

第10回コンピュータビジョン国際会議 ICCV2005 報告

古川 亮 (広島市立大学)

川崎 洋 (埼玉大学)

宮崎 大輔 (東京大学)

佐藤 智和 (奈良先端科学技術大学院大学)

内海 章 (ATR)

佐藤 洋一 (東京大学)

加藤 丈和 (和歌山大学)

三田 雄志 (東芝)

中澤 篤志 (大阪大学)

菅谷 保之 (岡山大学)

杉本 晃宏 (国立情報学研究所)

和田 俊和 (和歌山大学)

ryo-f@cs.hiroshima-cu.ac.jp

kawasaki@mm.ics.saitama-u.ac.jp

miyazaki@cvt.iis.u-tokyo.ac.jp

tomoka-s@is.naist.jp

utsumi@atr.jp

ysato@iis.u-tokyo.ac.jp

t.kato@ieee.org

takeshi.mita@toshiba.co.jp

nakazawa@ime.cmc.osaka-u.ac.jp

sugaya@suri.it.okayama-u.ac.jp

sugimoto@nii.ac.jp

twada@ieee.org

あらまし: 2005年10月17日~20日に北京(中国)で開催された第10回コンピュータビジョン国際会議の概要を報告する.

Report on the 10th IEEE International Conference on Computer Vision (ICCV2005)

Ryo Furukawa (Hiroshima City Univ.)

Hiroshi Kawasaki (Saitama Univ.)

Daisuke Miyazaki (Univ. of Tokyo)

Tomokazu Sato (NAIST)

Akira Utsumi (ATR)

Yoichi Sato (Univ. of Tokyo)

Takekazu Kato (Wakayama Univ.)

Takeshi Mita (Toshiba Corp.)

Atsushi Nakazawa (Osaka Univ.)

Yasuyuki Sugaya (Okayama Univ.)

Akihiro Sugimoto (NII)

Toshikazu Wada (Wakayama Univ.)

Abstract: This report gives an overview of the 10th IEEE international conference on computer vision (ICCV2005), which was held in Beijing, China, from October 17th to 20th, 2005.

1 はじめに

2005年10月17日から20日までの4日間にわたり、第10回コンピュータビジョン国際会議 (ICCV2005) が北京(中国)で開催された。ICCVは、IEEE Computer

SocietyのPAMI-TCが母体となり、ヨーロッパ圏、アメリカ圏、アジア圏の順に場所を移して西暦奇数年に開催される最大規模の国際会議である。当初、2003年に北京開催、2005年にニース開催と計画されていたが、2003

年に中国で大流行したサーズの影響で、2003年と2005年とで開催地を交換した経緯がある。ICCV2005では、Songde Ma(中国科学院)とHeung-Yeung Shum(マイクロソフトアジア)がGeneral Chairsを務め、Bill Freeman(MIT), Luc Van Gool(Univ. of Leuven), Subhasis Chaudhuri(インド工科大)がProgram Chairsを務めた。会議には、34ヶ国から831人(うち日本からは83人)が参加した。

投稿論文は1230件とICCV2003の966件を大きく上回り、過去最多となった。なお、10年前にボストンで開催されたICCV1995と比較すると、投稿論文数は倍増している。ICCVは、伝統的に、質の高い論文を厳選し、オーラル発表はシングルトラックで行うというスタンスをとっており、今回もオーラル発表45件、ポスター発表199件の計244件が採択され、採択率は19.8%であった。今回は、投稿論文数の増加を踏まえ、ポスター発表での採択を過去の開催に比べ50件ほど多くしたとのことである。第1著者の所属に基づいて、国別の発表件数を整理した結果を表1に示す。また、オーラルセッション、ポスターセッションの構成と発表件数を表2に示す¹。トラッキングや認識のための学習といったトピックが最近の流行となっているが、今回のオーラルセッションは、やや古典的な内容で構成されているという印象が残った。各セッションの詳細は次節以下を参照して頂きたい。

ICCVでは最優秀論文にMarr Prizeが授与される。今回は、Fredrik KahlとDidier Henrionによる“Globally Optimal Estimates for Geometric Reconstruction Problems”に贈られた。これは、3次元復元等の幾何計算に現れる最適化問題を半正定値計画問題として定式化し、緩和を利用することで大域的最適解を求める手法を提案している。半正定値計画緩和手法は数理計画法の分野で近年急速に進展しており、その成果をCVにうまく持ち込んだことが評価されたのではないかと思われる。なお、最終選考に残った論文として、Kiriakos N. KutulakosとEron Stegerによる“A Theory of Refractive and Specular Shape by Light-Path Triangulation”, Oren BoimanとMichal Iraniによる“Detecting Irregularities in Images and in Video”, Stefan RothとMichael J. Blackによる“On the Spatial Statistics of Optical Flow”の3編にHonorable Mentionが贈られた。

各ポスターセッションと平行してデモセッションが開

¹ハンドアウトのプログラム冊子と予稿集とで一部不整合があったが、実際の会議はプログラム冊子に基づいて進められたため、プログラム冊子をもとに分類した。

表 1: 国別発表件数一覧

国名	オーラル	ポスター	計
アメリカ	24	96	120
中国	1	24	25
イギリス	6	17	23
イスラエル	4	10	14
フランス	2	6	8
日本	0	8	8
カナダ	1	6	7
オーストラリア	3	2	5
スウェーデン	2	3	5
ドイツ	0	5	5
スイス	1	3	4
ギリシャ	0	4	4
韓国	0	2	2
台湾	0	2	2
トルコ	0	2	2
その他(計7ヶ国)	1	6	7
合計	45	199	244

催され、計18件の出展があった。また、本会議前後には、3日間にわたり、7件のショートコース、および、11件のサテライトワークショップが開催された。ベストショートコース賞が、Li Fei-Fei, Rob Fergus, Antonio Torralbaによる“Learning and Recognizing Object Categories”に贈られた。また、“Where Am I?”という名の位置同定をテーマにしたComputer Visionコンテストが企画されるなどICCVとしての新しい試みもあった。

(以上、杉本、佐藤(洋)、和田)

2 オーラルセッション

O1: Motion Segmentation

S. R. Rao(Illinois)らは、Hybrid Quadratic Surface Analysisを提案し、剛体オブジェクトを撮影した2枚の画像と、それぞれの画像上での対応点座標が複数与えられたとき、エピポーラ拘束を満たす基礎行列を精度良く推定する方法を与えている。従来、R. VidalらがGPCAを提案しているが、エピポーラ拘束は線形部分空間への分離問題に変換してから適用していた。しかし、本来2次形式の拘束式は2次曲面への当てはめを行う方が自然であり、例えば平面と2次曲面が混在する場合に適用できないという問題があった。提案手法によりGPCAを上回る性能が報告された。

J. Kang(Southern California)らは、移動するカメラによって撮影された映像から移動物体を検出・追跡する

表 2: セッション一覧

Oral Sessions		件数
O1:	Motion Segmentation	5
O2:	Feature Extraction and Image Structure	4
O3:	Tracking Deformable Structures	3
O4:	3D Acquisition	4
O5:	(Human) Pattern Detection	4
O6:	Segmentation	3
O7:	Pattern Detection and Grouping	4
O8:	3D Reconstruction	4
O9:	Color and Reflectance	3
O10:	Feature Matching	4
O11:	Tracking	4
O12:	Camera and Calibration	3
Poster Sessions		件数
P1:		42
P2:		59
P3:		57
P4:		41

問題を扱っている．従来よく用いられる方法では，シーンが平面に近いと仮定し，2D アフィン変換で 2 枚の画像を重ね合わせた上でマッチしない領域を動物体とする．しかし，奥行きのあるシーンでは parallax(視差)が生じるため，背景の立体物などを過剰に検出していた．提案手法では，4 枚のフレーム画像を用い，エピポーラ拘束および相対深さに基づく幾何拘束を導入することで問題を低減している．

O. Tuzel(Rutgers) らは，対応付けられた 3 次元空間内の 2 つの点の集合から，複数の異なった運動を求めるために，Lie 群上で Mean Shift を行う手法を提案している．Lie 群とは平行移動や回転といった運動を表現する集合である．単純に運動パラメータの空間内で Mean Shift を行うことはできないが，Lie 群を導入することでこの問題を解決している．

M. P. Kumar(Oxford Brookes) らは，人体のように異なった動きをする複数のセグメントからなる非剛体を記述するモデルを提案している．モデルは複数のレイヤーによって記述され，前のレイヤーに属するセグメントによって後ろのレイヤーのセグメントが遮蔽されるという場合を表現できる．歩行者を側面から撮影した動画シーケンスを入力すると，まず動きが似た領域を loopy belief propagation により分割する．これを初期モデルとして， $\alpha\beta$ -swap や α -expansion アルゴリズムによる形状の修正，セグメント単位での色の連続性評価，動きモデルの修正を繰り返しモデルを修正する．

O2: Feature Extraction and Image Structure

このセッションでは，画像の特徴抽出や構造についての解析や応用に関して 4 件の発表が行われた．

S.Roth(Brown 大) らは，オプティカルフローの空間的な統計量を学習する手法を提案している．空間的統計量は，(1) カメラモーション (2) シーンのデプス情報 (3) オブジェクトの個々の動き，により決まる．学習は，CVPR05 において著者らにより提案された Field-of-Experts モデルを事前確率とし，オプティカルフローの MRF モデルを機械学習することで行われる．学習により得られた事前確率を用いて，オプティカルフローの精度向上を行ったところ，従来の確率モデルを用いた手法に対し約 10%の精度向上が見られた．

T.Hazan(ヘブライ大) らは，画像のセットが与えられたときに，それらの画像に対する特徴抽出フィルタを得るための方法として，Non-negative Tensor Factorization(NTF) を利用することを提案している．Non-negative Matrix Factorization(NMF) と比較したときの NTF の利点としては，画像をラスタライズせずに済むために画素の隣接関係の情報が保存されること，唯一性があること，分解されたフィルタの局所性 (sparseness) が高いことが挙げられている．彼らは，NTF の計算アルゴリズムとして，因子が正を保つような最急降下法を提案し，その収束の証明を示している．

Y. Wang(UCLA) らは，知覚に基づくスケールスペースの概念およびそのアプリケーションを提案している．著者らは，画像を (i) オブジェクト境界の構造 (ii) テクスチャ情報，の 2 つに分け，D.Marr の primal sketch に基づいた sketch pyramid としてスケールスペースを表現した．知覚的なスケールスペースの利用例として，(1) スケールの異なるオブジェクトのトラッキング，(2) 知覚的な重要度に応じた画像解像度の決定，というアプリケーションを挙げ実装している．また，彼らは提案手法が超解像度化などにも応用可能であると述べている．

A.Davison(Imperial College, UK) の研究は，tracking などの問題において，観測対象の状態についてのそれまでの観測結果から，探索する特徴の決定や，探索範囲の決定を行う，guided search の手法に関するものである．こうした手法は，particle filter や，拡張 Kalman Filter などの形で提案されているが，彼は，この問題を，相互情報量を用いて定式化している．この方法では，観測によって得られる情報が，bit を単位とした情報量を用いて表される．彼は，このことにより，複数の異なる種類の特徴を用いる場合でもそれぞれの特徴の観測に

よる効果を比較可能であると述べている。

(以上, 川崎, 古川)

O3: Tracking Deformable Structures

このセッションでは3件の発表があり,それぞれターゲットが異なる追跡の発表となっていた。

1件目の発表は,非剛体のレンジデータのフレーム間の対応付けに関する研究である。Y.Wang(Stony Brook)らは,共著者の一人のS. Zhangらが開発した高速かつ高精度レンジファインダによって,ビデオレートで得られた高精度レンジデータを用いて,密な3次元点をフレーム間で対応付ける方法について発表した。Harmonic Mapと呼ぶ3次元の表面から二次元の円板へのマッピングを行うことで,3次元のレンジデータの対応付け問題を,二次元の画像間の対応付け問題に落して解いている。

2件目の発表は,多関節剛体モデルを画像に当て嵌め,単眼カメラを用いて人物の姿勢を追跡する研究である。R. Urtasun(EPFL)らは,多関節モデルの姿勢に対する統計モデルを使って人物の姿勢を追跡した。あらかじめ学習した統計モデルを使って,人物の姿勢の事前分布と推定しておくことで,本来困難な単眼の画像への多関節モデルの当て嵌めを安定に行っている。また,高次元のパラメータ空間を,カーネル関数を使って低次元の空間に埋め込むことで,小数のトレーニングデータからの学習を実現している。

3件目の発表は,流体のベクトル場を時系列フィルタによって推定する研究である。A. Cuzolらは,高次元のベクトル場を,うず状の成分と,収束発散の成分に分解することで,低次元のパラメータで表現し,パーティクルフィルターによって追跡した。(以上,加藤)

O4: 3D Acquisition

ステレオ視に関して4件の発表があった。

Ferisらは,カメラの周囲に4つのフラッシュを取り付け,それぞれを点灯したときの影から,オクルージョン境界を検出する装置を開発した。それを用いて,オクルージョン境界以外のデプスが滑らかになるような制約を付け,ステレオ視によりデプスを推定した。

Narasimhanらは,霧のある状況下でのスリット光投影法について発表した。霧の中で対象物体に向けてスリット光を照射したとき,一部の光は物体に到達する前に散乱して観測されてしまう。しかし,到達地点より先は光は観測されないため,物体の位置を検出でき,デプスが推定できる。また,霧の中でも5つの光源下での画像から照度差ステレオ法を用いて散乱パラメータとアルベドと法線が求まることも示した。

ステレオ視において大域解を求める問題はNP完全である。Meltzerらは,大域解を多項式時間で求める手法を発表した。ただし,大域解を求めることができて,コスト関数が適切でなければ真のデプスを求めることはできない。

Davisらは,任意のBRDFを持つ物体に対して,ステレオ視における対応付けを行った。2つの異なる光源下での観測輝度の比は,2台のカメラの間で同一であることを利用した。(以上,宮崎)

O5: (Human) Pattern Detection

C. Fuang(Tsinghua University)らは画像内の回転(RIP, rotation in plane)と観測方向の変化(ROP, rotation off plane)の両方の変化に対応する顔検出手法に関する研究について発表した。本研究はViola and Jones(2001)に代表される一連の高速顔検出アルゴリズム研究の流れの中にあるが,従来法の多くが顔の姿勢変化に対応するためにピラミッド構造や決定木を用いているのに対し,WFS tree structureと呼ぶ独自の木構造モデルの上で複数の可能性を残しながら探索を進めることで高速性と高い検出精度を達成している。また,この探索手法と相性のよい識別器として,Real AdaBoostアルゴリズムを多クラス多次元に拡張したVector Boostingという手法を同時に提案している。

L. Zhao(Maryland)らは,テンプレートを用いた人物(上半身)検出とカラーベースの領域分割を組み合わせることで物体検出と領域分割を同時に行う人物検出手法について発表した。人物の姿勢変化は階層的に配置された複数のテンプレートにより表現されており,また領域分割については非パラメトリックな色分布モデルによる繰り返し推定アルゴリズムを用いている。複雑な背景を持つシーンにおいて,提案手法によって領域分割と物体検出の双方について単独の処理に比べて性能が改善することを示している。

O. Boiman(Weizmann Institute of Science)らは,事前ルールを持たない異常(Irregularity)検出手法について発表した。ここでいうIrregularityとは,監視画像のように人の行動を長時間撮影した映像内に現われる“regular(規則的)でない”動きを指す。Boimanらは定義の難しいこの問題を,現在の入力画像(または画像列)が過去の観測データの組み合わせとして表現できるか否かという問題に置き換えた。入力(問い合わせ)動作を時空間画像内の小領域Patchの組み合わせとして表現し,類似の組み合わせが過去の観測内にあるか否かを探索する。豊富な実例により提案手法の可能性を示した。

X. Lan(Cornell University)らは,人物動作の解析手

法として一般的な多関節運動学モデル (kinematic tree model) では十分に表現されていない部位間の運動の相関性について、因子分析を利用して構成した小数のパラメータを追加することで効率的に表現し、運動復元の安定化を図った研究について発表した。提案法により、例えば歩行時の両手足間に存在する強い相関性を考慮した復元が可能となり、姿勢に関する MAP 推定の結果が改善することが示されている。(以上、内海)

O6: Segmentation

P. Kohli(Oxford Brookes) らは、動画像から高速にオブジェクトを切り出すことを目的として Dynamic Graph Cut を提案している。従来は、各フレーム画像から Graph Cut によりオブジェクトを切り出すという処理を繰り返していたが、隣接フレームではグラフの変化が小さいことを利用すると効率良く問題を解くことができる。同様の考え方に基づく Boykov らの方法では、原理的に terminal ノードとのエッジ (t-edge) の変化にしか対応できなかったが、提案手法によって画素ノードを結ぶエッジ (n-edge) の変化にも対応可能となった。フレームごとに Graph Cut を適用するのに比べて、処理を 3 倍程度高速化できたと報告している。

H. Mumim(Louisville) らは、隠れ、部分的な欠損、xy 方向で異なるスケール変化があっても高精度にオブジェクトを切り出すことを目的として、色と形状情報を用いた Level Set に基づくセグメンテーション手法を提案している。学習サンプルから事前に取得した訓練形状を、入力画像から得た色によるセグメンテーション結果にフィットさせる。訓練形状および色によるセグメンテーションはそれぞれ Level Set 関数によって表現されており、それらのフィッティングは Level Set 関数に基づくエネルギー最小化によって行われる。

J. Wang(Washington) らは、画像をオブジェクトと背景に分離するための混合比率 (α 値) を精度良く求める方法を提案している。従来は、Graph Cut と膨張収縮処理を組み合わせた方法などによって、オブジェクト、背景、それらの境界付近の 3 つの領域に分離した後、境界付近の α 値を算出していた。しかし、境界付近の画素で常に $0 < \alpha < 1$ となり、本来不要な場合でも混合される場合があった。提案手法では、オブジェクトと背景の分離と α 値を同時に推定する方法を提案している。(以上、三田)

O7: Pattern Detection and Grouping

本セッションでは、4 件の口頭発表が行われた。

J. Stahl(USC) らは、シーン中のラインセグメントから convex structure(凸包体) を発見する方法を提案して

いる。すなわち、画像から得られたエッジ情報を適当に補完し、より「良い」convex structure を発見する方法である。評価関数には想定される凸包体の面積、補完されるエッジ長、および凸包体内部の滑らかさ(一貫性) が用いられ、これを最小化することで解が得られる。実験では、実シーン中から得られたもっとも評価された凸包体が表示され、それがシーン中で意味のある物体であることが示されている。またこの手法を拡張することで、オープンバウンダリに対しても適用することができる。

V.S.N.Prasad と L.Davis(Univ. of Maryland) らは、画像から回転相似図形を検出する手法を提案している。彼らは、Snake 等で用いられる GVF(Gradient Vector Flow) を用い、画像中の任意の 2 領域において、それが回転した場合の回転角および回転中心を求め、投票する。同一パラメータに投票される領域の組みをグラフ化し関連付けすることで、回転相似パラメータ(何回転の相似であるかと回転中心位置) を求めている。実験では、花の画像に対して手法が有効に働くことを示している。

M.Shiliman(Microsoft Research) らは、文書のレイアウトおよび構造を自動認識する手法である。ここでは、文書のレイアウトを文法とみなし、この文法を形成するコスト関数を最小化する方法で、結果を得ている。また、様々な特徴量を機械学習によって動的に選択することで、特定の特徴量に依存しない認識を可能にしている。実験では、文書のレイアウトの認識、および数式の構造を認識し、LaTeX に変換する等のタスクを行っている。

M. Rochery(INRIA) らは、通常の Active Contour モデルの発展系である Higher-Order Active Contour(HOAC) と、その初期値を解決するための Phase Field(相転移) モデルを提案している。通常の Active Contour モデルでは、形に対する前提知識を入れることができないか、あるいは多少のズレしか許容できないのに対し、HOAC では、Contour の形状の性質自体(平行ラインなど) を制御することが可能である。一方で、通常の Active Contour と同様、HOAC においても Contour の初期状態が結果にクリティカルな影響を及ぼすが、これを物理シミュレーション等で用いられる Phase Field Model によって解決する手法を提案している。実験では、通常の Active Contour では難しい、衛星画像からの道路の抽出結果などが示されている。

(以上、中澤)

O8: 3D Reconstruction

F. Kahl(Lund 大) は L_∞ -norm を評価尺度として用いた複数画像からの 3 次元復元やカメラ校正の枠組を

提案した．彼は L_2 -norm を評価尺度とした従来の 3 次元復元やカメラ校正手法では解が局所的な解に収束することがあることを指摘し，提案手法が局所解に陥らずに最適な解が得られることを示している．論文では複数画像からの 3 次元復元や，射影変換行列の算出，カメラ校正を L_∞ -norm を用いて実現する方法についてまとめており，実画像による実験で提案手法の有効性を示している．彼の結論によれば L_2 -norm を用いて算出した結果を初期値としてバンドルアジャストメントのような反復手法を用いなくても，提案手法では十分な結果が得られていると述べている．

Q. Ke と T. Kanade(CMU) は F. Kahl と同様に L_∞ -norm を用いた 3 次元復元等の問題を扱っている． L_∞ -norm を用いた手法はアウトライヤに影響を受けやすいことから，アウトライヤが含まれていてもロバストに解を計算するために L_∞ -norm を拡張した新しい評価尺度を導入した． L_∞ -norm はデータごとに得られるある目的関数の最大値を評価尺度にするのに対して，提案した新しい評価尺度 F_m (論文中の記述に従う) は，データごとに得られるある目的関数の値を昇順に並び替えたうちの m 番目の値を用いるものである．そして新しい評価尺度を用いて 3 次元復元等の幾何問題を解く枠組を提案した．

また F. Kahl(Lund 大) と D. Henrion(LAAS-CNRS) は非凸関数の最適化問題を解く方法として階層的な *convex relaxation* に基づく手法を提案した．彼らは先の論文と同様に，バンドルアジャストメントのような反復手法では，精度のよい解を得るためには精度のよい初期値を用いる必要があり，初期値の与え方によっては解が局所解に陥ることを指摘している．それに対して，基礎行列の算出のような非凸関数の最適化問題で大域的に最適解を得る手法として，*linear matrix inequalities(LMIs)* を拡張した手法をまとめた．

P. Kovesei(Western Australia 大) は *shape from shading* 等の手法によって得られた物体面の法線情報から面を復元する手法として基底関数を用いた新しい手法を提案した．この手法が従来手法と異なるのは，基底関数に正規直交系ではない冗長な基底関数を用いていることである．そしてスケールの異なる基底関数を組み合わせることで詳細な面の復元を可能としている．更に法線方向の曖昧性を許容することで法線方向の情報がない場合にも比較的良好な結果が得られることを示している．
(以上，菅谷)

O9: Color and Reflectance

色や反射などに関して 3 件の発表があった．

Pitié らは，画像 A と画像 B が与えられたとき，画像 A を画像 B と同じような色合いに変色させた画像を生成する手法を発表した．従来法よりも自然な色づけを実現している．

Seitz らは，相互反射の影響を除去する手法を発表した．光源ビームをシーンに照射し，光源方向を変えながら複数枚の画像を撮影する．ビームを当てる位置と画素の位置はあらかじめ，1 対 1 の対応を取っておく．画素数と光源数(撮影回数)を n とする．まず，行を画素位置，列を光源位置として，撮影画像の輝度を並べた $n \times n$ 行列 T を構築する．次に， $C^1 = T^1 T^{-1}$ を計算する．ただし， T^1 は，各対角成分が T^{-1} の各対角成分の逆数となっている $n \times n$ 対角行列である．撮影画像を n 次元列ベクトル L_{out} としたとき， $C^1 L_{out}$ が相互反射の取り除かれた画像になる．

Kutulakos らは，透明物体や鏡面物体のデプスと法線を推定する手法を提案した．この論文は honorable mention の一つとなった．まず，光源として，液晶モニタを物体の背後に配置する．液晶モニタを配置する位置の数を M とし，そこから出た光が物体に K 回透過または反射し， N 個の視点から観測したとする．このとき，物体のデプスと法線を推定する問題を $\langle N, K, M \rangle$ 問題と呼ぶことにする．彼らは， $\langle 1, 1, 2 \rangle$ 問題と $\langle 2, 1, 1 \rangle$ 問題と $\langle 3, 2, 2 \rangle$ 問題が解けるということと， $\langle N, 3, 2 \rangle$ 問題が解けないことを示した．ダイヤモンドの形をしたガラスの物体のデプスと法線を推定した結果も示した．
(以上，宮崎)

O10: Feature Matching

K. Grauman and T. Darrell(Cambridge) はカーネルに基づく新しい分類手法を提案した．カーネルに基づく分類は複雑な決定境界を得ることができるが，一般的に計算コストが高い．これは複数の解像度のヒストグラムに特徴集合をマッピングするカーネルによって高速で，対象の配置が複雑な場合にもロバストな分類を可能にしている．そしてハリスオペレータによって検出した特徴点に対して SIFT オペレータなどによって特徴を分類したデータを用いて，提案カーネルを SVM に適用して分類実験を行っている．

H. Ling と D. W. Jacobs(Maryland 大) は物体の変形に対する不変量を定義し，空中になびく旗のように形が変形するシーンでの特徴点の対応付け手法を提案した．これは画像を 3 次元空間内の平面と考え，ある重みをかけて変形した画像の *geodesic distance* が重みを極限 1 に近づけると不変量になることに注目したものである．そして，対応付けの尺度として *geodesic-intensity*

hitogram と呼ばれる 2 次元のヒストグラムを提案して、変形のあるシーンを用いてその有効性を示した。

M. Leordeanu and M. Hebert(CMU) は特徴点の集合間の対応を求める手法として、特徴点間の対応度をノードとして持つ行列を定義し、接続の強いグループを分類する手法を提案した。これは *spectral technique* に基づく方法であるが、最適化過程で弛緩法を用いることで計算コストの低減を実現しているところが従来手法と異なる点である。また、データの誤差やアウトライヤに対してもロバストな分類を実現している。(以上、菅谷)

O11: Tracking

このセッションでは、4 件の発表があったが、その内の 2 件は同じ研究グループによる点や直線などの特徴ベースの追跡であった。

B. Han と L. Davis は、対象モデルを矩形領域内のピクセル毎の混合ガウス分布として表現し、ガウス分布の個数や平均、分散を更新しながら追跡した。対象をピクセル毎の混合ガウス分布として表現することで、詳細な対象表現を実現し、それを逐次更新することで対象の見えの変化に対応している。実際に人の頭などの追跡結果で照明変化や頭の方向変化に対して正確に追跡している様子を示した。

C. Shen らは、Mean-Shift による追跡を、初期の検出や対象が高速に移動する場合に対応するための改良手法について発表した。位置やスケールについて複数の初期値をランダムに選択しながら Mean-Shift を適用することで、局所解を避けて大域的な最適解を求めている。

C. Kemp と T. Drummond は、特徴集合のクラスタリングによって、カメラ姿勢を高速に追跡する手法を発表した。高次元の姿勢パラメータ空間を、部分的なパラメータ成分を決定する特徴集合にクラスタリングし、少数の特徴から低次元の姿勢パラメータの推定する問題に分解した。ヘリコプターで撮影した迷路の画像からカメラの姿勢を推定する結果などを示した。

E. Rosten と T. Drummond は、点と直線の特徴を組み合わせた高速なカメラ姿勢の追跡アルゴリズムについて発表した。点の対応付けによっておおまかな姿勢を決定し、それを元に初期値と姿勢の事前分布を与えて直線追跡によって詳細な姿勢を決定している。また、対応点を高速に探索する方法についても議論している。発表では、高速に振り回しているカメラで撮影した映像からリアルタイムでカメラの姿勢を追跡する結果が印象的であった。(以上、加藤)

O12: Cameras and Calibrations

R.I.Hartley(National ICT, Australia), S.B.Kang

(MicrosoftResearch) は、平面のキャリブレーショングリッドを撮影した複数画像からレンズ歪み、歪み中心の推定を行う手法を提案している。このテーマについては既に広く研究されているが、この研究では、 $1 + k_1 r^2 + k_2 r^4 + \dots$ に代表されるようなレンズ歪みモデルを一切用いずパラメータなしでの推定を行うところに特徴がある。同手法では、まず歪み中心をエビ極として算出し、次にキャリブレーショングリッドから理想カメラへのホモグラフィを算出する。ここで、画像上における同一半径の円周上付近における歪み補正距離の分散が最小となるようにホモグラフィを決定することで、レンズ歪みをパラメータなしで推定している。著者らは、特定のレンズ歪みモデルを用いないことで様々なカメラに対応できるとしている。

K.Tieu(MIT) らは、重複領域を持たない複数の道路監視カメラに対して、動体の進入退出情報を統計的に解析することで、カメラの設置地点(ノード)間の接続情報、移動時間分布を推定する手法を提案している。この研究では、監視カメラ間を結ぶ道路を通過する車両や歩行者の移動時間には相関があるという仮定に基づき、各ノード間が接続されている可能性を移動時間の分布のエントロピーから算出している。また、Markov Chain Monte Carlo 法を用いて、対応関係が未知の進入退出情報からノード間の接続可能性を統計的に推定している。

O.Drbohlav と M.Chantler(Heriot-Watt,UK) は、Photometric Stereo における未知光源下での物体法線と反射率の推定の任意性について、光源方向の異なる 2 枚の画像上における鏡面反射を用いることで解決を図る手法を提案している。この研究では、画像間で最低 2 画素の鏡面反射を特定できれば、Photometric Stereo を線形に校正でき、パラメータの任意性を解消できることを理論的に証明している。(以上、佐藤(智))

3 ポスターセッション

Poster session 1

本セッションでは、全 42 件の発表が行われた。執筆担当者らが分類したところ、その内容は、顔に関する研究 2 件、人物行動認識 5 件、カメラキャリブレーション関係 3 件、トラッキング 3 件、物体認識・識別 8 件、領域分割 4 件、BRDF や光源や形状に関する研究 7 件その他 10 件である。

まず、顔および動作認識についていくつかの研究を紹介する。

Fan(Shanghai Jiao Tong Univ.) らは、顔画像からそ

のビューおよびクラスを認識する手法である。認識の高速化および効率化のため、SVMの入力として認識に有効な部分特徴を選択する Feature Selection 法 (SVM-DFS) を提案している。

Yang(Hong Kong Polytechnic Univ.) らは、PCA と ICA の組み合わせによる顔認識手法 2 種を比較している。PCA-I では、画像からその平均を引いたものを学習に用い、PCA-II では通常の PCA を白色化 (PCA における分散の係数を揃える) したものをを用いる。PCA-I,II の結果をそれぞれ ICA に入力し、ICA-I, ICA-II 手法とし比較している。FERET データベースを用いた実験結果からは、PCA-I と ICA-I の間には差がみられず、ICA-II が通常の PCA に比べても良い結果を示すものだった。しかしながら著者らは、認識においては PCA の処理法の影響が最も大きく、ICA による効果は少ないと結論付けている。

Y.Ke(CMU) らは、人物の動作を、時空間の 3 次元ボリュームにより認識しようとする研究である。彼らは、Viola&Jones の rectangle feature を 3 次元空間に拡張した手法を提案し、Stand Up, Sit Down 等の基本動作についての認識を行っている。また、この空間において動作をクラスタリングする試みも行っている。

Yang(Univ. of Maryland) らは、Particle Filter の高速化を図った論文を発表している。この追跡手法では、色ヒストグラムおよびエッジヒストグラムを用いているが、そのヒストグラムマッチングを Harr-like rectangle features を用いて高速化し、また、サンプルの選択について、明らかに低い確率をとるものは直ちに棄却し、高い確率をとるものはより詳細にチェックするという階層的選択を行うことで、高速化を図っている。

また、BRDF や光源や形状に関する研究はバラエティに富んでいた。Sinha らは、photo-hull の問題を graph cut で解いた。佐藤らは、任意光源下の見えを再現するために拡張光源を使ってサンプリングした。Kim らは、影から光源環境を推定した。Agrawal らは、integrability 制約を用いて法線から高さを計算した。

Vogiatzis らは、境界点を利用して物体形状と BRDF パラメータと光源環境を求める手法を発表した。物体表面上のある点の接平面が 2 つのカメラのカメラ中心を含むとき、その点を境界点という。境界点は輪郭抽出とエッジ拘束により求めることができ、境界点での 3 次元位置と法線も求まる。まず、この境界点だけを用いて、光源環境と BRDF パラメータを推定した。その後、物体表面上の全ての点で、照度差ステレオ法を用いて物体の表面法線を推定した。

Atkinson らは、表面が滑らかな不透明物体を回転し

て 2 視点から偏光データを取得し、物体の表面法線を推定する手法を発表した。偏光データから、各点につき法線の候補が 2 つ求まる。視点 1 の点 A の周囲で形状を 2 通り求め、視点 2 の点 B での 2 通りの形状のどちらかと一致した場合、点 A と点 B を対応点とする。対応点を利用して、2 つの法線候補のうち、正しい法線を決定した。

Goldman らは、物体の法線と BRDF を求める手法を発表した。物体表面の各点は、2~3 種類の BRDF の内挿で表すことができると仮定する。例えば、葉っぱの場合、濃緑色の葉の部分と黄緑色の茎の部分の 2 種類の BRDF で表すことができる。それ以外に、BRDF パラメータの滑らか拘束なども用いて、物体の形状と BRDF パラメータを推定した。(以上、宮崎、中澤)

Poster session 2

このセッションでは、幅広い分野に関して合計 59 件のポスター発表が行われた。以下では、いくつかの印象に残った論文について紹介する。

Y.Wang(UCLA) らは、2 個の、同一のトポロジーを持ったポリゴンメッシュのマッチングを行う手法について述べている。手法の特徴として、それぞれのメッシュの幾何学的な特徴 (conformal factor と mean curvature) を、conformal mapping(等角写像) によって 2 次元パラメータ空間上の画像として表現し、マッチングをその空間上で行う点が挙げられる。画像上でのマッチングは、相互情報量エネルギーの極値を求めることで行われる。

S. K. Zhou(シーメンス) らは、高次元の画像データを、Haar-like 特徴と Boosting による特徴選択によって低次元の特徴ベクトルで表現し、画像から出力パラメータへ写像を学習する方法を提案し、顔画像からの年齢推定や CT 画像画像からの腫瘍の検出などの結果を示した。画像から直接出力パラメータへの写像を学習する技術は、一見複雑な問題を簡単に扱うことができ今後の発展が期待できる。

D.Nehab(プリンストン大) らは、Rectification されたステレオ画像が与えられ、窓画像によるマッチングを行う場合に、subpixel 単位で対応点を定める方法について述べている。既存手法では、この問題は、片方の画像で窓を固定し、もう一方の画像上で窓を一定間隔で (通常 1 画素) 動かしながら、離散的な 1 変数のコスト関数を求め、曲線当てはめによって極値の位置を求めることで解かれる。Nehab らは、この方法ではマッチングの推定値に偏りが生じること、2 変数のコスト関数に曲面当てはめを行う手法で偏りを解消できることを示した。

H.Stewenius らは、3 画像の対応点から、triangulation

によって3次元復元を行う際の、最適解の求め方について述べている(同様の問題は、2画像についてはHartleyらによって解かれている)。彼らは、射影された点と観測値との差の自乗和によるコスト関数を最小化するという条件から、3個の3変数6次多項式によって構成される連立方程式を得、これをGroebner基底を用いて解いている。これによって、全てのコスト関数の極値と停留点が得られる。彼らは、これを実データ及び合成データに適用し、極大値、極小値、各種停留点の数の変化について報告している。

A. Banno(東大)らは、固定されていないレンジセンサで計測した複数の3次元データの歪みを除去し、位置合わせする手法を提案している。初期位置合わせ用の外部パラメータは、計測装置に固定したカメラで撮影した画像列から因子分解法により得られる。また、彼らの開発した、空中から3次元計測するためのFloating Laser Range Sensor (FLRS) や、計測結果によるデジタルアーカイブ等も注目されている。

D. Feldman(エルサレム大)らは、全方位画像に円形クロススリット(X-Slits)を用いることで、リアルタイムなイメージベース・レンダリング(IBR)を実現する手法について述べている。X-SlitsによるIBRはデータのコンパクトさ等から注目されているが、X-Slits画像は、パースペクティブではないため、本質的に歪むという問題があった。著者らは、シーンの粗い3D構造を用いることでこれを軽減し、さらに、IBRを背景とするシーン中に仮想物体をレンダリングする手法を提案している。

H. Tao(カリフォルニア大)らは、二値のマスク画像で表現される非直交基底による部分空間の特性を分析し、このような基底を用いて主成分分析やテンプレートマッチングへの応用について検討した。二値の非直交基底は、Integral Imageを使って高速に内積を計算することができ、またHaar-like特徴など近年その有効性が注目されており、そのような基底で表現される空間の特性の分析は重要と考えられる。(以上、加藤、川崎、古川)

Poster session 3

C-P. Chen(Taipei)らは、顔認識の精度向上を目的として、照明変動を補正する手法を提案している。従来は、同じ人物に対して異なる光源下で撮影された複数の画像が必要であったが、提案手法では一枚の画像と事前に他の人物から学習したモデルを用いることで実用性を高めている。事前学習では、まずWeissら(ICCV2001)の方法により、Appearance image(学習用の画像)をReflectanceとIllumination成分に分解する。

次に、取得した複数のIllumination成分にPCAおよびNMF(Nonnegative Matrix Factorization)を適用することで照明モデルを得る。認識時は、入力画像から照明による成分を除き、Reflectance成分のみを用いる。

(以上、三田)

M.Blank(Weizmann,Israel)らは、人物の動作を撮影したビデオからシルエットを抽出し、シルエットを時間方向に並べることでspace-time空間を用いた動作のボリューム表現を行っている。このようなボリュームデータを解析することによって、比較的解像度が粗く雑音の多い画像に対しても高速かつ高精度な動作認識が可能であるとしている。J. Starck, A. Hilton(Surrey,UK)は、多数のビデオカメラで撮影された動画像によって復元された非剛体運動を伴う三次元サーフェスモデルに対して、密な対応点探索を高速に行なう手法を提案している。この手法では、三次元モデルを球面上に展開し、球面画像上で2次元的に対応探索を行なう。実験では、様々な動作を伴う人物の三次元モデルにスケルトンモデルを自動で当てはめ、モーションキャプチャへの利用可能性を示している。

L.Z.Manor(California)らは、球面画像を平面状に展開する際の問題点について議論している。一般に、球面画像から平面への展開には透視投影モデルが用いられることが多いが、透視投影では広角の画像を生成する際に画像周辺で像が引き伸ばされてしまうために、視覚的な違和感が生じる。この研究では基礎的な解決手法として、複数面から成る投影面への透視投影画像を合成する手法が有効であることを示している。さらに画像の前景、背景を分離して別々に投影することで、人間の視覚にとってより違和感の少ないパノラマ画像を生成している。ただし、現時点で手法を利用するためにはいくらかの手作業が必要である。A.Ilie, G.Welch(NorthCarolina)らは、マルチカメラシステムにおけるカメラ間の色の校正手法を提案している。この手法では、各カメラで同一のカラーパターンを撮影することで校正を行うが、一般的に行われているソフトウェアによる校正だけでなく、ハードウェアの撮影パラメータ(明度、彩度、ホワイトバランス、ゲイン、ガンマ等)をも最適化することで、可能な限りカメラ間の色を一致させる。

S.Park(KAIST,Korea)らは、レンジファインダを用いて計測・生成される3次元モデル上の形状欠損の問題に対して、形状とテクスチャを同時に修復する手法を提案している。この手法では、まず3次元点群に対してdeformable surface modelを用いることで表面形状を推定するが、このステージにおいて形状の欠損箇所を自動判定する。次に欠損箇所周辺のコンテキスト

に近い形状・テクスチャを持つ箇所を同一の3次元モデル内から探索し、それを欠損箇所周辺に変形させてはめ込むことで形状・テクスチャの同時修復を行っている。G.Zeng(HKUST,HongKong)らは、複数画像からの3次元復元問題に関して、space curving法に滑らかさの制約を導入する手法を提案している。この手法では、滑らかさの制約を導入する際に問題となるNP-hardな最適化問題をgraph cutを導入することで解決している。更に、space curving法を基礎としているにもかかわらず、3次元モデルを微小パッチの集合として復元でき、最終的に微小パッチを統合することでなめらかな表面を持つ3次元パッチモデルを生成している。

(以上、佐藤(智))

Poster session 4

M. I. A. Lourakis と A. A. Argyros はバンドルアジャストメントの実装にレーベンバーグ・マーカート法が最も有効かというタイトルで M. Powell が提案した *dog leg* アルゴリズムとレーベンバーグ・マーカート法の比較を行った。*dog leg* アルゴリズムがレーベンバーグ・マーカート法と異なるのは *trust region* と呼ばれる領域を用いて反復解を制御することである。彼らの実験によると解の精度はどちらの方法を用いても同等であるが、実行時間は *dog leg* アルゴリズムが優れているという結論が得られている。

Y. Zheng と D. Doermann は2次元の非剛体に対するロバストな点対応手法を提案した。これは変形後も近傍の点との関係はそれほど変化しないと仮定し、点対応問題をグラフマッチング問題として定式化したものである。対応処理に Rosenfeld らによって提案された *relaxation labeling* を取り入れることで、*shape context* 法や TPS-RPM アルゴリズムよりも良好な結果を得ている。

(以上、菅谷)

N. J. W. Morris と K. N. Kutulakos は2台の固定カメラによる観測によって屈折率が未知の液面の3次元形状を復元する手法を提案した。槽の底面に既知パターンを置き、液面が静止した状態での観測を出発点として観測パターンのフローをサブピクセル精度で追跡し、スネルの法則を利用して反射率と液面形状を推定する。その際、両カメラの観測と既知パターン上の対応点により定義される refractive disparity を最小化することにより液面の高さに依らず安定した解が得られることを示した。さらにバンドルアジャストメントによって復元モデルによる再投影誤差を最小化し復元精度を向上させている。

T. Mita と T. Kaneko, O. Hori は、複数の Haar-like

Feature を組合わせてそれらの共起性を評価する Joint Haar-like Features による顔検出について発表した。識別に適した Haar-like Feature の組合わせの探索手法、見つかった組合わせを弱識別器とする AdaBoost アルゴリズムの構成手法を紹介し、テストデータを用いた Viola and Jones 法との比較により提案手法の優位性を示している。

K. Hara と K. Nishino, K. Ikeuchi は、実際的な条件下で Torrance-Sparrow 反射モデルのよい近似を与える Spherical Torrance-Sparrow Model を提案し、同モデルを初期推定に用いることで既知形状を撮影した単一の画像から物体の反射特性と光源を同時に推定する手法を示した。従来法と異なり、光源の数についても推定を可能としている。

R. Fergus と L. Fei-Fei, P. Perona, A. Zisserman は、検索エンジンである google の画像検索結果を画像認識の学習画像として利用するという試みについて報告した。機械学習の枠組みである probabilistic Latent Semantic Analysis を位置と大きさの変動を考慮して拡張した TSI-pLSA と呼ぶモデルを提案し、画像認識に適用した。学習結果は良好とは言えないが、いくつかのキーワードについて学習により画像検索の精度が向上することを示すなど興味深い結果を紹介している。

(以上、内海)

4 おわりに

第10回コンピュータビジョン国際会議 (ICCV2005) について、その概要を出席者有志によってまとめた。会議の全体の様子と研究動向をできるだけ正確に報告できるように努めたつもりであるが、紙面の都合上、全ての発表を網羅しているわけではない。より詳細な内容については、[1] を参照して頂ければ幸いである。

最後に、ICCV2009の開催に関して、General Chair: 松山隆司(京大), PC Chairs: 横矢直和(奈良先端大), Roberto Cipolla(ケンブリッジ大), Martial Hebert(CMU), Xiaoou Tang(マイクロソフトアジア) として日本が京都開催を提案し、採択されたことを記して本報告を締めくくることにする。

(以上、杉本, 佐藤(洋), 和田)

参考文献

- [1] *Proceedings of the 10th IEEE International Conference on Computer Vision*, IEEE CS Press, 2005.