

The 11th IEEE International Conference on Computer Vision (ICCV2007) 報告

池田 聖 (奈良先端科学技術大学院大学) 岡部 孝弘 (東京大学)
佐藤 智和 (奈良先端科学技術大学院大学) 阪野 貴彦 (東京大学)
向川 康博 (大阪大学) 山崎 俊太郎 (産業技術総合研究所)
佐藤 洋一 (東京大学)

jsei-i@is.naist.jp takahiro@iis.u-tokyo.ac.jp
tomoka-s@is.naist.jp vanno@cvl.iis.u-tokyo.ac.jp
mukaigaw@am.sanken.osaka-u.ac.jp shun-yamazaki@aist.go.jp
ysato@iis.u-tokyo.ac.jp

あらまし: 2007年10月16日～19日にブラジル リオデジャネイロで開催されたコンピュータビジョンに関する国際会議 ICCV2007 の概要を報告する。

Report on the 11th IEEE International Conference on Computer Vision (ICCV2007)

Sei Ikeda (NAIST) Takahiro Okabe (University of Tokyo)
Tomokazu Sato (NAIST) Atsuhiko Banno (University of Tokyo)
Yasuhiro Mukaigawa (Osaka University) Shuntaro Yamazaki (AIST)
Yoichi Sato (University of Tokyo)

Abstract: This report gives an overview of the 11th IEEE International Conference on Computer Vision (ICCV2007) which was held in Rio de Janeiro, Brazil from October 16th to 19th, 2007.

1 はじめに

2007年10月16日から19日までの4日間にわたり、第11回IEEEコンピュータビジョン国際会議(The 11th IEEE International Conference on Computer Vision (ICCV2007))がブラジル リオデジャネイロのホテル Sofitel Rio de Janeiro で開催された。ICCVはIEEE Computer SocietyのPAMITC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り、今回はその11回目にあたる。ICCV2007では、Larry Davis(University of Maryland), Patrick Bouthemy(INRIA), 池内克史教授(東京大学)の3名がGeneral Chairsを務め、Dimitris Metaxas(Rutgers University), Baba Vemuri(University of Florida), Amnon Shashua(Hebrew University), Harry

Shum(Microsoft Research)の4名がProgram Chairsを務めた。会場には世界中から700名を超える研究者や学生が集まり、4日間にわたって活発な研究発表や討論が行われた。ICCVは質の高い論文を厳選し、シングルトラックのオーラル発表を行うことで知られているが、今回も1190件の投稿論文に対し、全体の採択率が23.5%、オーラル発表に至っては3.9%と極めて狭き門となった。テーマ的には物体認識や学習ベースの手法に関する論文が多数発表され、少し前に活発に発表されていた多視点幾何やトラッキングに関するものが減少する傾向が見受けられた。

投稿論文数の増加に伴い、査読の負担が増していることを受け、ICCV2007では2段階査読という新たな仕組みが導入された。第1段階では各論文あたり10名の査読者がコメントを書かずに採択または不採録かだけを判断し、ここである程度以上の評点

を得た論文のみが従来と同様の査読を行う第2段階へと回された。査読の負荷を少しでも軽減すると同時に査読精度を向上させることを目指し導入された2段階査読システムであったが、残念ながらその評判はあまり良くなかったようである。そのためか、次回以降のICCVではこれまでと同様の査読のやり方に戻ることが予定されている。

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を表2に示す。各セッションの詳細は次節以下を参照して頂きたい。

ICCVにおける最優秀論文賞のMarr PrizeはBrad Davis(University of North Carolina)らによるPopulation shape regression from random design dataに贈られた。次次のHonorary Paper Mentionsは、Ying Nian Wu(UCLA)らによるDeformable template as active basis, Abhijeet Ghosh(University of British Columbia)らによるBRDF acquisition with basis illumination, Manmohan Chandraker(UC San Diego)らによるGlobally optimal affine and metric upgrades in stratified autocalibrationの3件に論文に贈られた。また、第1回目のICCVが開催されてから20年目の節目にあたる今回からThe Azriel Rosenfeld Lifetime Achievement Award in Computer VisionとComputer Vision Significant Researcher Awardが新設され、それぞれCMUの金出武雄教授とUCLAのDemetri Terzopoulosに授与された。

(以上、佐藤(洋))

2 オーラルセッション

O1: Learning I

従来、グラフマッチングの研究では、NP困難な最適化問題として知られている二次割当問題の近似解法の研究が盛んに行われていた。Caetano (ANU)らは、グラフマッチングを学習の問題として捉え直して、対応付けの与えられたグラフ対から類似度関数を推定する枠組を提案している。単純な一次割当てと学習の組合せにより、従来の二次割当てと同程度の精度でかつ遥かに高速なマッチングが実現できることが報告されている。

Frome (UC Berkeley)らは、k-最近傍決定則に基づくカテゴリ識別法を提案している。画像間の距離

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	27	110	137
中国	3	26	29
フランス	3	21	24
イギリス	3	13	16
イスラエル	2	8	10
カナダ	2	7	9
スイス	2	4	6
ドイツ	1	10	11
オーストラリア	1	6	7
シンガポール	1	1	2
ベルギー	1	1	2
オーストリア	1	0	1
スウェーデン	0	5	5
インド	0	4	4
日本	0	3	3
韓国	0	3	3
チェコ	0	2	2
その他 (計9カ国)	0	9	9
合計	47	233	280

表 2: セッション一覧

Oral Sessions
O1: Learning I
O2: Stereo and Reconstruction
O3: Features
O4: Segmentation & Grouping
O5: Vision and Graphics
O6: Statistical Methods
O7: 3D Reconstruction
O8: Recognition and Registration
O9: Tracking
O10: Learning II
O11: Calibration and Localization
Poster Sessions
P1: Learning and Recognition I
P2: Low Level Vision & Applications
P3: Tracking & Segmentation
P4: Cameras, Stereo, Pose Estimation, Video & Shape
P5: Learning and Recognition II
P6: Registration, Shape, Illumination, Texture, Segmentation

を局所特徴量間の距離の線形結合として定義し、同一カテゴリに属する画像間の距離が異なるカテゴリに属する画像間の距離よりも小さくなるように、三枚組の学習画像から、マージン最大化により線形結合係数を求めている。Caltech101 データセットに対して、従来手法を上回る識別性能が得られている。

Vedaldi (UCLA) らは、窓関数により平滑化されたパターンにアダプストを適用する Parzen AdaBoost を提案している。並進、回転、スケール、コントラストなどの変動に対する頑健性を高めるために、画像の微小変動を反映した非等方的な窓関数を利用している。また、最急降下法による弱識別器の最適化や Haar ウェーブレットを用いた効率化などの工夫も行っている。通常のアダプストなどと比較して、少数個の弱識別器の組合せで高い識別性能が得られることが報告されている。

Stein (CMU) らは、時系列画像から、見えと動きの両方を手掛かりにして遮蔽輪郭線を抽出する手法を提案している。画像を過剰に領域分割して遮蔽輪郭線の候補を求めたのち、ロジスティック回帰により求めた輪郭線らしさと輪郭線の交点の整合性から、MAP 推定の枠組で遮蔽輪郭線を抽出している。見えだけでなく動きを併用すること、および、大域的な整合性を考慮することの有効性が実験により示されている。(以上、岡部)

O2: Stereo and Reconstruction

このセッションでは3件の発表があった。セッション名からは少し想像しにくいですが、幾何学的な形状復元法ではなく、いずれもフォトメトリックステレオ・透明物体の光線追跡・鏡面反射のフローなどの反射屈折を手がかりとした形状推定法である。反射屈折の現象は、未だに完全にはモデル化できていないため、新たな光学現象を利用したり、特殊な光学器機を用いるなどの工夫によって、新しい手がかりに基づく形状復元法を発見することができる。フォトメトリに関する研究が、このように形状復元法としてまとめられているのはおもしろい。

Alldrin と Kriegman (Univ. of California) は、BRDF の左右対称性を利用したフォトメトリックステレオを提案した。フォトメトリックステレオは、光源や BRDF にどのような条件を課すかによって、解き方や、どこまで曖昧性が残るかが異なる。本研究は、等方な BRDF は視線方向と法線を含む平面に対して左右対称であるという、良く知られてはい

るものの従来はあまり利用されてこなかった特性を利用したことが特徴である。これにより、一般性をさほど損なうことなく、放射輝度分布が左右対称となるような法線を見つける問題として定式化した。なお、Kriegman はフォトメトリックステレオ関係だけでも、ICCV2007 では他に1件のポスター発表があり、さらに CVPR2007 でも3件(うち2件はオーラル)を通しており、これらから最新の研究動向が伺える。

Morris と Kutulakos (Univ. of Toronto) は、内部材質が不均一な透明物体の形状を復元するために、光源位置を変えたときの表面上の放射輝度の変化を解析する手法を提案した。液晶モニタを照明代わりに使用し、表面上の放射輝度を、光線空間のスライスである Scatter-Trace と呼ぶ画像で表現した。この Scatter-Trace から、ある奥行きを仮定したときの物体表面の鏡面反射成分のみを取り出し、鏡面反射の軌跡の振る舞いから仮定した奥行きを検証した。物体表面の鏡面反射を手がかりとするという点では、Specular Stereo に似ているが、内部材質が不均一な場合にも適用できることが特徴である。

Adato (Ben-Gurion Univ., Israel) らは、環境が物体表面に鏡面反射として歪んで映り込むことを利用し、そのフローから物体形状を推定する方法を提案した。例えばグリッドパターンなどの既知物体を環境に配置して、その映り込みを利用する方法は以前から存在するが、本研究では環境は無遠に存在しカメラは平行投影であると単純化することで、環境が未知の場合の解法を示した。論文中で示された実験結果は、残念ながら鏡面加工した球面の形状復元結果のみであるが、一般物体上で生じる映り込みから、実際にどれくらいの精度が出るのか興味がある。(以上、向川)

O3: Features

Ahuja (UIUC) らは、平面状とみなせる3次元空間中(2.1D)を撮影した画像に繰り返し出現するテクスチャ要素(Texel)の自動抽出を行った。入力画像に対し、多重解像の領域分割を行い、画像を Segmentation Tree として表現する。この木構造から類似したブランチを抽出することで、Texel のモデルを特定する。さらに、各 Texel が空間中に繰り返し発生する頻度を、確率モデルとしてパラメータ推定することにより、画像中の各 Texel を抽出した。

Wong (Univ. Cambridge) らは、画像列からグ

ローカルな情報を利用した特徴点抽出を提案した。まず画像列を2次元行列として表現し、その行列にNon-negative Matrix Factorization (NNMF) を適用し、分解した2つの行列、Subspace 画像と係数行列から特徴点を抽出している。実験において、提案手法と他の手法とを用いて特徴点抽出を行い、ジェスチャー認識に行ったところ、提案手法による特徴点を利用した方が認識率が高いことを示した。

Pham (Nanyang Technological University) らは、カスケード型の顔識別器において、学習にかかっていた計算時間を大幅に短縮する手法を提案した。提案手法では Haar 特徴を用い、学習には平均、分散といった統計量で行うことによって、従来の $O(NT \log N)$ (ただし、N: サンプル数、T: 特徴数) であった処理時間を $O(Nd^2 + T)$ (ただし d: サブウィンドウのピクセル数) に短縮することができた、と述べている。

Zheng (Siemens) らは心臓の3次元CTデータから、左右の心房・心室の自動的なセグメンテーションを行った。対象物体の位置、姿勢、スケールの推定には、提案手法である Marginal Space Learning を用いて、少ない次元数からの効率的な計算を行った。また、グローバル・ローカルな特徴を抽出するためには Steerable Features という手法を提案している。これらの手法を用いて、柔軟物体である心臓の3次元データに対して、精度良くセグメンテーションを行った。(以上、阪野)

O4: Segmentation & Grouping

Sinop (CMU) と Grady (Siemens) は、画像のセグメンテーションの統一的な枠組みを提案している。この枠組みは、代表的な手法である Graph-Cut 法と Random Walker 法を特殊解として含んでいる。またそこから自然に導かれる第3の手法として、 L_∞ ノルムを最小化することによって、新しいアプローチが存在することを示している。この方法は、初期化時に与えられる seed の数が多い場合にも安定して処理を行えるなどの特徴がある。

Xiao (Hong Kong University) らは、多視点画像を入力として、再構築した3次元点集合と対応する2次元画像の同時セグメンテーションに対する新しいアプローチとして、重み付きのグラフに対するラベル付け問題に帰着するアイデアを提案している。まず3次元点と2次元画像の類似性を用いてグラフ構造を構築し、階層的で疎な類似性伝播アルゴリズムで最適なラベル付けを行う。ただし全自動のアル

ゴリズムで高い精度が実現できないため、教師データありの類似性伝播アルゴリズムを用いて精度を改善する。

Zhu (University of Pennsylvania) らは、輪郭線検出によって得られた2次元線分情報をグループ化し、位相を復元する方法を提案している。untangling cycles と呼ばれるグループ化の方法を用いてランダムウォークグラフを構成し、その固有値問題を解くことによって連続した主要な輪郭線を抽出している。(以上、山崎)

O5: Vision and Graphics

Bai と Sapiro (University of Minnesota) は、静止画像と動画に適用可能な、アルファマットの抽出を対話的に行う方法を提案している。前景と背景を定義する scribble をユーザーが与え、画像の勾配情報で重み付けされた測地線距離マップを、最適な線形アルゴリズムで計算する。この測地線情報からアルファマットを計算することを示している。局所的にマットの精度を向上させるための、ユーザーインターフェースについても述べている。

Ding (Pennsylvania State University) らは、動画の欠落部分を埋めるための inpainting の手法を提案している。フレーム画像中のデータが作る行列のランクを最小化することにより、各フレームの最適な descriptor を設計することができる。これを用いると、非周期的な運動や、背景やカメラに動きのある場合にも精度よく inpainting することができる。実験では、動的なテクスチャや物体の遮蔽をうまく再現することに成功している。

Goesele (University of Washington) らは、撮影環境の変化に対して頑健な multi-view stereo の手法を提案している。ライティングや画像サイズが異なり、対象シーンと無関係な遮蔽物を多く含む、オンラインの写真データベースから対象形状を復元することを目指している。提案法では、Structure from Motion によって得られたカメラ情報を元に、シーン復元に用いる画像を適応的に順次選択することで、頑健に形状復元を行っている。

Pritch (Hebrew University) らは、監視カメラなどによって撮影された、非常に長くほとんど変化のない動画から、要旨動画を作成する方法を提案している。まず動画中に観測される移動物体や活動を抜き出してデータベースを作成し、次にユーザーの要求にしたがってそれらを編集して望ましい特徴を

持つ要旨動画を生成する。個々の活動それぞれの発生時刻とは無関係に抜き出されるために、互いに因果関係がある場合などの工夫などに関して考察されている。

Ghosh(University of British Columbia)らは、BRDFを計測する新しい手法を提案している。提案法では、BRDFを基底関数の線形和表現し、基底に対応する光を対象物に直接照射することで、基底係数を計測している。著者らは実際にBRDFの計測装置を作成し、小型の簡単な装置を用いて、高速にBRDFを計測できることを実証している。

(以上、山崎)

O6: Statistical Methods

本セッションの発表から Mar Prize が選出されているが、それ以外の発表についても質が高いものが多く、専門分野が遠い出席者にも注目されたセッションだったと思われる。本節では、Mar Prize を受賞し、既に多くの人々が注目する Davis(University of North Carolina)らの研究紹介をあえて避け、特にシンプルで汎用性の高い Sheikh らの研究を簡単に紹介する。

Sheikh(Carnegie Mellon University)らの発表は、大規模なデータをクラスタリングするために mean-shift 法を改良したアルゴリズム modoidshift 法の提案であった。両手法とも、データ点の局所密度が最大となる点を見つけるために、繰り返し処理の中で、ある探索点とそれを中心とした一定範囲内における複数のデータ点との距離が最小となる点を探し、それを次の探索点としてこの処理を繰り返す、というアプローチをとっている。両手法の相違点は、mean-shift 法では、任意の点が探索点となりうるのに対し、modoidshift 法では、データ点のみが探索点となる、ということだけである。

この小さな違いによりもたらされる modoidshift 法の利点は複数あり、増加し続けるデータに対して効率よくクラスタリングできるという点やクラスタの線形分離可能性が必要でないという点、あらゆるメトリックで動作するという点、クラスタ数を指定する必要がないという点が挙げられる。特に、最初の特徴は発表中で強調していた部分であり、最大の特徴である。

(以上、池田)

O7: 3D Reconstruction

本セッションの発表内容は、4件中の3件が複数

の画像間もしくは画像と光源間の幾何拘束を用いる伝統的な 3D Reconstruction であったが、最後の1件 Hoiem(Carnegie Mellon University)らの研究は、そのような拘束が前提とされていないという意味で他の研究とは異なり特徴的であった。

Hoiem らの研究では、1枚の画像から複数の物体間のオクルージョンエッジを抽出し、物体間の奥行き関係や擬似的な depth-map 生成する手法を提案している。ただし、画像内に含まれるオブジェクトを空、地面、複数のオブジェクトの3種類と仮定し、水平線は画像のおよそ中央を水平に通ることやオブジェクトは地面に設置しているという、一般的な写真撮影における風景写真を前提としている。オクルージョンエッジの抽出では、単純な指標を基に細かく分割された領域を再統合する際に、各領域の境界線がオクルージョンによるものか否か、境界線両側の表面の隠蔽関係はどうなっているかを判定する。これらの判定は、各領域の2次元的特徴や3次元的な隠蔽関係などが与えられたときのエッジの接続関係の学習を基にしている。

初期のコンピュータビジョンにおける線画解釈の議論に近いようにも感じられるが、大きな違いは、最新の領域分割手法や学習理論を踏まえたものであり、従来に比べて遥かに複雑な情景に対し、我々に近い奥行き認識結果が得られていることである。

(以上、池田)

O8: Recognition and Registration

Rabinovich (UCSD) らは、シーンのコンテキストに着目した物体の識別・検出のための手法を提案している。画像を領域分割して、標準的な BoF により各領域のカテゴリラベルの候補とそれらの信頼度を求めたのち、条件付確率場の枠組を用いて、カテゴリの共起に基づきラベルを修正している。PASCAL と MSRC のデータセットを用いて、カテゴリの共起を組込むことの有効性を確認している。

Savarese (UIUC) らは、見えと形状に基づく3次元物体のカテゴリ識別と姿勢推定のための手法を提案している。カテゴリ内の物体形状の変動を吸収するために、物体表面を構成する部分領域ならびに部分領域間のホモグラフィ変換を用いて、より抽象的に物体形状を記述している点が特長である。姿勢変動の大きい画像からなるデータベースを作成して、提案手法の性能を評価している。

Jhuang (MIT) らは、局所的な時空間特徴に基づ

く動作識別法を提案している。視覚野における運動情報処理の神経生物学的モデルに着想を得て、単純な画像処理を階層的に組合せて特徴を抽出したのち、SVMを用いた識別を行っている。KTH, Weizmann, および、UCSDの人物や動物の映像データセットに対して実験を行い、従来手法と同程度かそれを上回る識別性能を得ている。

従来、単眼映像に基づく紙面などの非剛体のレジストレーションにおいて、二次元ワープに重なりを生じるような自己遮蔽を扱うのは困難であった。Gay-Bellile (Blaise Pascal) らは、自己遮蔽を遮蔽輪郭線への二次元ワープの収縮とみなして、重なりに対するペナルティ項を導入することで、安定な画像間の対応付けを実現している。単純な外れ値処理に基づく手法と比較して、提案手法の自己遮蔽に対する頑健性が示されている。(以上、岡部)

O9: Tracking

このセッションでは以下の6件の発表があった。Salzmann(EPFL) らは、連続的に変形する3次元の表面形状に対して、あらかじめ定義された2Dのサーフェスモデルをあてはめて追跡する手法を提案した。この手法では、従来から用いられてきた表面形状の滑らかさに関する制約を用いず、フレーム間での高速なトラッキングを実現することで、曲率が不連続となるような複雑な形状に対するモデルの変形あてはめを実現している。

Joshi(UCSD) らは、マルチカメラを用いることで、70%程度のオクルージョンが発生する状況においても、対象物体のトラッキングを実現する手法を提案した。この手法では、従来から用いられてきた synthetic aperture 法をトラッキングに応用している。具体的には、マルチカメラによって異なる位置から同期撮影された複数の画像を合成することで仮想的に被写界深度が非常に狭い画像を生成すれば、トラッキング対象以外がぼけてしまうことに着目し、オクルージョンに頑健なトラッキングを実現している。

Zhao(UC Santa Cruz) らは、Differential Earth Mover's Distance と呼ばれる新たな評価尺度を用いた照明変化に頑健なトラッキング手法を提案した。この手法では、従来提案されていた Earth Mover's Distance の弱点であった計算コストの問題を解決し、30fps でのトラッキングを実現している。

Zang(Inst. Automation, China) らは、物体の追跡問題について、追跡領域の輝度値の時間変化がお

よそ Gaussian で近似できるのに対して背景領域は多様に変化することに着目し、追跡対象と背景をオンラインで分離・学習しながら追跡を行う手法を提案した。

Sharma(Ohio State Univ.) らは、事前に与えられた人物の画像とそれ以外の画像による学習結果を用いて、人物の位置とその人物のシルエット形状を画像中から同時に抽出する手法を提案した。この手法では、学習データに基づくトップダウンな情報と、シルエット形状の尤もらしさに基づくボトムアップな情報を、Markov Random Field を用いて組み合わせ、抽出の精度を向上させている。

Lee(Rutgers Univ.) らは、関節モデルで表現される人体姿勢とシルエットを関連づけるモデルを用いて、人物領域を追跡する手法を提案した。

(以上、佐藤(智))

O10: Learning II

このセッションでは以下の3件の発表があった。Huang(Rutgers) らは、カメラで撮影した動画像において、追跡対象となる物体上のテクスチャが連続的に変化している場合においても頑健なトラッキングを実現する手法を提案した。この手法では、追跡によって特定のフレーム上に位置合わせされた複数の画像から平均画像を生成し、その平均画像の輝度勾配の分布が元の画像と同様の特性を持つように、追跡結果を最適化している。

Tsin(Siemens) らは、動画像中における三次元エッジモデルのあてはめ・追跡問題について、オンラインでの学習によって追跡のロバスト性を向上させる手法を提案した。この手法では、エッジ周辺の輝度パターンの変化を randomized forest と呼ばれる複数の木構造によってオンラインで学習・分類することで、照明条件の変化やカメラの運動によるエッジ周辺のパターンの変化に対してロバストな追跡を実現している。

Tuytelaars(K.U.Leuven) らは、多次元の特徴ベクトル空間を一樣な格子で分割するという非常に単純な方法が、学習および分類を行う上で効率的かつ効果的であると主張している。この手法では、通常このような格子による分割で問題となる bin 数の発散をハッシュにより解決し、学習データに依存しない均一な格子により特徴空間を分割することで、従来手法に対して画素単位での物体認識の精度を向上させている。(以上、佐藤(智))

O11 Calibration and Localization

Tardif (Univ. Montreal) らは平面シーンを撮影した複数枚の画像からレンズ歪を除去し、カメラの自己校正を行う手法を提案している。この手法では、レンズ歪モデルを画像中の歪中心まわりに対称であるとし、光学中心を一致させて撮影した画像間でのホモグラフィを利用し、最適化計算を凸計画問題に帰着させている。ただし、撮影シーンに大きな平面が写っていること、光学中心が一致していること（推定すべき外部パラメータが回転のみであること）および、ひずみ中心が既知であることを前提としており、自己校正法としての汎用性が高いとは言えない。

Chandraker (UCSD) らはカメラの自己校正において、最適化計算に緩和法を用いた手法を提案している。射影復元からアフィン復元にアップグレードする際の modulus constraint と、アフィン復元からメトリック復元にアップグレードする際の DIAC の最適化計算とを、半正定値計画問題に緩和して大局解を求めている。これまでの最適化計算では、目的関数そのものを最小化しており、局所解に陥るといふ問題点があった。そのため、大局解を効率良く求めるため、非線形計算を凸計画問題に緩和する手法が最近多くみられるようになってきている。

Williams (Oxford Univ.) らは単眼カメラを自由に動かして画像をリアルタイムでトラッキングする際、ブレ、急激なカメラ運動、オクルージョン等で発生するトラッキングの失敗からの回復手法について提案した。特徴点追跡に失敗する前までに学習した画像パッチと、Randomised List を用いて高速にマッチングを行った。特徴点抽出、マッチング、カメラ位置・姿勢の復元までに要する時間は 1 フレームあたり 19ms に抑えられており、リアルタイムで処理できることを示している。

Moreno-Noguer (Lausanne 工大) らは PnP 問題を非繰り返し法によって、 n のオーダーで推定する手法を提案した。この手法では、3次元点を 4つの制御点の重み付き和による表現に変換することによって、点数 n のオーダーでの計算時間を可能にしている。繰り返し手法と同程度の精度を達成するための計算時間が、およそ 20 分の 1 に短縮された。ただし、結果を見る限り、絶対的な精度では繰り返し法の方が良いようである。 (以上、阪野)

3 ポスターセッション

P1: Learning and Recognition I

Kapoor (Microsoft Research) らは、一般物体認識における学習画像のラベリングの労力を軽減するために、確率モデルに基づく能動学習を提案している。ガウス過程により推定されるラベルの不確定性を手掛かりにしてラベル付けすべき画像を決定している点が特長である。教師付きガウス過程や SVM と比較して、少数のラベルを与えるだけでも良い識別性能が得られることが報告されている。

Simon (Univ. Washington) らは、インターネット上に大量に存在する画像から、観光地などのシーンの要約を作成する手法を提案している。具体的には、SIFT 特徴の共起に基づいて画像の類似度を定義して、シーンを代表する画像セットをクラスタリングにより抽出している。また、画像に付随するノイズを含むラベルから、シーンのラベルを推定している。

Li (UIUC) らは、画像からイベントを識別する手法を提案している。局所特徴に基づくシーンの識別と局所特徴とレイアウトに基づく物体の識別を、生成モデルの枠組で統合している点が特長である。8種類のスポーツイベントからなる画像セットを作成して、シーンと物体の両方の情報を組合せることの有効性を示している。

Omercevic (Ljubljana) らは、高次元特徴量の最近傍決定則に基づく画像マッチング法を提案している。“意味のある”最近傍パターンは、距離の増加に伴い急速に数の増える背景パターンの分布に対して外れ値になることに着目している。また、高速化のための近似アルゴリズムも提案している。

(以上、岡部)

P2: Low Level Vision & Applications

Chum (University of Oxford) らは、画像を検索鍵とした大規模画像検索の性能を向上させる方法を提案している。大規模検索では通常、最初の検索で得られた関連性の高い結果画像を、鍵として再度利用することで検索漏れを防ぐが、同時に結果の正確さが損なわれる。著者らは、最初の鍵と、次の鍵の間の幾何的な拘束を考慮し、また関連のある画像間に共通する隠れた特徴量を抽出することにより、検索精度を向上させている。

Hartley (Australian National University) と

Kahl(Lund University) は、現在用いられている最適化手法で大域最適化が可能な問題集合に関して考察している。特に、 L_∞ ノルムを用いた essential 行列の推定問題が、大域最適化可能であることを証明している。

Khan(University of Central Florida) らは、平面間のホモグラフィだけを用いて Shape from Silhouette により形状復元する手法を提案している。具体的には、3次元空間を走査する平面を考え、全入力画像をその平面にワープさせて平面上の積集合を計算している。

Hasinoff と Kutulakos(University of Toronto) は、カメラの開口を変化させて計測した複数の画像から、defocus,dynamic range, ノイズを同時に解析する方法を提案している。対象となるシーンを深度に応じて平面分割して occlusion を検出し、最低3つの異なる開口画像から、計測後に画像の focus を変更する手法について述べている。

Koppal と Narasimhan(CMU) は、未校正の近接点光源下で計測された画像列から、対象物の形状を復元する方法を提案している。光源と対象物表面の距離に応じて反射光が低減することを利用し、反射が最大となる時に光源と表面の距離が最短であることから対象物の形状に関する拘束を導いている。

Yamazaki(産総研) らは、点光源下で物体が作る投影像から、Visual Hull を計算する方法を提案している。物体やカメラを固定のまま、大量の投影像を計測できるため、従来不可能であった微細で複雑な形状物を計測することができる。また投影像の epipolar 拘束を用いることにより、従来の校正方法と比較してはるかに高精度の形状復元を実現している。

Åström(Australian National University) らは、1次元の自由度を持つカメラで計測した画像列から、最適に Structure from Motion を行う方法を提案している。著者らは L_∞ ノルムを用いて、dynamic programming を適用することにより再投影誤差を大域的に最小化できることを証明している。

Yuan(Hong Kong University) らは、高速シャッター撮影によって得られた blur の無い画像と、低速撮影で得られた blur を含む画像の位置あわせについて述べている。カメラの blur カーネルが疎であることを事前知識として、ピラミッド型のマッチングを用いて両画像の位置あわせを実現している。

(以上、山崎)

P3: Tracking & Segmentation

Hernandez(Cambridge) らは、異なる3地点に設置されたRGBの光源と1台のカメラを用いて、テクスチャのない服の変形を追跡する手法を提案した。この手法では、入力画像に対する密な法線マップの推定と、連続フレームに対するオプティカルフローの算出を行っている。ポスター会場では、推定結果のビデオ映像が繰り返し再生されていたが、非常に完成度が高く常に人だかりができていた。

Li(Simens) らは、三次元のワイヤフレームモデルを動画像上にあてはめて追跡する問題について、遮蔽境界のエッジを追跡に用いる手法を提案した。この手法では、従来手法においてアウトライヤとして扱われてきた遮蔽境界のエッジを積極的に利用することで、トラッキングの精度を向上させている。

Wu(Univ. Maryland) らは、動画像上での特徴点の追跡問題に対して、時間軸に対する順方向および逆方向の追跡結果の整合性を同時に考慮した手法を提案した。提案手法では、このアプローチをKLTトラッカに組み込み、オリジナルのKLTとの比較を行っているが、おおむねKLTと同等の処理時間を保ちつつ、追跡精度を大幅に改善できるとしている。

(以上、佐藤(智))

P4: Cameras, Stereo, Pose Estimation, Video & Shape

40件の発表の傾向を無理矢理言うとするれば、3次元的な幾何学問題を取り扱うものが多かったという事であるが、セッション名の長さから分かるように発表内容の幅は広く、カメラ自体の開発、マルチカメラ、イベント検出、3次元復元、形状解析など多岐にわたっていた。これらの中から、特に問題設定に目新しさが感じられた2件をピックアップして紹介する。

1件目は Jacobs(Washington University) らの研究で、地球規模で分散している屋外用ネットワークカメラの自己位置推定である。Jacobs らは、カメラの位置推定手法として、太陽光の照射強度マップを用いる手法と衛星写真を用いる手法の2つを提案している。基本的には、両手法ともカメラが屋外に固定されていることを前提としており、ネットワークカメラにより得られる画像の明るさと太陽光の照射強度マップや衛星写真の各画素の輝度変化を比較することにより地球上の位置を特定する。太陽光の照射強度マップを利用する場合は、地球上の太陽光に

よる明るさが周期的に変化することを利用しており、赤道から離れるにしたがって昼夜の時間の差が大きくなることから緯度が分かり、周期変化の位相から経度が分かる。衛星写真を利用する場合は、これに加えて気象変化に伴う明度変化を捕らえられることから、より高い精度でカメラの位置が推定できる。

2件目は Wolf(Tel-Aviv University) らによる研究で、動画像の Retarget 手法である。動画像の Retarget とは、入力動画像中の撮影対象物など重要な領域を残し、ワーピングにより小さいサイズの動画像を生成することである。Wolf らは、重要な領域を判定するために、画像中の各画素位置の勾配、顔検出位置、動物体検出位置を特徴として用いている。ワーピングでは、判定された重要度が高ければ画素の間隔を広く、低ければ狭くなるように画像を変換する。ワーピング後のアスペクト比をコントロールすることもでき、携帯電話のテレビ機能や動画のサムネイル等に應用できると考えられる。

(以上、池田)

P5: Learning and Recognition II

Yan (Central Florida 大) らは、3次元の物体モデルを用いて、与えられた画像中の物体のクラス分類を行った。画像中から SIFT によって抽出した特徴を、候補となるクラスの 3D モデルに投影することによって、クラス分類を行っている。

Nowozin (Max Plank Institute) らは、画像列中の人物の動作識別において、動作の出現する順序を考慮した表現方法と識別方法を提案している。この手法では、LPBoost を組み合わせることによって、識別関数の学習と特徴空間からの特徴抽出を行い、画像の部分列から動作の識別をしている。

Nedović (Amsterdam 大) らは、単画像の構図(たとえば地面と空、テーブルと人物と壁、地面と壁など)の分類を行った。あらかじめ、分類すべき構図のテンプレートをもっておき、入力画像から特徴を抽出し SVM によって画像構図を識別している。ただし、平均の識別率は 50% を割っており、現時点では有効に動作するとは言い難い。

Zhang (Stony Brook 大) らは画像中の物体認識をリアルタイムに行うための手法を提案した。人物、自動車、バイクを抽出するため、多重解像空間からカスケード型に候補となるウィンドウを破棄していき、最終的には SVM で対象物体を検出している。320×240 のサイズの画像について、既存手法が

3~4fps であるのに対し、提案手法は 25fps を達成している。

Zhang (復旦大) らは、画像を低次元空間に写して、効率的に分類していく手法を提案した。画像識別のプロセスをグラフ構造とし、各エッジにおいて線形識別を利用して分類していく。このとき、各識別子で、クラス内をコンパクト化しクラス間の距離を大きくするコスト関数を設定し、学習させている。実験結果では、フィッシャー判別分析と同程度の精度であることが示されている。

Ling (Princeton 大) らは、過去の顔写真と現在の顔写真とが同一人物であるかを判別する手法、結果について報告した。これはパスポートの写真が、本人であるか否かを判別することを目的としており、顔写真の勾配から特徴ベクトルを算出し、SVM で判別する手法をとっている。取り立てて目新しい方法ではないが、着眼点がユニークであり、4年以上経過した写真の判別が難しくなる等、示唆的な結果が得られていた。

(以上、阪野)

P6: Registration, Shape, Illumination, Texture, Segmentation

このセッションは、その名前が示すように、様々なジャンルの合計 39 件の発表があった。その中から、特に聴衆者の多かった色に関する発表を 2 件紹介したい。Park(Hanyang Univ.) らは、分光分布の異なる LED 照明を用いて、ビデオレートでのマルチスペクトラル映像の生成を可能とした。2種類の照明下で RGB カメラ (60fps) で撮影した情報から、多数の分光分布を主成分分析して得られた直交基底ベクトルの係数を決定することで、30fps のマルチスペクトラル映像を得るという比較的シンプルな方法である。分光分布の異なる 5 種類の LED を、S/N 比がもっとも高くなるように組み合わせ、2種類の照明を決定したところに工夫がある。

Laolonde と Efros (CMU) は、色の整合性によって画像の写実性を評価する方法を提案した。ある画像から切り出した物体を他の画像に合成すると、人間は色に関する不整合を感じることもある。これを定量的に評価するために、人間が自然か不自然かを分類した大量の合成画像を用いて、その色分布を最近傍法で学習した。さらに、不自然な画像の色分布を自動的に修正することで、合成画像を自然な見え方に変換する方法についても提案している。

(以上、向川)

4 おわりに

第 11 回 コンピュータビジョン国際会議 (ICCV2007) 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、プロシーディングスを参照して頂ければ幸いである。

次回の ICCV2009 は松山隆司教授 (京大) が General Chair となり、2009 年 9 月 29 日～10 月 2 日の予定で国立京都国際会館において開催されることになっている。ブラジルは日本からみて地球の反対側と非常に遠かったこともあり、残念ながら ICCV2007 では日本からの発表はあまり多くなかった。久しぶりの日本開催となる ICCV2009 では少しでも多くの論文が日本から発表されることを期待したい。最後に、ICCV2007 で開かれた PAMI-TC Meeting での投票の結果、ICCV2011 の開催地が僅差でスペインのバルセロナに決まったことを記して本報告を結ぶ。

(以上、佐藤 (洋))