

## CVPR2009 報告

岡谷 貴之<sup>†1</sup> 石川 博<sup>†2</sup> 大塚 和弘<sup>†3</sup>  
齋藤 英雄<sup>†4</sup> 日浦 慎作<sup>†5</sup> 勞 世紅<sup>†6</sup>

2009年6月22日~24日に米国フロリダ州マイアミで開催された国際会議 CVPR2009 の概要を報告する。

### Report on CVPR2009

TAKAYUKI OKATANI,<sup>†1</sup> HIROSHI ISHIKAWA,<sup>†2</sup>  
KAZUHIRO OTSUKA,<sup>†3</sup> HIDEO SAITO,<sup>†4</sup> SHINSAKU HIURA<sup>†5</sup>  
and SHIHONG LAO<sup>†6</sup>

This report gives an overview of 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2009) which was held in Miami Beach, Florida, USA from June 22th to 24th, 2009.

†1 東北大学

Tohoku University

†2 名古屋市立大学

Nagoya City University

†3 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories, NTT Corporation

†4 慶応大学

Keio University

†5 大阪大学

Osaka University

†6 オムロン株式会社

OMRON Corporation

### はじめに

CVPR は、ICCV とともにこの分野を代表する国際会議である<sup>\*1</sup>。CVPR の正式名称は、IEEE Computer Society Conference on Computer Vision and Pattern Recognition であり、欧州の ECCV やアジアの ACCV に相当する北米の地域会議として位置付けるのが一般的だと思われるが、欧州はもちろんアジアからも多数の参加者があり、国際色は豊かである。

フロリダ州マイアミで開催された CVPR2009 もその例外ではなかった。CVPR 本体は6月22日から24日の3日間、前後ワークショップを含めると20日から25日までの6日間の日程で、Miami Beach にある Fontainebleau Resort というホテルで開催された。General Chairs に Daniel Huttenlocher (Cornell 大)、Gerard Medioni (Southern California 大)、James Rehg (Georgia Tech.) 各教授、Program Chairs に Irfan Essa (Georgia Tech.)、Sing Bing Kang (Microsoft Research)、Marc Pollefeys (ETH Zurich/North Carolina 大) 各教授・博士という組織構成で運営された。参加登録者数は全 1,110 名 (CVPR2008 は 1,253 名) と発表されている (うち学生参加者は 39%)。

論文の投稿数は全 1,464 本であり、うち 61 本 (4.1%) がオーラル発表として、322 本 (22.1%) がポスター発表として採択されている (両発表を合わせた採択率は 26.2%)。参考までに過去 4 年の CVPR (2005-2008) での採択率を記すと、それぞれ 27.9、28.1、28.2、31.9%であった。なお、論文投稿時に著者が選択する “Primary Subject Area” ごとの投稿数、ポスターおよびオーラル各セッションの採択数と率が公表されているので、表 1 に示す。

査読から採否の決定までのプロセスは、例年同様に入念なものである。まず、double blind 方式で、論文 1 本につき 3 名の査読者が査読を行う。その査読結果に対して、著者は反論の機会を 1 度与えられる。その反論内容は査読者側に通知されるとともに、査読者間で結果に整合性がない場合、査読者同士で意見交換を求められることもある。これらすべてに基づいて採否は決定されるが、その決定は常に 2 名以上の Area Chairs の確認を経て行われたということである。

\*1 Computer Science Conference ランキング ([www.cs-conference-ranking.org](http://www.cs-conference-ranking.org)) 2009 年 2 月版によれば、Topics II のカテゴリー (Artificial Intelligence/Machine Learning/Robotics/Human Computer Interaction) において、ICCV、CVPR は (701 会議中) 4,5 位にランキングされている。ちなみに 1-3 位は、American Association for AI National Conference(AAAI)、Neural Information Processing Systems(NIPS)、Intl Joint Conf on AI(IJCAI)。

表 1 分野別の採択数

分野 (Primary Subject Area)	投稿数	ポスター数	オーラル数	採択率
Applications	85	15	1	19
Biometrics	22	4	0	18
Camera Calibration	22	2	0	9
Color and Texture	26	6	1	27
Computational Photography and Video	33	11	2	39
Early and Biologically-inspired Vision	26	5	1	23
Face and Gesture	109	24	5	27
Illumination and Reflectance Modelling	12	5	1	50
Image/Video Search and Retrieval	62	12	4	26
Image-Based Modelling	35	9	1	29
Matching and Recognition	76	18	5	30
Medical Image Analysis	62	15	2	27
Motion and Tracking	141	19	8	19
Multi-view Geometry	19	8	1	47
Object Detection and Recognition	162	34	7	25
Performance Evaluation	10	3	0	30
Range Image Analysis	8	3	0	38
Segmentation and Grouping	116	23	3	22
Sensors	12	1	0	8
Shape Representation and Matching	44	14	2	36
Shape-from-X	36	10	3	36
Statistical Methods and Learning	154	45	5	32
Stereo	31	6	3	29
Structure from Motion	25	8	1	36
Video Analysis for Activity Recognition	75	19	4	31
Vision for Graphics	12	1	0	8
Vision for HCI	9	0	0	0
Vision for Robotics	25	2	1	12

セッション構成は、17 のオーラルセッションと 6 つのポスターセッションからなる。ポスターセッションは 3 日間を通じて、午前と午後の 1 日 2 セッション行われた。オーラルセッションは 17 のうち 16 セッションがパラレル、1 セッションは特別なセッションでシングルトラックで行われ、受賞論文を含む 4 本の論文が発表された。これらに加えて、最近博士課程を卒業したかあるいは卒業間近の学生によるポスター採択論文について、短い発表の機会を与える Doctoral Spotlight セッションも行われた。

論文賞各賞と受賞論文については、後の “Oral Session 11: Paper Awards” を参照されたい。

以下、オーラルセッションでの発表論文を中心に、その概要を報告する。

(文責：岡谷貴之)

## Oral Session 1: Image and Video Search

Image and Video Search のセッションでは 4 本の論文があった。その中で、1 件は人の姿勢情報を使った検索手法で、そのほかの 3 件は従来の “bag-of-words” 手法を改良して物体の幾何形状情報を活用できるような手法を提案している。チェコ工科大学の Matas の研究室からは 2 件の発表があった。

Ferrari (ETHZ) らはビデオから特定の姿勢の人の映像を検索する手法を提案している。画像から特定姿勢の人を検索には Pictorial Structure モデルを使って、頭部、胴体および手の位置を検出し、それらの位置関係を検出する。また HOG を使って学習した姿勢検出器も使う。

Ped'och (Czech TU) らは、従来の局所領域データを使った検索アルゴリズムには大量のメモリが必要という課題に対して、楕円空間で変換するとき、射影誤差を最小化するように変換することで、1 つの特徴量を 24 ビットで表現しても、同等の性能を達成できると報告している。

Czeth TU の Chum らからのもう一件の発表は Geometric min-Hashing を使った物体探索手法である。Geometric min-Hashing は min-Hashing を拡張した手法で、min-Hashing より高い精度が期待できる。その有効性を確認するために、大量画像の中の小さい物体を検出するのに有効であることを示した。また、オクルージョンや大きさの変化にロバストな手法である。

Wu ら (MSR) は部分的に複製や編集した画像を検索するのに有効な手法を提案している。MSER で検出された局所的な領域の中に存在する複数の SIFT 特徴量をまとめて複合的な特徴量を作ることでオクルージョンや大きさの変化などに対応した画像探索を実現した。(担当：勞世紅)

## Oral Session 2: Optical Flow and Image Registration

Wu (Northwestern Univ) らは、空間的な文脈を制約条件とした頑健なオプティカルフロー推定法を提案した。ある画素の文脈を近傍の色やエッジ方向などの特徴の事後確率として表現し、その空間勾配・時間差分を用いて文脈の不変性に基づくオプティカルフロー拘束式を定義し、複数の文脈による過剰決定方程式系の最小二乗解としてフローを計算する。

Brox (UCB) らは、大きな変位にも対応可能なオプティカルフロー推定法として、エネルギー

ギー最小化と記述子照合の各々の利点を統合した方法を提案した。まず、階層的画像分割により得られたパッチ間の対応付けにより仮説を生成し、その仮説に対応したフローの制約条件と既存の制約条件とを組み合わせたエネルギー関数の最小化によりフローを推定する。小さい物体の大きな変位を含む密なフロー場を高精度に推定することができる。

Myronenko (Oregon Health and Science Univ.) らは、正確な画像位置あわせのための類似性尺度として、従来、前提とされていた輝度の空間的独立性・定常性を排し、輝度の非定常な空間歪みを反映した新しい類似性尺度を提案した。

Roberts (Georgia Tech) らは、移動カメラから得られる疎なオプティカルフローを入力とし、そこから直接、密なオプティカルフローと自己運動推定を頑健に行う方法を提案した。部分空間拘束に基づくフロー推定法を一般撮像系に拡張し、シーン幾何の統計的規則性を捉える線形部分空間を確率的 PCA を用いて学習している。

(担当：大塚和弘)

### Oral Session 3: Stereo

Smith らは、各画素の位置、色、テクスチャ、フィルタバンクへの応答等を特徴ベクトルとし、画像をこの特徴ベクトル空間内での点群として扱うことにより、ステレオマッチングを 2 点群間のマッチングとして捉える手法を提案している。点群を特徴ベクトル空間における距離によって比較的疎なグラフにすることにより、自然に階層的セグメンテーションの構造を生成し、事前分布としてカーネル密度推定によって色などの画像特徴と奥行きを相関させたものを使い、グラフカットでエネルギーを最小化している。

Sizintsev らはステレオ画像の時系列をマッチングする動画像ステレオ問題において、3 次元時空内における各点で方向毎のガウス関数 2 階微分フィルタとそのヒルベルト変換に対する 3 次元時空画像の応答を求め、これをコンパクトな特徴とするため 2 次曲面として近似し、それらをマッチングしている。

Bleyer らはステレオにおける各視点からの元画像において、奥行きの不連続な点における画素では、奥行きが異なるがその視点からは隣接して見える点の色が混ざり、この混ざり方が視点毎に異なることが、不連続点におけるマッチングの不正確さを招くことを指摘している。またこの問題を解消するため、各視点からの元画像をまず色によってアルファ層に分解し、それらをマッチングすることにより、ステレオペアから新視点画像を合成する方法を提案している。

(担当：石川 博)

### Oral Session 4: Image Enhancement and Restoration

Lu らの論文は、手持ちカメラで撮影した露光パラメータの異なる複数枚の画像から、ハイダイナミックレンジ (HDR) 画像を合成する手法に関するものである。手持ちカメラであることにより画像同士の位置ずれだけでなく、露光時間の長い画像では手ぶれが発生する。そこで最尤推定法に基づきブラーカーネル推定、HDR 再構成、カメラレスポンス関数推定を繰り返し、これら 3 者を求める。

Banerjee らは激しく劣化した文書画像を復元するために、MRF の学習に基づく手法を提案している。学習により各文字に対応したプロトタイプが生成され、出力結果は各文字をプロトタイプに置き換えたような結果となる。このため超解像も同時に達成されるが、修復誤りは OCR における誤認識のように全く別の文字として現れる (論文では、本来  $v$  の文字であったところが  $a$  になっている例が示されていた) が、OCR とは異なり言語 (文字種) を問わず適用可能であるところがおもしろい。

空気中の霞などによる散乱光は部分偏光しているため、これを偏光フィルタにより軽減することは写真撮影において広く用いられている。しかしこのことは被写体の SN 比改善に有効であるとは限らない、という基本的な指摘が Treibitz によりなされた。偏光フィルタは霞の向こうの被写体からの光も減ってしまうため、露光時間が余計にかかってしまう。その露光時間中に偏光フィルタを用いず複数の画像を撮影し、合成すれば、実際の散乱光の偏光比を考えると SN は後者の方が良いとの結論である。もっとも被写体輝度と散乱光の比率を知ることが出来ないが、これに対する回答は同会議の Best Paper が与えているので是非併せて参照されたい。

(担当：日浦慎作)

### Oral Session 5: Statistical Methods and Learning

Lampert ら (Max Planck Inst.) は、一般物体認識において、学習サンプルのない物体クラスをどうやって認識するかという問いに取り組んでいる。学習サンプルの代わりに、物体の各クラスに属性を与え、その属性の項目が全クラス間で同一であるような状況を考える。彼らは、50 の動物に対し、85 項目の属性 (例えば「白い」、「茶色い」、「水」、「魚を食べる」など) を定義したデータを実験で使用している。このとき、クラスの集合を 2 つにわけ、一方には学習サンプルがあり、残りにはそれがないという状況のとき、ない方のクラスに属するテスト画像を見せられたとき、それを認識する方法を提案している。前

者のクラスの学習サンプルを使って (学習画像) → (属性) の関係を学習し, その結果から (未知のクラスのテスト画像) → (属性) を計算し, 属性からクラスラベルを求めるという方法である.

岡谷ら (東北大) は, SFM のような, CV で良く扱われる画像から幾何学量を推定する問題に対し, 最尤推定 (あるいはそれを根拠とするバンドル調整) を適用したとき, 推定量が持つ偏差について述べている. エピポーラ条件などの幾何学的拘束が観測空間において超曲面を与えるとき, 最尤推定量の偏差が, 超曲面の曲率と誤差分散の大小関係によって決まることを述べ, その補正方法を示した.

Saffari ら (Graz Univ.) は, 複数クラス分類のための半教師ありブースティングの方法を提案している. 複数クラス教師あり学習を説明する損失関数に, ラベル無サンプルについてのクラスタ正則化項およびマージン最大化損失 (項) を追加し, これを最適化するブースティングアルゴリズムを提案している. Pascal VOC2006 のデータセットで実験を行い, 従来法に比べて計算効率と精度がともに向上したと報告している.

Munoz ら (CMU) は, 3 次元あるいは 2 次元画像 1 枚が与えられたとき, その各ピクセル (あるいはボクセルや領域) にラベル付けをする問題に対して, 新たな方法を提案している. ピクセル間のインタラクションを  $M^3N$  (Max-margin Markov Network) モデルで表現し, その学習を手でラベル付けしたデータによって行う際に (現在の標準手法である parametric gradient 法ではなく) functional gradient 法を用いる方法を示した. 実装が簡単で, 2 次や線形のソルバーを必要とせず, 要求される記憶域が小さいため, 大規模データを使った学習が可能となる.

(担当: 岡谷貴之)

### Oral Session 6: Tracking

Babenko (UCSD) らは, 適応アピランスモデル学習による物体追跡法を提案した. これは「検出による追跡」の一つであり, 物体と背景を分離する識別器をオンラインで学習するために, 従来の教師あり学習に代わり, 多事例学習を用いており, それによりラベルの不正確性に起因する識別器の性能低下を防ぎ, より頑健な実時間追跡を可能としている.

Kwon (ソウル国立大) らは, 対象追跡の方法として, 物体テンプレートの 2 次元アフィン運動をパーティクルフィルタにより推定する方法を提案した. 2 次元アフィン変換群  $Aff(2)$  上での観測関数のテイラー展開により, 幾何的に最適された重要関数 (Importance Function) を導出することで, 追跡の効率と性能の向上を図っている.

NTT の三上ら (登壇は大塚) は, 対象追跡のためメモリーベースパーティクルフィルタを提案した. 非マルコフ性ダイナミクスを反映した事前分布予測を狙いとして, 過去の状態履歴からの時間的サンプリングにより事前分布生成を行うという新しい定式化を提案している. 急激な運動方向変化や追跡失敗からの復帰において顕著な効果を確認している.

Mahadevan (UCSD) らは, 対象と背景との判別による追跡 (discriminant tracking) のため, 生物の視覚系を参考とした中心周辺顕著性の原理に基づいて, 初期化, 特徴選択, 対象検出の 3 つの問題を統合した枠組みを提案した. トップダウン顕著性により特徴選択と対象検出が行われ, ボトムアップ顕著性により初期化が行われる.

(担当: 大塚和弘)

### Oral Session 7: Object Detection & Recognition

このセッションでは, 4 本の発表があった. 今年は Pictorial Structure や Hough 変換のような従来手法を再考の論文が多いのが一つの特徴.

Andriluka ら (TU Darmstadt) は pictorial structures のモデルを使った一般姿勢の人体検出と人体姿勢推定システムを提案している. 従来の人体検出の多くは, 歩行者や成立している人を対象にしている. この研究は人体の各部位を検出する検出器を学習することと, 人体各部位の位置関係を学習してあるので, 直立してなくても, 人の検出ができる. 同時に, その人の姿勢も推定できる. 実験では歩行者の人体検出, 上半身の姿勢推定, そしてダンスや体操をしているような全身の自由な姿勢の推定を行い, 従来手法よりも良い性能を得ている.

Gall ら (ETHZ) は Hough Forest による物体検出法を提案している. コードブックを使った従来の Hough 変換の手法と違って, Random Forest のフレームワークを使ってパーツのアピランスを Hough 変換の票にマッピングする方法を使う. これによって検出性能と速度両方が従来の手法より改善されることが報告された.

UC Berkeley から 2 本の論文が発表された. Gu らの Recognition using Regions は画像領域を使って検出とセグメンテーションを行うフレームワークを提案している. 一般的には正確なセグメンテーションが難しいため, 領域を物体検出の有効な特徴量として使うのが難しい. 本研究では, 領域の輪郭, エッジ, 色, テクスチャーを使って, 木構造の "bag-of-regions" で領域を表す. 物体を検出する時は Hough 変換で投票した後, 識別器を使って確認の上, 更にセグメンテーションを行う. Maji らの研究も Hough 変換による物体検出手法である. マージン最大化のフレームワークを使って, Hough 変換の投票に最適な重みを学習するこ

とで、検出器の性能を高める手法を提案している。

(担当：勞世紅)

### Oral Session 8: Texture, Symmetry, and Shape

人が歩いたときの足跡のように、ある軸に沿って左右対称であるが、その位相がずれているような図形を Glide Reflection と言う。さらにこの対称軸が湾曲しているような場合について、その対称軸を画像から見つけ出す方法を Lee らが提案している。対応点の検出には SIFT が用いられている。

Salzmann らは、部分的にオクルージョンが発生するほどきつく曲げられた面の変形を復元する手法を提案している。面は曲がるだけで伸縮しない(測地線距離は変化しない)という仮定が従来用いられていたが、著者らの方法ではこれを「縮みはするが伸びはしない」という制約に置き換えることで凸問題として定式化出来ることを示している。

Chan らは動画画像から異なるダイナミクスを持つ領域をセグメンテーションする手法として、変分近似に基づく手法を提案した。彼らはまた、セグメンテーションされた領域が時間的に変化するモデルも提案している。

表面における拡散反射成分が観測されにくい鏡面物体や透明物体の形状を計測する手法として、Johnson らは透明エラストマー(ゴム状の弾力性を有する透明材料)の表面に塗装を施した器具を用いることを提案している。計測対象をこの塗装面に押しつけ、塗装面の反対側から透明エラストマーを観察すると、形状は計測対象そのものだが反射特性は塗装面のものになる。形状計測は Photometric Stereo の原理によるが、反射特性が既知であることから安定な計測を行うことが出来る。指紋や紙幣印刷の凹凸なども良好に計測でき、彼らが CVPR において行ったデモ展示は Best Demo Award を授与された(SIGGRAPH および ICCV でも展示があった)。

(担当：日浦慎作)

### Oral Session 9: Face Recognition

このセクションには3本の発表があった。顔の向きや表情変化にロバストなアルゴリズムが発表されている。

Wright と Hua (MS) の論文は、従来の顔認識手法は顔特徴点の検出精度に大きく依存しているため、特徴点の位置検出に誤差がある場合の認識性能に限界があるという課題に注目した。そのための対策として、特徴点の位置がずれた場合でも影響が受けにくい認識手法

を提案した。密に抽出された局所特徴量に位置座標を合わせた高次元のデータに対してランダムプロジェクトを使って量子化できたヒストグラムでのマッチングは間接的に顔画像の Elastic Matching を実現できるので、特徴点の検出ずれや向きのずれに対するロバスト性が向上できるアイデアである。

Li と Prince (Univ College London) の論文も Wright の論文と同じ、特徴点の位置が正確に検出できない課題への対策に関する論文である。従来の顔認識アルゴリズムは特徴点を検出してから、認識を行う2つのステップに分かれている。本論文は、それらを一体化して考える。まず、登録画像の助けを得て、テスト画像の中の特徴点の検出精度を高める方法を検討した。更に、特徴点の位置を明示的に求めず、特徴点の位置を隠れ変数として扱う確率的な認識手法を提案している。顔の向きが大きくずれた場合の認識性能を向上させた。

Nagesh と Li (Arizona Univ) は Compressive Sensing 理論を顔認識に応用し、表情変化にロバストな顔認識のアルゴリズムを提案している。同じ人の異なる表情の複数の画像から、1) 個人の共通性を含む特徴画像と2) 表情変化を構成する特徴画像の2枚の画像に集約することができる。更に大幅な次元削減も可能であること示した。これによって、顔認識のデータ量を大幅に減らすことができる。実験結果では、近年注目されている sparse representation に基づく顔認識よりも良い性能を得ている。

(担当：勞世紅)

### Oral Session 10: SFM and Geometry

Ponce は "What is a Camera?" と題した講演で、線形性を満たすカメラの一般幾何モデルについて数理的な考察を発表した。現在用いられているほとんどのカメラは一般に透視投影であり、ピンホールカメラモデルを用いて説明できる。このモデルに基づきエピポーラ幾何や多視点幾何などの幾何学的拘束に関する理論が発展を見たが、この論文では逆にこれらの拘束が成立するカメラとしては透視投影のほかにもどのようなものがあるか、またそれらを統一的に表現する数学的フレームワークについて述べられている。

Akhter らは動画画像からの非剛体形状復元 (Nonrigid Structure from Motion) 問題に関する発表を行った。シーン形状を少数の基底形状の線形結合により表すことで、Tomasi-Kanade の因子分解法の自然な拡張として非剛体形状復元を行うことが出来るが、ここで用いられている正規直交制約が解の不定性を生じていると従来考えられてきた。この考え方は正しくなく、全体の回転に関する不定性を除く唯一の解を正規直交制約のみから決定可能であることを Akhter らは指摘している。基底形状とその結合係数の間に不定性はある

が、これは求められた形状に不定性を生じない。なお余談であるが、この発表は CVPR において初めてのパキスタンからの口頭発表であった。インド出身の Raskar 座長が、前々日のクリケットワールドカップでパキスタンが初優勝したこととあわせて紹介し、その荣誉をたたえた一幕には清々しい印象を受けた。

(担当：日浦慎作)

### Oral Session 11: Paper Awards

ここでは、ベストペーパー賞、学生ベストペーパー賞と、それぞれの次点である Honorable Mention を受賞した論文が発表された。

まず、ベストペーパーを受賞した、Single Image Haze Removal Using Dark Channel Prior(The Chinese University of Hong Kong, Microsoft Research Asia) は、1 枚の入力画像から、Haze を除去するという研究である。特に何も事前情報は必要ないが、Dark Channel Prior という考え方を事前知識として利用するところが本論文の重要なアイデアであり、このアイデアをベースにして、Haze を除去することができる手法を提案しているところが受賞のポイントなのだと思う。

この Dark Channel Prior を利用するというアイデアは、単純ではあるが非常に巧妙である。Dark Channel とは、理想的には、光の強度が 0 の点を撮影した画素である。このような点を撮影した画素は、もし Haze が無ければ輝度値が 0 になる。しかし、Haze があると、Haze により乱反射した照明光の影響で輝度値が 0 にはならず、その輝度値そのものが、その画素に対する Haze の度合い(つまり、物体までの距離)を表わすことになるわけである。このことに着目し、本研究では、自然画像においては、ある程度の局所領域には、かならず Dark Channel となっている画素が一つは存在するという仮定をすることにより、各画素毎にその近傍領域内で最も暗い画素値が Dark Channel という想定で、Dark Channel に対する輝度値の分布を作る。Haze が無い画像に対しては、この Dark Channel の画像は画像全体で 0 になり、一方、これを Haze の画像に対して行くと、Haze の度合いの分布となる。これは Haze によって散乱された光の強さの分布ということになり、カメラから物体までの距離によってその明るさが決まっていることになるため、距離画像にもなる。このことを利用して、Haze の強さから距離画像が得られるという副産物付きで、Haze を除去した画像を合成できるのである。

この研究は、プレゼンテーションも非常に良く構成されており、きわめて了解度が高かった。個人的にも、プレゼンテーションでもベスト、内容でもベストペーパーに相応しい論文だ

と思った。

さて、次は Best Paper の Honorable Mention は、Understanding and evaluating blind deconvolution algorithms(MIT 他) に与えられた。これは、ボケのブラインドデコンボリューションアルゴリズムの最近の手法を取り上げ、それらを比較検討したものである。非常に良く整理されたサーベイにもなっているうえ、いろいろな観点からブラインドデコンボリューションを比較したものであり、このトピックを扱う研究者にとって非常に参考になる論文である。

Student Best Paper は、Nonparametric Scene Parsing: Label Transfer via Dense Scene Alignment(MIT) である。これは、一般画像のセグメンテーションの問題を解決する手法の論文である。前提として、事前に正しくセグメンテーションされた画像データベースがあるとす。そして、このデータベースから、入力画像に最も近い画像をデータベースから見つけてくる。この際に、本論文で提案している Coarse-to-Fine SIFT flow という SIFT 特徴量を利用したオプティカルフローベクトルを利用して、できるだけデータベースから入力画像と似たような画像見つけてくる。そして、その見つかったデータベースに存在する画像に事前に与えられているセグメンテーション結果を、Coarse-to-Fine SIFT flow ベクトルを利用して変形する。そうすることによって、入力画像のセグメンテーションが良好に行えるというものである。

画像のセグメンテーションは、撮影されているシーンの内容によって、アルゴリズムのパラメータや設定を変えたいものであるが、それを知るためにセグメンテーションをしようとしているわけであり、そこが難しい。これを解決するために、事前にセグメンテーションされているデータベースを使うというアイデアと、そのデータベースから入力画像と似た画像を見つけてくるための手法に工夫がなされている論文である。

学生ベストペーパーの Honorable Mention は、A Tensor-Based Algorithm for High-Order Graph Matching(Ecole Normale Supérieure, INRIA, KAIST) である。2 つの異なる形状間で、視覚的に似通った点を見つけるための新しい形状特徴量の提案とその対応付け手法が提案されている。

(担当：斎藤英雄)

### Oral Session 12: Video Analysis

Loy ら(Queen Mary Univ. of London) は、複数の、視野を共有しない未校正カメラで撮影した、人の流れの激しいパブリックスペースでの、人々の行動間の相関をモデル化する

方法を提案している。3つのタスクを実行する。カメラネットワークの時空間トポロジーを推定し、よりロバストで正確に個人を同定することを容易にし、視覚的なエビデンスを異なるカメラ間で結びつけることにより、大域的な行動をモデル化し、ビデオの時間セグメンテーションを行う。

Liuら (Univ. of Central Florida) は、これまでのアクション認識の研究が対象としてきたような、撮影条件のコントロールされたビデオではなく、例えば YouTube のような多様な条件下 (カメラの動き、背景クラッタ等) で撮影されたビデオを対象に、人のアクションを認識する方法を提案している。具体的には、映像から動きの特徴量と静的な特徴量を抽出し、2つを組み合わせることで認識性能の向上を狙い、その際、動きの統計量を使って不適切な特徴を取り除いている。また学習の際には、静的特徴量で情報量豊富なものを PageRank アルゴリズムを使って探し当てることと、コンパクトで識別力に優れた vocabulary を見つける情報理論に基づく方法を提案している。

Sunら (National Univ. of Singapore) は、Liuら同様に多様な条件下で撮影されたビデオを対象としたアクション認識を実現するには、特徴量の時空間コンテキストをより有効に活用することが必要だとしている。ビデオ映像上で SIFT の軌跡の集合を抽出し、1) 各軌跡上での濃淡勾配の統計量、2) 各軌跡における SIFT 特徴点の遷移、3) 軌跡間の近さ、をこの順に階層をなす特徴量とする。(2) と (3) に現れるコンテキストの情報を Markov 過程の遷移行列として表す。複数チャネル非線形 SVM を用いてこれら特徴量を融合する。HOHA および LSCOM データを用いた実験で、既存手法を上回る性能を確認した。

Guptaら (Univ. of Maryland) は、ビデオ映像が与えられたとき、そのストーリーライン (人のアクションの因果関係) を認識する問題を扱っている。論文では、野球の中継映像を対象に、ピッチャーが投げ、バッターが打つあるいは打たない、打った場合には野手が走ると同時にバッターも走り、その後野手が打球を捕り... という、個々のアクションが因果関係を持ちながらつながってゆく様子を、グラフとして表現し、そのグラフそのものを学習する方法が提案されている。ビデオ映像と行動を記述したキャプションのペアを学習サンプルとし、その集合から、個々のアクションの見え方とストーリーラインのモデルを同時に学習する。

(岡谷貴之)

### Oral Session 13: Segmentation

Vazquez-Reinaら は、神経細胞断面の電子顕微鏡写真における、複数の等輝度領域が入

れ子状になった特徴的な形状のセグメンテーションのために、複数のレベルセット関数を柔軟に結合して使うことを提案している。

Mukherjeeら は、2つ以上の画像中にある類似した領域の同時セグメンテーションであるコセグメンテーションのために、領域内の輝度のヒストグラムを画像間でマッチングする方法を提案している。そのために、各画像で前景とされた画素のヒストグラムを  $L^2$  距離で比較するエネルギーを定義し、離散ラベル  $\{0, 1\}$  を区間  $[0, 1]$  に連続化して線形計画法を使う場合に、 $L^1$  距離の場合よりも容易に最小化できることを報告している。

Liら は、ユーザによってタグ付けられたウェブ上の画像などのノイズの多い情報から頑健なシーンモデルを学習できる、全自動学習の枠組を提案している。

Jojicら は、CVPR2004 の probabilistic index model (PIM) の一般化を提案している。例えば一連のボールの画像があったとして、このクラスの特徴を学習したいが、いろいろな色のボールが写っているために、例えば各画素の色という特徴を直接比べることができないとする。このようなときに、カラーパレットを使って、ボールの色は全ての画像で同じインデックスになるようにすれば、ボールの形状といった幾何的な特徴を比べられる。同一クラス内の多数の画像上に共通の幾何学的な特徴を見つけられるようにカラーパレットの上の確率分布を学習することにより、色の確率分布を学習する場合より柔軟にクラスの幾何学的特徴を抽出することができる。これが PIM であるが、今回は、同じインデックスを持つ画素の集合を *stel* と呼び、*stel* の確率分布を学習することで、クラス内に共通する画像中の空間的な相関を学習することを可能にしている。

(担当: 石川 博)

### Oral Session 14: Matching & Alignment

3つの発表があった。最初は A Robust Shape Model for Multi-view Car Alignment (CMU) であったが、驚くことにラストオーサの金出先生の発表であった。金出先生の「招待講演」は数知れず聞いたことがあるものの、このような通常の口頭発表を拝見するのは初めてであった。限られた時間の口頭発表だからこそ、出来るだけの情報を伝えるために、事前の準備を完璧に行ってこられたんだな、ということがひしひしと伝わってくる発表であった。この論文では、入力画像に含まれている任意方向から撮影された「自動車」を検出するための、複数の点群とそれらを連結するための線分により表わされる「自動車」モデルの提案と、そのモデルにマッチングをして自動車を検出するための手法が提案されている。画像に撮影されている自動車が一部隠れてしまうことによるマッチングの失敗を防ぐために

RANSAC の考え方を巧みに利用しながら、自動車が撮影されている方向や自動車の種類に依存しないマッチングの仕組みに工夫がある。質疑の時に、計算時間などについての詳細についての質問があったときには、「さっき見せたビデオで紹介したような感じ」という答えしかなく、実際にプログラムを実装して実験した第一著者にしかわからない詳細についての回答を避けておられたのが微笑ましかった。

さて、この CMU の研究のように、画像に撮影されている「自動車」を検出するための研究が、このセッションではないがポスターセッションでも発表されていた。それは、Predicting High Resolution Image Edges with a Generic, Adaptive, 3-D Vehicle Model (Brown University) という研究であり、少数のポリゴンからなる 3 次元メッシュモデルにより「自動車モデル」というのを定義しておき、その自動車モデルのサイズや形状を変化させながら入力画像のエッジとマッチングさせることによって自動車を検出するものである。人間が扱う画像に最も多く撮影されているものとして「顔」や「歩行者」があり、その検出法などが盛んに研究されてきたが、「自動車」も同様に多く撮影されている基本物体の一つである。自動車の場合は、「顔」や「歩行者」と異なり、見る方向による 2 次元画像の変化が極めて大きいので、ここで試みられているように、見え方の変化にも対応できるようなモデル化と検出法が重要であり、これらの 2 つの研究はそのあたりがしっかりと検討された手法になっていると思う。

さて、このセッションの他の 2 つの発表は、特に物体は限定しないが、物体を撮影する方向やサイズの違いに影響されない Invariant な特徴量を独自に提案している物体マッチングの手法についての研究発表であった。

(担当：斎藤英雄)

### Oral Session 15: Features and Medical Image Analysis

Feng ら (Siemens) は、3 次元超音波画像上で胎児の顔を検出する問題を扱い、3 次元および 2 次元情報を組み合わせた学習に基づくアプローチを提案している。また、顔以外の部分を取り除く自動カーヴィングの方法も提案している。

Goh ら (John Hopkins Univ.) は、生体組織内での水分子拡散を計測する HARDI (high angular resolution diffusion imaging) の画像に対し、内挿やフィルタリングといった基本演算および主成分分析などの統計的演算等の実行を可能にするための数学的な枠組みを提案している。HARDI は拡散 MRI の一種だが、各ボクセルでの水分子拡散の方向性を、従来の DTI (diffusion tensor imaging) より高い角度分解能で捉える。HARDI は、各方向に

対する水分子拡散の確率を表す ODF (orientation distribution function) を与えるが、提案手法はこの ODF がリーマン多様体上にあることに基づいて、その上での演算を定義することで、上述の目的を果たす。

Chandrasekhar ら (Stanford Univ.) は、画像間対応を求めるのに SIFT や SURF などの局所特徴記述子が良く使用されるが、分散カメラネットワークや大規模な検索などの応用では、データサイズが小さいことが求められるとし、CHoG とよぶ、低ビットレートで転送可能な特徴記述子を提案している。そこでは、勾配ヒストグラムを効率的に圧縮できる木構造として表現する方法と、圧縮したまま記述子間の距離をデコード (解凍) することなく効率的に計算できる方法を示している。

(担当：岡谷貴之)

### Oral Session 16: Tracking and Learning

Ta (Georgia Tech) らは、物体認識と特徴記述子ベースの追跡を同時かつ高速に行う方法を提案した。2 次元画像上での追跡に代わり、尺度空間上で注目点の追跡を行う。特徴検出と特徴記述のタスクを分離し、運動推定のための特徴記述を省力化し、さらに尺度空間上での運動予測によりフレーム間照合時の探索空間を削減することで高速化を実現した。

Hinterstoisser (ミュンヘン工大) らは、SLAM などを想定した高速かつ頑健なパッチ平行化の方法を提案した。環境中の特徴点の同定と姿勢推定のため、特徴点まわりのアピランスを平均パッチ集合により表現して照合を行う方式において、平均パッチの計算を PCA と線形予測子を用いて高速に行う点に新規性・優位性がある。

Li (USC) らは、雑踏中の歩行者追跡を対象として、tracklets の対応付けのための新しい affinity model の学習法を提案した。彼らはこの対応付け問題がランク付けと識別との融合問題であることに着目し、RankBoost と AdaBoost を統合した新しい Boosting アルゴリズム (HybridBoost) を提案し、多様な特徴の自動選択・統合により頑健な追跡を実現した。

Buehler (Oxford) らは、放送映像とその字幕情報から英国手話を自動的に学習する方法を提案した。手話映像と字幕間での時間ズレ、順序や表現の違いなどに対処するため、両手の軌跡、形状、方向、接触関係を含むシーケンス照合のための距離関数の定義、及び、多事例学習によるスコア関数の最適化を行っている。

(担当：大塚和弘)



## Oral Session 17: Radiometry and Optimization Techniques

Gupta らは、アクティブシーン復元における周囲照明の影響を論じ、プロジェクト光のデフォーカスポケによる形状復元に際して、周囲照明は光輸送を直接モデル化あるいは計測することなく分離することが可能であることを示している。

Tan らは、固定された正射投影で可視な表面法線のなす半球を実射影平面で表現することを提案し、双方向反射率分布関数 (BRDF) の持つ対称性をこの空間上で幾何学的に解析している。またその結果を照度差ステレオにおける曖昧さの削減に応用している。

Komodakis らは高階マルコフ確率場 (MRF) のエネルギー最小化問題において、メッセージ伝播アルゴリズムの双対分解法 (ICCV2007) により、与えられた MRF をより簡単な MRF に分解することでエネルギーを最小化することを提案している。

同じ問題において石川は、任意階のエネルギーを同値な 1 階のエネルギーに変換することにより、QPBO などのグラフカット手法を使って高階エネルギーを最小化する方法を提案している。

(担当: 石川 博)

### その他目を引いたポスター論文

Markerless Motion Capture with Unsynchronized Moving Cameras (MPI Informatik 他)

これは、手持ちのハンディカメラを利用して、何かしらの動作をしている人を複数 (数名) 人数で複数視点から動きながら撮影した多視点画像から、対象人物のモーションキャプチャを行う研究である。通常は、スタジオなどに固定され、しかも時間的な同期の取れた多視点カメラから行うモーションキャプチャを、複数の手持ちカメラで実現したものであり、結果はリアルタイムで得られるものではないが、非常に完成度が高いと感じた。

技術的ポイントはいくつかあるが、まず、複数カメラ間の同期を取るために、ビデオ撮影時に映像と同期して記録される音声チャンネルを利用するという点がポイントである。単純なアイデアであり、そのための処理も、異なるカメラ間で記録された音声信号の相互相関のピークから、カメラ間のタイミングのずれを計測することによりフレーム同期を取るものである。音声チャンネルの時間サンプリングレートはビデオのそれ (15FPS とか 30FPS) に比べて極めて高い (40KHz ぐらい) ので、音声信号パターンの時間的ズレを相互相関という単純な信号処理で見つけるだけで、十分に精度のよいフレーム同期を実現できるのであ

る。このようにして時間的な同期がなされた多視点ビデオについて、各カメラで SFM をバンドルアジャストメントを利用して実行することにより静止している背景形状を復元し、それらを異なる視点間でマッチングすることにより、多視点カメラの 3 次元的位置合わせが行われる。そして最後に、複数カメラ映像に撮影されている人物シルエットに人物 3 次元モデルを当てはめることによりモーションキャプチャを実現している。

(担当: 斎藤英雄)

### おわりに

以上 CVPR2009 について、出席者有志によりその概要をまとめた。紙面都合上、基本的にオーラルセッション発表論文のみの報告となったが、当然ながらその他の論文や紹介した論文の詳細は、プロシーディングスを参照頂きたい。

次回 CVPR2010 は、Larry Davis と Jitendra Malik が General Chair となって、2010 年 6 月 13 日から 18 日までの日程で、サンフランシスコにて開催の予定である。

(文責: 岡谷貴之)