは、関節を持つ物体認識への一般化Hough変換の適用を行い、2次元回転のみを対象とし、関節点を参照点とすることで、1関節を含む物体認識法を実現している。Waynerは、多面体を対象として、直交投影(rotation, translation, scale)において画像情報から計算可能なindex functionにより画像特徴集合とモデルの対応付けを求める方法を提案し、3種類のindex function(うち2つは画像から得られる不変量)を用いた実験結果を報告している。

Pose estimation and Face Recognition

Shakunagaは回転自由度を持つ関節と剛体からなる構造の単眼視姿勢推定について基本問題集合を示すと同時に、推定アルゴリズムを与えた。Haralickらは、P3P問題に対する最初アルゴリズムが1841年に与えられていることを示すと同時に、その後に与えられた種々の解法の間関係を整理した。Kang-Ikeuchiは、従来、法線方向のヒストグラムとして表現されていたExtended Gaussian Imageを複素化し、距離情報を位相情報として取扱うことにより、平行移動をも容易に推定できることを示した。Turk-Pentlandは、顔集合の主成分分析の結果得られる固有ベクトル(Eigenfaces)空間を用いた顔同定の実時間アルゴリズムを提案し、最優秀論文賞を獲得した。顔同定アルゴリズムには問題があるが、動画像から背景除去・顔の切りだしを含めて、実時間での顔同定の可能性を示した点が評価されたのだと思われる。