

第5回コンピュータビジョン国際会議 ICCV'95 報告

和田俊和 (岡山大学)、志沢雅彦 (NTT マルチメディア総合研究所)、
内山俊郎 (NTT データ通信)、藤原久永 (岡山県工業技術センター)、
三浦 純 (大阪大学)、李 七雨 (イメージ情報科学研究所)、
鄭 絳宇 (九州工業大学)、浅田尚紀 (広島市立大学)

twada@chino.it.okayama-u.ac.jp, shizawa@nttcvg.hil.ntt.jp,
uchiyama@lit.rd.nttdata.jp, fujiwara@okakogi.go.jp,
jun@cv.ccm.eng.osaka-u.ac.jp, cwlee@senri.image-lab.or.jp,
zheng@sein.mse.kyutech.ac.jp, asada@its.hiroshima-cu.ac.jp

1995年6月に米国マサチューセッツ州ケンブリッジのMITで開催された第5回コンピュータビジョン国際会議の概要を報告する。

A Report on ICCV'95

Toshikazu WADA(Okayama University), Masahiko SHIZAWA(NTT),
Toshio UCHIYAMA(NTT Data), Hisanaga FUJIWARA(Okayama Industrial Center),
Jun MIURA(Osaka University), Chil-Woo LEE(LIST),
Jiang Yu ZHENG(Kyushu Institute of Tech.), Naoki ASADA(Hiroshima City University)

This report describes an overview of the fifth International Conference on Computer Vision, which was held at MIT, USA in June 20-23, 1995.

1 はじめに

第5回コンピュータビジョン国際会議 (ICCV'95) が6月20日から23日の4日間、米国マサチューセッツ州ケンブリッジのMITで開催された。ICCVはヨーロッパ、アメリカ、アジアの順に場所を移して2年おきに開催されるが、今回は米国IEEEのTechnical Committee PAMIが母体となって会議を主催し、MIT AI labのEric GrimsonがGeneral Chairを務めた。ICCVは論文を厳選しシングルトラックで質の高い発表と議論を行う会議としてCV分野では高い評価を受けている。ICCV'95では応募論文総数600件に対して161件の論文が採択され(採択率は27%)、その内訳は口頭発表56件、ポスター105件であった。表1にICCV'95のプレナリーセッション、ポスターセッション、ワークショップの一覧を示す。会議には22ヶ国から582人が参加した。今回は米国東海岸のMITで開催したため、ICCVとしては通常よりも応募論文数、参加者数とも多かったようである。

ICCVでは最優秀論文にMarr Prizeが授与される。今回は優劣がつけ難いということでColumbia大のM. Oren and S. Nayarの“A theory of specular surface geometry”、および岡山大のT. Wada, H. Ukida, and T. Matsuyamaの“Shape from shading with interreflections under proximal light source - 3D shape reconstruction of unfolded book surface from a scanner image-”の2編にMarr Prizeが贈られた。前者が、鏡面反射特性を持つ物体表面を回転撮影し、その画像系列から表面テクスチャと映り込みを区別して物体の3次元形状を復元する問題に関して美しい解を与えた理論研究であるのに対して、後者は、見開きにした書籍をイメージスキャナで撮影した画像から書籍の3次元形状を求める問題を相互反射、近接光源、表面反射、不均一反射率分布、移動光源という複雑な条件下でのShape from Shadingの問題として定式化し問題の解法を示した応用研究である。T. Wadaらの受賞は日本人初のMarr Prizeというだけではなく、理論に偏り過ぎた過去のCV研究に対して理論と応用のバランスのとれた研究の重要性を示した点でも大きな意味があるといえるだろう。なお、Marr Prizeの最終選考に残った論文としてY. Xiong and S. Shafer (CMU)の“Hypergeometric filters for optical flow and affine matching”、A. Heyden (Lund大)の“Reconstruction from image sequences by means of relative depths”、P. Viola and W. Wells III (MIT)の“Alignment by maximization of mutual information”の3編にHonorable Mentionが贈られた。

今回の会議の特徴としては、物理ベースビジョンと3次元幾何学に関する研究が着実に進んでいる一方で、顔やジェスチャーの認識という応用指向の研究と学習をCVに適用する研究が多くなったことがあげら

表1: セッションおよびワークショップ一覧

プレナリーセッション		
S1:	IU Environment, Recognition	4件
S2:	Calibration, Navigation	5件
S3:	Shape Recovery (1)	3件
S4:	Shape Recovery (2)	3件
S5:	Face and Gesture Recognition	5件
S6:	Curve Matching, Shape Completion	5件
S7:	Pose and Correspondence	3件
S8:	Deformable Models	3件
S9:	Color, Texture, Specularities	5件
S10:	Motion (1)	5件
S11:	Learning, Modeling	3件
S12:	Representation, Geometry	3件
S13:	Motion (2)	5件
S14:	Stereo, Robot Vision	5件
ポスターセッション		
P1:	Stereo, Texture, Low Level Vision, Color, Calibration, Motion	36件
P2:	Recognition, Application	35件
P3:	Deformable Models, Surfaces, Learning, Geometry, Modeling, Analysis, Sensors, Active Vision	34件
ワークショップ		
W1:	Context-Based Vision	
W2:	Physics-Based Modeling in Computer Vision	
W3:	Visual Information Management	
W4:	Representations of Visual Scenes	
W5:	Shape and Pattern Matching in Computational Biology	

れる。全体の印象としては、ノイズや環境変動に対する処理の安定性、実現可能性、応用をする際の有効性を踏まえた研究が相対的に多く、また議論もこの点に集中するケースが多かった。個別の研究テーマの傾向としては、Snakes などの変形モデルの拡張に関する研究が盛んに行われている、Photometryに関する研究では鏡面反射が生じている物体の形状復元の研究が行われている、幾何学的解析では床面の平坦さに関する仮定を導入し平坦ではない部分(障害物)を検出する研究が行われている、Appearance-Based Vision は一つの研究の流れとして定着している、などがあげられる。(以上、和田、志沢、浅田)

2 プレナリーセッション

S1: IU Environment, Recognition

C.Kohl(Amerinex)らは、CVの標準的なプラットフォームとして米国を中心に開発が進められているImage Understanding Environment (IUE)と既存の画像理解ソフトウェアとの整合性を保つための研究について述べ、具体的にはKhorosのVisual Programming LanguageであるCantataを利用して、IUEプロセスとKhorosのプロセスを相互に接続して画像の解析を行うという例を示した。

次の2件は2Dあるいは3Dモデルと画像との対応付けに関する研究である。R.Basri(Weizmann)らは複数の領域から構成される対象のモデルと画像の領域を対応付けるアルゴリズムを示した。この手法では、対象を構成する複数の凸領域(凹領域の場合はその凸包)によってモデルを記述し、モデルの領域が画像から抽出された領域に含まれるようにモデルと画像間の幾何学的変換を求めている。一方、P.Viola(MIT)らは物体表面の光学的特性を用いることなく、輝度情報をもとにして物体の幾何モデルと画像の対応付けを行う手法を示した。この手法では、モデルと画像の相互情報を定義し、それを最大化するモデルの配置を求めている。

最後のR.Rao(Rochester)らの研究は、画像のインデキシングに関するものである。スケールを変えたガウシアン関数をフィルタ関数として用い、これらのフィルタ出力から回転とスケールの変化に不変な局所特徴ベクトルを構成し、それをインデックスとして用いている。異なる視点間のインデックスを補間するために、Kanervaによって提案されたSparse Distributed Memory(SDM)をもとにしたメモリを提案し、これにインデックスを格納するという点が特徴である。(以上、和田)

S2: Calibration, Navigation

M. Li(KTH)らの"Head-Eye Calibration"はKTHで作成されているアクティブビジョン実験システムのキャリブレーションの問題について述べたものである。D. Stevenson(Iowa)らの"Robot Aerobics: Four Easy Steps to a More Flexible Calibration"はカメラ運動を利用したカメラキャリブレーションについて述べたものである。カメラの運動、特徴点の位置の予測、カメラパラメータの修正を繰り返すことでキャリブレーションを行う方法で、エアロビクスという言葉はキャリブレーションの過程で同じカメラ運動を繰り返すことから来ている。この方法の問題点はレンズ歪みの扱いが難しいことである。これに対しては、歪みが軸対象であることだけを仮定し、カメラを少しずつ回転させ各位置で特徴点(最初は光軸上に配置しておく)の動きを求めて歪みを直接評価することで対処している。

ポスターセッションのG. Stein(MIT)の"Accurate Internal Camera Calibration Using Rotation, with Analysis of Sources of Error"も同じ考え方に基づくものである。この場合はカメラの動きを回転運動に限定し、内部パラメータのキャリブレーションだけを行なうように単純化している。しかしレンズ歪みの扱いが若干不十分なような印象を受けた。同じくポスターセッションのM. Qui(中国)らの"The Nonparametric Approach for Camera Calibration"は逆にレンズ歪みだけを扱ったものである。レンズ歪みは、通常は歪みの関数形があらかじめ分かっていると仮定してパラメトリックに評価される。M. QuiらはKohonenアルゴリズムという学習アルゴリズムの一種を用いて、レンズ歪みをノンパラメトリックに評価する方法を提案している。(以上、藤原)

S3: Shape Recovery (1)

Marr Prizeを受賞したT. Wada(岡山大)らの研究発表はこのセッションで行われた。複雑な条件下でのShape from Shading問題の解法を示した点に加えてCV研究が実際的な問題に役立つことを明確に示した点が高く評価されたものと思われる。

他の2編の論文は、画像から物体の密な形状情報を記述するグラフィックスモデルの構築であり、両方とも物体の回転画像を利用したものである。物体の反射特性に関しては、それぞれ鏡面反射と乱反射を主に利用している。J.Y. Zheng(九工大)らの"Shape and model from specular motion"は鏡面反射の物体表面における視覚特徴の画面移動に注目し、その移動の形式と物体の表面形状との関係を明らかにした。単一または複数の光源の下で、簡単な直接計算によって鏡面反射の物体形状を推定し、物体のグラフィックスモデルを作成できることを示した。これによって、現在広く使われているレーザー投光で測定不可能の強い鏡面反射を持つ物体の計測が可能となる。J. Lu(UBC)らの"Reflectance function estimation and shape recovery from image sequence of a rotating object"は同軸光源による物体の濃淡情報とシルエットの境界情報を組み合わせて、物体表面の幾つかの特異点の3次元位置を求め、これらの特異点から物体形状を近似的に計算する方法を提案した。物体の反射率が均一であるという条件の下で反射率の推定も行っている。(以上、鄭)

S4: Shape Recovery (2)

このセッションの発表は主にShape from Motionの研究であり、エッジ点の3次元での位置推定を目的としている。その中でS. Seitz(Wisconsin)らの"Complete scene structure from four point correspondences"は空間中非同一平面での4点を異なる2枚の画像での投影で対応付けたことによって、画像間の運動をaffine変換でまとめ、エッジ線を定め、

特徴点のマッチングを安定にする方法を示した。そして3次元位置の最適計算も施している。一般的に画像間各特徴の対応付けは、対象特徴によって安定にできるものとできないものがある。安定に対応できるものを用いて、不安定な特徴の対応に情報を与える発想は従来の無差別の対応より良い結果をもたらすと思われる。(以上、鄭)

S5: Face and Gesture Recognition

K. Matsuno, C.W. Lee, S. Tsuji (イメージ情報科学研究所)らはポテンシャルネットを用いた表情認識に関する研究を発表した。この内容は、ECCV'94およびICPR'94で発表した内容を探索のところまで拡張したものである。顔表情を1枚の静止画像から認識することができるアルゴリズムで、従来生理学で研究された網膜上のポテンシャル場をその背景にしている。

A. Essa, A. Pentland (MIT)らは顔の物理モデルを用いた表情認識に関する研究を発表した。認識方法はCVPR'94で発表した有限要素法による認識方法をベースにしており、今回はその認識結果を時空間のパターンとして認識率を上げたことであり、アルゴリズムの大きな変化は見られなかった。まず eigen face 方法によって目尻、口、鼻先などを探して顔の wire-frame モデルを正規化し、画像列から求めた optical flow 情報を加えて有限要素法で顔の各筋肉の動きを推定し、(1)特徴ベクトルのマッチングと、(2)optical flow のテンプレートマッチングの二つの方法によって認識を行うものである。発表翌日の Media lab の公開時にはビデオによるデモを見せていた。

A. Lanitis (Manchester)らの研究は、顔画像を形状と濃淡値に分けてあるパラメータ空間に記述し、顔の appearance を総合的に認識しようとするものである。入力画像が与えられると、まず ASM というアルゴリズムより特徴点を探し、その結果を幾つかのモデル画像より求めたパラメータで表現し、特徴点付近の濃淡情報を画像ワーピングアルゴリズムによって正規化された形状 (shape-free patch) に直して認識に使っている。これは各個人の個性、ポーズ、表情、照明などによる画像記述の乱れを一つの枠組みのなかで解決しようとする試みである。一つ一つの課題に対して明確な公式化が行われていないという問題はありますが、総合的に顔問題を形状と濃淡値を融合して解決するという点は評価すべきである。

M.J. Black (Xerox)らの研究は、CVPR'94で発表した内容を拡張したものである。彼らの従来のアルゴリズムは NTT の間瀬らが行った optical flow による表情認識を FACS によって解析したものだったが、今回はその情報を2次元 flow を使った簡単なパラメータ空間に表現し、optical flow 情報よりそのパラメータを推定し、その結果によって入力画像から特定部分(4点で定められる目、口、眉など)をテンプレート

として登録し、テンプレートマッチングで追跡を行う方法を示した。特徴は、頭の大きな運動に対してもトラッキングが安定することであるが、テンプレートを更新しながら探索を行う場合、累積されるエラーをどう処理するかに関してさらに工夫が必要と思われる。

A.F. Bobick, A.D. Wilson (MIT)らはジェスチャ認識に関する研究を発表した。ジェスチャを状態空間において連続した状態の集合であると定義し、その状態遷移の中での反復性と変化ルールを見つけることが認識であると考えている。通常、入力データからの有効な状態を決めるのは難しいが、この研究ではメンバシップ関数を用いて推定している。認識には前向き DP アルゴリズムを使っており新しさはない。実験ではマウスに2次元位置データ、磁気センサによる手の3次元位置、方向データ、手の動画像を用いて認識を行っている。このような研究は最近日本でも良く見られることであり、状態遷移に関しては HMM アルゴリズムを用いた研究が既に報告されている。ファジイ理論を用いたのは目新しいがその効果はこれから評価すべきであろう。

顔、手、人体の認識に関しては、ポスターセッションの Recognition, Applications にも幾つかの研究が発表され、人間の表情や行動の情報抽出の研究が注目を集めていることは確かである。また、このテーマに関しては最近リアルタイムで認識を行う研究も発表されるようになっており、今後の発展がさらに期待される。(以上、李)

S6: Curve Matching, Shape Completion

E. Pauwels (Leuven)らの研究は、Geometry-Driven Diffusion を用いて閉じた軌跡の特徴を強調する方法について述べたものである。Geometry-Driven Diffusion はもともとエッジを保持したままスムージングを行うために提案された方法である。全く同じ考えを閉じた軌跡に適用し、重要な特徴(エッジ)を保持したまま小特徴の除去(スムージング)を実現している。この研究では Diffusion 自体も関数最適化で実現できるように工夫を施している。

S. Zhu (Harvard)らの研究は、Snakes、統計的類似度テストにより領域を併合する Region Growing、Bayes/MDL、などの各種のセグメンテーション手法を、統計的評価量を変分原理を用いて最小化を行うという観点から統一的に眺めることを試みたものである。Bayes を用いたセグメンテーションと違って MDL を用いたセグメンテーションというのは、まだそれほど一般的ではないように思われるが、各種のセグメンテーション手法を数学的に統一した見地から眺めようとする試みは注目に値する。(以上、藤原)

S7: Pose and Correspondence

3D → 2D の rigid な対象の対応付け、2D → 2D の non-rigid な対象の対応付け、2D → 2D の領域ベース

の対応付けに関する研究の発表が行われた。

R. Horaud (INRIA) らは、weak perspective モデルを用いて、3次元物体と2次元の点集合との対応付けを反復計算によって行い、極限的には perspective モデルの下での対応付けが求められるという Dementhon と Davis によって提案されたアルゴリズムを、paraperspective モデルを用いるように拡張した。

H. Tagare (Yale) らは、閉曲線の形状を基にした Non-rigid な対応付けを行う際の基準に関する研究を発表した。従来の曲率を基準とした対応付けでは、(1) 曲率は rigid な輪郭に関する不変量であって、non-rigid な輪郭の対応付けに用いるには問題がある、(2) 二つの輪郭 C_1 と C_2 の対応付けを行う場合に、 $C_1 \rightarrow C_2$ の対応付けと、 $C_2 \rightarrow C_3$ の対応付けの結果が等しくなるという保証はない、という問題がある。これらの問題を解決するために、新たな評価関数を定義し、これを最小化することによって(1)(2)の問題を解決した対応付けアルゴリズムを提案した。

C. Wang, K. Abe (静岡大) らは、領域隣接グラフを用いて異なる視点から撮影した画像の領域を対応付ける手法に関する研究を発表した。領域隣接グラフは視点の変化に対してその構造が変化する。このような構造の変化を許容し、かつ属性値の類似性が最大となるような対応付けを求めることがこの研究の目的である。領域隣接グラフを2連結成分に分解することによって、葉の部分と2連結成分となる木構造を得ることができる。この木構造と2連結成分は、それぞれ領域の包含関係と隣接関係を表している。このような表現を用いて、2つのグラフの2連結成分の対応付けから開始してグラフ全体のコストが準最適となるような対応付けを求めるアルゴリズムを提案している。(以上、和田)

S8: Deformable Models

A. Giachetti, V. Torre (Genova) らは、Deformable Object (例えば人工的に合成された雲、黒い色素を含んだ水)の動きを線形的に近似することによって動きを解析するアルゴリズムを提案した。これは平面(リジッドな)の動きから求めた optical flow が一つの singular point をもつこと、deformable な動きのフローは3つの singular point をもつことに着目し、それを基準にして Jacobian を求め、線形的に解釈する方法である。幾つかの実験結果に関して示したが、結果の評価に関して不明確なところが多く実世界のオブジェクトに対してどれくらい安定するかの疑問が残った。

J. Park, D. Metaxas, L. Axel (Pennsylvania) らは、3次元 Deformable Model を用いた左心室の動きを認識する研究を発表した。この研究は Terzopoulos の 3D symmetric model と似ているところが多い。発表内容は二つに分けられる。一つは6つのパラメータ関数による3次元左心室モデルの作成に関するも

のである。このモデルは3次元のパッチによって作られるが、パラメータを制御することによって形状を変えることができる。もう一つは画像情報からモデルを変形する力を求めるアルゴリズムである。ここでは SPAMM (SPATIAL Modulation of Magnetization) のデータ(横と上から撮影して得られるような2枚の画像)を用いており、入力画像が与えられると2次元 Snake によって初期の形状を求めた後、提案するアルゴリズムによってモデルを変形していく。変形が完了して動きがとまるとモデルからパラメータを推定しそれを分析することによって認識を行っている。6つのパラメータが持つ意味(定性的な表現を含めて)は医学的には大変重要であるらしい。

V. Caselles (Genova) らは、オブジェクトの特徴情報を有効に取り出す Active Contour モデルに関して発表した。一般的に Active Contour モデル (Snake) は境界線のヒューリスティックな情報、つまりある形状を表すために重要と思われる特徴点を意識せずにセグメンテーションを行うが、彼らは形状を表すための最短距離を制約として用いることによってさらに有効と思われるアルゴリズムを提案している。そのため Snake の表現を Riemannian Space での最短曲線表現に直してエネルギーの最小化を行っている。この手法は従来の2次元形状の最適表現と Snake アルゴリズムを結合したものといえる。(以上、李)

S9: Color, Texture, Specularities

G. Healy (UC Irvine) らは、色に関して今回合計4つもの論文を発表をしている。ポスターセッション P1 で発表した "The Illumination-Invariant Recognition of Color Texture" は Color Texture 特徴を特異値分解を使ってきれいに処理している点が興味深く参考になった。一方、このセッションで発表した D. Panjwani, G. Healy の "Results using Random Field Models for the Segmentation of Color Images of Natural Scenes" は、Color image segmentation を GMRF を用いて行う手法を扱っているが、CVPR'93での同様の発表との違いが少なく新規性に乏しい。質問として境界がなめらかでないというものが出たがこれは Texture の特徴を推定する Window の大きさに起因している。

Texture の発表では"きめの勾配"を扱った J. Gårding (KTH) の研究がレベルが高いと感じた。アイデアは ECCV'94 において Malik & Rosenholtz が提案した隣接する Texture 間のアフィン変換を求めて勾配を推定する方法を基本として使っている。Malik らの方法が表面パラメータの探索空間が5次元であるのに対して、彼らが提案する枠組では1次元と、大幅に小さくなっているのが特長である。Gårding は ECCV'94 の発表において同じ ECCV'94 に発表された Malik & Rosenholtz の論文をリファアー(つまり未発表論文のリファアー)しており、両者は Texture 研究の最先端でし

のぎを削っている。

Specularities の発表は、詳細な物理モデルを導入して鏡面反射問題を解いた M. Oren, S. Nayar (Columbia) の “A Theory of Specular Surface Geometry” 1 件であり、この発表は Marr Prize を受賞した。(以上、内山)

S10: Motion (1)

Motion のセッションで印象に残った発表は以下の 2 編である。

一つは、Y. Xiong, S. Shafer (CMU) の “Hypergeometric Filters for Optical Flow and Affine Matching” で、マッチングのための新しいフィルタの提案をしている。位置と周波数の不確定性のことまで考慮に入れて議論をしており、緻密な研究という印象を受けた。もちろんこの H-filter 自体 Gabor 関数に近く、再帰的に定義できることなどの良い性質を持っており、他の問題にも適用できそうである。一方、S. Ayer (Swiss FIT), H.S. Sawhney (IBM) らは、動画像をいくつかのレイヤーに分解する研究について述べている。しかし近景に隠された遠景に何があるのかは推定しておらず、レイヤー同士の前後関係(近景と遠景)を考慮した分離ではないのが残念である。(以上、内山)

S11: Learning, Modeling

3 件の発表のうち 2 件は CV のための学習手法 (visual learning) に関するものであり、残りの 1 件は人工生命への CV の応用、あるいは CV への人工生命の応用である。

B. Moghaddam, A. Pentland (MIT) は、物体検出のための教師なし学習方法 (確率密度推定法) を提案した。これは固有値分解を用いた主成分分析と密度推定のための EM アルゴリズムを統合し、多モード分布に対応したものである。このアルゴリズムを画像データベースからの多スケールの顔検出と手話理解へ応用した。手話理解では、分布が多モード(つまり多義解釈)になるため、混合ガウス分布で多モードを表現している。

S. Mukherjee, S. Nayar (Columbia) は、visual learning のための関数近似手法である GRBF を、wavelet 変換を用いて自動的に生成する方法を提案した。従来の GRBF の学習法で ad hoc に決められていた基底関数の数と係数を、wavelet 展開の係数から与えられた近似精度を満たすように決定する公式を導いた。この公式を用いて 3 次元物体の認識と姿勢推定を学習する実験を報告した。

D. Terzopoulos, T. Rabie (Toronto) は、人工生命 (animat) の魚に CV の技術を用いてアクティブビジョンの視覚機能を模倣的に実現し、それを用いて人工生命のグラフィクスシミュレーションを行うシステムのデモを示した。このシステムは CV の立場から見

るとアクティブビジョンのシミュレーション実験と見ることができ、CV の新しい応用としての価値があると思われる。

これら 3 本の論文は、いずれも CV というよりもパターン認識、ニューラルネット、人工生命の学会で発表されてもおかしくない内容である。彼らは CV という枠に縛られることなく、CV は自分らの研究の単なる発表の場の一つと見なしている。

ビジョンにおける学習の研究動向は、パターン認識的な手法やニューラルネットの応用が中心であるが、計算論的な研究と学習理論的な研究をバランスよく進めて最終的には両者を統合することが今後の CV 研究において重要になるであろう。(以上、志沢)

S12: Representations, Geometry

形状表現に関して 1 件、Fundamental Matrix(F) の計算法に関して 2 件の発表があった。

C. Dorai, A.K. Jain (Michigan State) は、自由曲面を持つ物体の新たな表現法を提案した。2 つの主曲率から形状パラメータを定義し、このパラメータが変化しない限り最大の面パッチ (Constant-Shape Maximal Patch: CSMP) の集合として効率的に自由曲面を表現することができる。さらにパッチ間の隣接関係を保持することにより、non-convex の物体も扱うことが可能である。

B. Boufama, R. Mohr (LIFIA-INRIA) は特徴点の対応の組から F を直接計算するのではなく、新たなパラメータ表現のもとで 1 つの epipole と 2 自由度の homography を求める問題に変換することにより、効率的にしかも安定に F が求められることを示している。

また P.H.S. Torr, A. Zisserman (Oxford) らは、 F を求める際に、複数の動きモデルを並列にロバスト当てはめアルゴリズムを用いて検討することにより、データ中に outlier が存在しても複数解の存在する状態 (degeneracy) を識別する手法を提案している。(以上、三浦)

S13: Motion (2)

Motion に関する 5 件の発表があった。2 次元運動の解析に関する S.A. Niyogi (MIT) の研究は、時空間 Junction 解析を用いて運動によるオクルージョン (Kinetic occlusion) の検出を行う方法を提案した。本方法では、オクルージョンの物理モデルを用いて、時空間可操蛇フィルタの出力から時空間 junction を検出し occlusion と disocclusion を同定しているのが特長である。

G. Chou (MIT) は、Kinetic occlusion から図地分離を行う運動検出モデルを提案した。運動境界の運動とパターンの運動が一致する時に、その運動を図の運動と解釈する。この判定を Bayes 則を用いて行った。

A. Heyden(Lund) は、カメラキャリブレーションを行わずに、 m 枚の画像から空間中の n 点の配置を復元する問題を扱った。この問題を解くために、4 個のパラメータで決まる reduced fundamental matrix を導入した。

R.I. Hartley (GE) は、2 枚の画像間の点対応から運動パラメータと構造パラメータを復元する Longuet-Higgins の 8 点線形アルゴリズムを改良し、ノイズに強いアルゴリズムを提案した。8 点線形アルゴリズムはノイズに弱いとされ、種々の最適化手法や非線形手法が提案されてきた。この研究は、簡単な正規化によってノイズに対する耐性が著しく向上することを示している。

J. Costeira, T. Kanade (CMU) は、多フレームからの 3 次元運動と構造推定のための因子分解法を多物体の複数運動へ拡張した。新たに、shape interaction matrix を導入し、そのブロック対角化操作によって、複数物体の複数運動の推定が可能であることを示している。従来、T. Boult と L. Brown によって提案されていた複数物体に対する因子分解法においては、セグメンテーションのために多数回因子分解を実行しなければならなかった。提案方法では、因子分解のための特異値分解は、一度だけの実行で十分である。しかし、ブロック対角化のための計算量がどの程度必要なのかを今後評価する必要がある。(以上、志沢)

S14: Stereo, Robot Vision

ステレオ関連で 2 件、ロボットビジョンで 3 件の発表があった。なお、後者ではいずれも実際のロボットやカメラヘッドを動かしているビデオが紹介された。

P. Fua(SRI) は、複数視点から得られるステレオデータを統合して、物体の 3 次元形状復元を行う方法を示した。まず物体表面を小さな円盤状の面素の集合として表し、得られた複数のステレオデータから面素集合を初期化し、さらに物体の滑らかさの拘束および 3 次元データと各ステレオ画像対との整合性を評価する関数を定義して、それを最小化するように各面素の位置を調節する。簡単な距離に基づいたクラスタリングで複数物体が分離できることを示している。

D. Bhat, S. Nayar (Columbia) は、鏡面反射成分が存在するときのステレオ対応付けを扱っている。対応点の濃度値の差がある範囲内に収まるという条件のもとで、できるだけ depth の解像度を高めるようなカメラの最適な配置を求めている。求められた最適配置は法線と照明の方向に依存しない。

S.B. Kang(DEC), K. Ikeuchi(CMU) は、人間の物体を組み立てる動作を観察、理解し、同様の作業を行うロボットプログラムを生成するシステム (Assembly Plan from Observation システム) を構築した。物体と手の位置は実時間ステレオで、指の姿勢はデータグローブで計測する。得られた時系列データを自動的に

分節した結果から、作業の実現に必要な物体の動きおよび手の把持姿勢を抽出する。抽象化されたレベルでそれらの情報を表現しているため、さまざまな種類のロボットハンドへマッピングして実際のプログラムを生成できる。システムとしての完成度は高い。

S.M. Fairley, I.D. Reid(Oxford) らは、ステレオ動画像から移動物体を抽出し、アクティブカメラで追跡するシステムを構築した。前回の ICCV で提案した affine transfer に基づく注視点の移動アルゴリズムを拡張し、得られたすべての特徴点を用いることにより信頼性を高め、さらに注視点の移動には structure 情報が不要であることから、特徴点の数にかかわらず常に一定の処理時間で注視点制御に必要な motion 情報のみを得るアルゴリズムを考案した。これは実時間処理の観点から有用である。さらにステレオ動画像にこのアルゴリズムを適用して、2 フレームで必要な計算が行えることを示した。Transputer ベースのシステムを構築し良好な追跡を行っていた。

J. Miura(大阪大), K. Ikeuchi(CMU) は、組立作業をガイドするための視覚認識戦略を、作業の目的を考慮して自動的に生成する手法を提案した。従来の視覚プランニングの研究では、視覚特徴の候補があらかじめ与えられている場合がほとんどだが、ここでは作業の記述から候補を自動生成している。また、可能な視覚認識戦略の中から最適なものを選ぶ際の指標として、ある認識戦略の作業の遂行に対する寄与度、具体的にはある認識戦略をとった時の予測される作業の成功確率を用いることを提案した。ロボットビジョンでは、視覚情報の価値はそれ単独(例えば計測精度だけ)では評価できないことが多いため、このような観点から指標を設定することは重要である。(以上、三浦)

3 ポスターセッション

ポスターセッションは最終日を除いて毎日 1 回、合計 3 回行われた。ポスター会場は非常に混雑したが、連日 2 時間の熱心な議論が続けられた。上記の報告中でも関連のあるポスター論文について触れているが、その他に 3 回のポスターセッションを通じて印象に残った論文を以下に紹介する。

S. Nayar (Columbia) らの "Real-time Focus Range Sensor" は、フォーカスを変化させて得た 2 枚の画像からビデオレートで距離画像を得るシステムで、距離精度も RMS 誤差 0.34 % とかなり完成度の高いものである。フォーカスを利用した距離計測法は texture の無い面の距離計測を行うことができないが、この問題に対してパターン光を照射することで対処している。エッジやノイズの処理が明らかではないが、通常の画像センサの空間分解能の距離画像を実時間で得るシステムを実現したことは高く評価できる。

N. Asada (岡山大) らは、従来のたたみ込み積分に基づくボケモデルに対して新たに逆投影に基づくボケ

モデルを提案し、occluding edge のボケ現象を詳細に解析して実験的に検証した。occluding edge のボケ解析については、T. Nguyen, T. Huang らがたたみ込み積分を多重化する方法を ECCV'92 で提案しているが、この研究ではたたみ込み積分に基づくボケモデルの限界を明らかにし、より一般的なボケモデルの導出を行っている。

S. Smith (Defence Research Agency, UK) は、動画像で特徴点の追跡結果から得られるオブティカルフローを分割して運動物体領域を抽出し、さらにエッジ情報を加えてより精密な領域を得る手法を提案し、Transputer ベースの実時間システムを構築した。自動車や飛行機などをかなり安定に追跡していたが、その能力は特徴点の追跡能力に依存する。

Z. Zhang (INRIA) は、2 視点から得られたシーン中の直線セグメントの対応を利用し、それらが 3 次元空間内で overlap する位置にあるべきだという拘束を利用して形状及び動きを推定する手法を提案した。2 画面間の対応から解が求められることが特徴である。

H.-Y. Shum, M. Herbet (CMU) らは、自由曲面を持つ物体の多視点から得られた距離画像からの再構成を扱っている。各視点での距離画像から得られた球面表現間の対応を resampling によって行い、さらに統計的最適化手法を用いることにより精度よく自由曲面を表現する手法を提案している。

この他に、カメラの観測方向が変化している時の連続のビデオ画像系列の中で、共通視野を持つ画面を重ね、全体視野を作り出す Mosaic 画像の研究発表が 2 件ほどあった。その目的は、画像データの圧縮やビデオの目次を作ることであり、マルチメディアにおけるビジョンの応用の一つと考えられる。オブティカルフローや affine 変換を計算することによって実現できる。(以上、藤原、三浦、鄭、浅田)

4 ワークショップ

ICCV'95 では会議の前に 4 件、後に 1 件の合計 5 件のワークショップが行われた。以下では報告者が参加した 2 件について概要を述べる。

W1: Context-based vision

私が期待したのは、Context を与えることで何かの推定や処理がうまくいくことであった。しかし内容は期待外れであった。Context を使うというより、限定された世界でなら Computer Vision はうまく行くという話が多かったと思う。例えば、最初から入力される画像がスーパーマーケットの棚に限定されるといふ意味でコンテキストを使っていると主張していた。これに対してはマイクロマウスにこんなことができるかという冗談めいた話がでていた。(以上、内山)

W2: Physics-Based Modelling in Computer Vision

実際的な研究が重視された今回の ICCV でも、特に顕著な特徴は Physics-Based Vision に関する研究が高く評価されたことであろう。それは今回の Marr Prize が Physics-Based Vision の基礎と応用の 2 件に与えられたことから分かる。Physics-Based Vision に対しては、実際問題に適用しようとするパラメータの数が多くなり過ぎるなどの批判はあるものの、ICCV 直前の 2 日間で行われたこのワークショップは盛況であった。

ワークショップ中驚いたのは、パネル討論中に Physics-Based Vision に関する悲観的な意見が聞かれることが多いことであった。しかしそれに対して K. Ikeuchi (CMU) は「Physics-Based Vision に対する悲観的な意見が多く聞かれるが、これは定義の違いから来るものであろう。私自身は Physics-Based Vision の意味を広くとらえている。パラメータの数が多くなり過ぎるという問題に関しては、全世界 (entire world) を完全にモデル化するということがもともと無理なのであり、どのような問題に適用するのかを明確にする必要があるだろう。そうすれば Physics-Based Vision は実際問題にも適用できるはずであると信じている。私自身は今 Physics-Based Vision と Physics-Based Graphics のクローズドループによって何か面白いことができなかと考えているところだ。」と述べていたのが印象的であった。(以上、藤原)

5 むすび

本報告は、ICCV'95 の概要を有志によってまとめたものである。会議全体の様子と最新の研究動向をできるだけ正確に報告できるように、執筆者で分担して 14 のプレナリーセッション全部について概要をまとめた。そのためにポスターセッションとワークショップについては十分な紙面を割くことができなかった。個々の論文に興味を持ち、詳細な内容を知りたい方は直接 proceedings[1] を参照していただくか、本報告の執筆者に問い合わせをしていただければ幸いである。最後に、本報告の機会を与えていただいた CV 研究会主査の京都大学 松山隆司 教授に感謝する。(以上、浅田)

参考文献

- [1] Proceedings of the Fifth International Conference on Computer Vision, IEEE Computer Society Press, June, 1995.