

## IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2008) 報告

川崎 洋† 清水雅夫†† 高松 淳††† 田中正行†† 中澤篤志††††  
延原章平††††† 古川 亮†††††† 芳 世紅††††††† 八木康史††††††††

† 埼玉大学工学部 〒338-8570 埼玉県さいたま市桜区下大久保 255  
†† 東京工業大学 大学院理工学研究科 〒152-8550 東京都目黒区大岡山 2-12-1  
††† 奈良先端科学技術大学院大学 情報科学研究科 〒630-0192 奈良県 生駒市 高山町 8916-5  
†††† 大阪大学 サンバーメディアセンター 〒560-0043 大阪府豊中市待兼山 1-32  
††††† 京都大学 大学院情報学研究科 〒606-8501 京都市左京区吉田本町  
†††††† 広島市立大学 情報科学研究科 〒731-3194 広島市安佐南区大塚東 3 丁目 4 番 1 号  
††††††† オムロン (株) コアテクノロジーセンター 〒619-0283 京都府相楽郡木津町木津川台 9 丁目 1 番  
†††††††† 大阪大学 産業科学研究所 〒567-0047 大阪府茨木市美穂ヶ丘 8-1  
E-mail: †kawasaki@cgv.ics.saitama-u.ac.jp, ††mas@ctrl.titech.ac.jp, †††j-taka@is.naist.jp,  
††††mtanaka@ok.ctrl.titech.ac.jp, †††††nakazawa@cmc.osaka-u.ac.jp, ††††††nob@vision.kuee.kyoto-u.ac.jp,  
†††††††ryo-f@cs.hiroshima-cu.ac.jp, ††††††††lao@ari.ncl.omron.co.jp, †††††††††yagi@am.sanken.osaka-u.ac.jp

**あらまし** IEEE Computer Society 主催, Computer Vision and Pattern Recognition 会議が, 6月22日~28日, 米国アラスカ州アンカレッジにおいて開催された. その概要を参加者9名が分担して報告する.  
**キーワード** 会議報告, CVPR, コンピュータビジョン, パターン認識

### Report on CVPR2008

Hiroshi KAWASAKI†, Masao SHIMIZU††, Jun TAKAMATSU†††, Masayuki TANAKA††, Atsushi  
NAKAZAWA††††, Shohei NOBUHARA†††††, Ryo FURUKAWA††††††, Shihong LAO†††††††, and  
Yasushi YAGI††††††††

† School of Engineering, Saitama University Saitama, Saitama, Japan  
†† Graduate School of Science and Engineering, Tokyo Institute of Technology Meguro, Tokyo, Japan  
††† Graduate School of Information Science, Nara Institute of Science and Technology Ikoma, Nara, Japan  
†††† Cybermedia Center, Osaka University Toyonaka, Osaka, Japan  
††††† Graduate School of Informatics, Kyoto University Sakyo Kyoto, Japan  
†††††† Graduate School of Information Sciences, Hiroshima City University Asa-Minami, Hiroshima, Japan  
††††††† Core Technology Center, Omron Corporation Souraku, Kyoto, Japan  
†††††††† The Institute of Scientific and Industrial Research, Osaka University Ibaraki, Osaka, Japan  
E-mail: †kawasaki@cgv.ics.saitama-u.ac.jp, ††mas@ctrl.titech.ac.jp, †††j-taka@is.naist.jp,  
††††mtanaka@ok.ctrl.titech.ac.jp, †††††nakazawa@cmc.osaka-u.ac.jp, ††††††nob@vision.kuee.kyoto-u.ac.jp,  
†††††††ryo-f@cs.hiroshima-cu.ac.jp, ††††††††lao@ari.ncl.omron.co.jp, †††††††††yagi@am.sanken.osaka-u.ac.jp

**Abstract** IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2008) was held in Anchorage, USA, June 22-28 2006. This is a report on CVPR2008 by nine participants.

**Key words** Conference Report, CVPR, Computer Vision, Pattern Recognition

## 1. ま え が き

CVPR は, ICCV と並びビジョン業界を代表する IEEE Computer Society 主催のコンピュータビジョン会議である。2008 年は, Narendra Ahuja 教授 (University of Illinois at Urbana-Champaign) と Linda Shapiro 教授 (University of Washington) が実行委員長, Kim Boyer 博士 (RPI), Mubarak Shah 教授 (University of Central Florida), Tanveer Syeda-Mahmood 博士 (IBM Research) がプログラム委員長を努める組織構成で, 6月22日から28日の一週間, 米国アンカレッジ市にて開催された。

会議には, 35カ国から1208名の参加者があり, 名実ともに国際色豊かな会議である。参加者の半数強は, 米国からの参加ということで, アラスカ州が米国人にとって特に魅力的な場所であるとの印象を受けた。日本からの登録者数は約4パーセントに当たる50名である。また, 表1に国別の参加登録者数を示す。

査読は, Full paper 投稿, Double blind で行われた。約1600件の投稿に対して, オーラル発表62件, ポスター発表445件と計507件が採択された。採択率は, 31.7%と例年よりやや上がっている。投稿論文を対象とした賞に関しては, 学生を対象としたポスター賞が増え, 次点を含め, 合計で9件が表彰された。それら以外には, H Schneiderman, and T Kanade と C Bregler and J Malik の2グループが, Longuet-Higgins Prize を受賞した。さらに, 特に有意義な査読を行った33委員が, 優秀査読者として表彰された。詳細は, 以下のとおりである。

### • Best Paper Awards

#### (1) Best Paper

Beyond Sliding Windows: Object Localization by Efficient Subwindow Search, Christoph H. Lampert, Matthew B. Blaschko, Thomas Hofmann

#### (2) Best Paper

Global STereo Reconstruction under Second Order Smoothness Priors, Oliver Woodford, Ian Reid, Philip Torr, Andrew Fitzgibbon

#### (3) Best Student Paper

Fast Image Search for Learned Metrics, Prateek Jain, Brian Kulis, Kristen Grauman

#### (4) Best Student Paper Runner-up

Epitomic Location Recognition, Kai Ni, Anitha Kannan, Antonio Criminisi, John Winn

#### (5) Best Poster

The Patch Transform and its Applications to Image Editing, Taeg Sang Cho, Moshe Butman, Shai Avidan, William Freeman

#### (6) Best Poster Runner-up

Viewpoint-independent object class detection using 3d feature maps, Joerg Liebelt, Cordelia Schmid, Klaus Schertler

#### (7) Best Student Poster

Robust dual motion deblurring, Jia Chen, Lu Yuan, Chi-Keung Tang, Long Quan

#### (8) Best Student Poster Runner-up

Manifold-manifold distance with application to face recognition based on image set, Ruiping Wang, Shiguang Shan, Xilin Chen, Wen Gao

(9) Best Demo Real-time Semantic Segmentation with Semantic Texton Forests? Jamie Shotton, Matthew Johnson, Roberto Cipolla, Microsoft Research, Cambridge and University of Cambridge

### • Longuet-Higgins Prize

表1 学会参加者

Table 1 Attendance

| 国                            | 人数   |
|------------------------------|------|
| Australia                    | 22   |
| Austria                      | 18   |
| Belgium                      | 9    |
| Brazil                       | 1    |
| Canada                       | 50   |
| Chile                        | 2    |
| China                        | 64   |
| Czech Republic               | 14   |
| Denmark                      | 9    |
| Finland                      | 3    |
| France                       | 39   |
| Germany                      | 60   |
| Greece                       | 4    |
| India                        | 7    |
| Ireland                      | 3    |
| Israel                       | 29   |
| Italy                        | 17   |
| Japan                        | 50   |
| Korea                        | 21   |
| Luxembourg                   | 2    |
| Mexico                       | 2    |
| Netherlands                  | 19   |
| Norwa                        | 4    |
| Portugal                     | 7    |
| Russia                       | 2    |
| S. Korea                     | 6    |
| Singapore                    | 23   |
| Slovenia                     | 2    |
| Spain                        | 19   |
| Sweden                       | 8    |
| Switzerland                  | 21   |
| Taiwan                       | 12   |
| Turkey                       | 4    |
| UK                           | 35   |
| USA                          | 669  |
| Venezuela                    | 1    |
| Registered for conference    | 1258 |
| No shows (visa problems etc) | 50   |
| Total in attendance          | 1208 |

(1) H Schneiderman, T Kanade

Probabilistic modeling of local appearance and spatial relationships for object recognition. A significant advance in object recognition through probabilistic modeling and multiple-view training, yielding a state-of-the-art face detection technique.

(2) C Bregler, J Malik

Tracking people with twists and exponential maps An inspired application of kinematic modeling techniques from robotics to the challenge of tracking people in motion from a single camera view, including a memorable model-based analysis of the Muybridge motion study videos.

#### • Outstanding Reviewers

Rene Vidal, Bastian Leibe, Stefan Roth, Peter Sturm, Ross Beveridge, Michael Brown, Ahmed Elgammal, Joachim Weickert, Leonid Sigal, Athinodoros Georghiades, Peursum Patrick, Nassir Navab, Douglas Lanman, Marc Pollefeys, Sanjiv Kumar, C.V. Jawahar, Bernt Schiele, Nuno Vasconcelos, Moray Allan, Ghassan Hamarneh, Elli Angelopoulou, Lihi Zelnik-Manor, Moshe Ben-ezra, Ali Shokoufandeh, Ko Nishino, Richard Wildes, Sing Bing Kang, Jean-marc Odobez, David Demirdjian, Sudeep Sarkar, Ales Leonardis, Jan-Olof Eklundh, Rahul Suthankar

本会議は、24日からの3日間で、各曜日いずれも、午前中にポスターそして2並列のオーラルセッション、ランチを挟んで、並列でパネルとポスター、その後、2並列のオーラルが2回という構成で実施された。ポスターセッションのセッションタイトルは、表2のとおりである。オーラル、ポスターともに Geometry や Radiometry などから、Statistical Method, Learning, Segmentation, Categorization, Recognition に話題の中心がシフトしている印象を受ける。今回の会議の特徴としては、招待講演の代わりに、教育、将来、応用を各々キーワードとした3件のパネルが行われた。パネルの概要は後述する。また、デモも活発で、午前と午後の半日づつに分かれ、三日間で合計12件実施された。デモは半日単位で行われたこともあり、オーラルやポスターの間に見に行く事ができた。CVPRは例年、本会議前後にワークショップとチュートリアルを実施している。今回も22, 23, 27, 28の4日間に、19件のワークショップと10件のチュートリアルが開催され、本会議に劣らず盛況であったと感じる。

## 2. Oral Session O1A-1: Color, Illumination, and Reflectance

このセッションでは、色や照明に関する解析やノイズ除去に関して5件の発表が行われた。

K.Sunkavalli(Harvard)らは、屋外で撮影した画像列において、時間変化に加えて、光源のスペクトル変化も考慮することで、シーンの恒常的な色を推定する手法を発表した。光源として、スカイライト成分および太陽光(影とシェーディングを生成)の2つを仮定することで、全てのピクセルの反射パラメー

表2 ポスターセッション

Table 2 Poster session

| セッション名  | 発表件数 |
|---|------|
| P1A-1: Statistical Methods and Learning   | 73   |
| P1P-1: Segmentation and Grouping  | 72   |
| P2A-1: Motion and Tracking  | 75   |
| P2P-1: Object Recognition and Color & Texture   | 75   |
| P3A-1: Stereo and Structure from Motion, Image and Video Retrieval, Object Detection and Categorization | 76   |
| P3P-1: Selected Topics  | 74   |

タ、およびシーンの法線情報やカメラの位置情報などを同時に推定する。応用として、シーン中の影の検出や、求めた法線情報による形状復元、背景画像の作成などが示された。

形状復元手法の一つに、焦点のボケた画像群からシーンのボリューム情報を推定する研究が、トモグラフィーやCVなど複数の分野で行われているが、P.Favaro(Edinburgh)らは、フーリエ解析を用いて、これを解析する手法を発表した。P.Favaroらによると、入力枚数が密な場合には、単純なデコンボリューションにより、復元形状の空間解像度を見積ることができ、入力データが疎な場合には、周波数領域におけるSVDにより効率的にボリュームの復元が実現できることが報告された。

Y.Zheng(Delaware)らは、単画像からピネッティング(周辺減光)を校正する手法を提案した。これは、画像中心から放射方向への輝度微分は、自然画像においては一般にシンメトリックであることから、撮影画像上で観測される偏りを用いて推定する手法となっている。本論文ではこれを、非パラメトリックにLM法で最適化する手法、および、モデルベースで多項式近似する手法の2つが提案された。提案手法のメリットとして、セグメンテーションなどの画像処理が不要であることが報告された。

M.Harker(Leoben, Austria)らは、フォトメトリックステレオなどにより得られる画像の勾配空間から、勾配の二乗誤差を最小化する表面形状を復元する、新しい手法を発表した。M.Harkerらによると、勾配空間からの形状復元はこれまでポワソン方程式を解くことにより行われていたが、これは最小二乗解ではなく、本論文により初めて厳密解を解析的に得られる手法が示された。また、過去の手法では1次近似であったのに対し、提案手法は2次以上の近似であり、精度の高い復元が実現できる。本手法は大きなシーンでも安定して、繰返し処理なく形状復元が可能のため、実用性の高い手法とこのことである。

L.Heらは、高感度撮影時に生じる画像のカラーノイズを除去する手法を発表した。これまででは、このようなノイズの除去にはRGB空間における色情報を用いる手法が主であったが、本手法では、HSI空間で輝度情報を用いることで、効率の良いノイズ除去手法が報告された。この時、画像の輝度エッジの方向を考慮することで、エッジを保存しながらノイズのみが除去される。実験では、感度を下げて撮影したノイズの少ない画像との比較により、過去の手法を上回る結果が得られたことが報

告された。

(担当：川崎 洋 )

### 3. O1A-2: Segmentation and Grouping

G. Sfikas (Univ. of Ioannina) らは、空間的な滑らかさを加味した混合ガウスモデル (Gaussian Mixture Model: GMM) に基づく画像セグメンテーションを提案している。事後確率最大化法と EM アルゴリズムを用い、すべての学習パラメータを自動的に推定する方法を提案している。

P. Kohli (Microsoft Research, Cambridge) らは、高次 Conditional Random Field (CRF) に基づく教師なし学習ベースの画像セグメンテーションを提案している。高次項として、 $P^N$  Potts モデルにおける段階的な変化を直線的な変化に拡張した Robust  $P^N$  モデルを用いることを提案している。この高次項は従来の  $\alpha$ -expansion 等のグラフカットに基づく最適手法でも適用可能であるという特徴がある。

G. Zeng (ETH) らは、画像中に繰り返し現れる類似な模様 (Repetition elements) の情報を利用した画像セグメンテーションを提案している。グラフカットによるセグメンテーションに通常用いられるエネルギー関数に、相互情報量 (Mutual Information: MI) の概念を用いて各対応点間ごとに定義された類似度合を表す項を加えることで実現している。

O. Duchenne (ENS / INRIA) らは、Transductive 学習に基づく、画像セグメンテーションを提案している。Transductive 学習では、クラスラベルが定義された訓練データ (例えば、前景・背景とユーザーによりラベル付けされた画素) と、されていないテストデータ (それ以外のデータ) を同時に利用し学習をおこなう。学習には Graph Laplacian 法を利用している。計算量もグラフカットと同程度であり、高速である。

M. Maire (UC Berkeley) らは、抽出されたエッジを利用した交差点 (Junction) の検出・位置推定をおこなう手法を提案している。学習ベースの局所エッジ抽出と一般化固有値問題によるセグメンテーションによる大域的なエッジ抽出の結果を組み合わせることで、精度の向上を実現している。頂点と各エッジとの距離の重み付き和を EM ベースの最適化、つまり 1. 重みの推定と 2. 最少となる頂点位置の推定の 2 つのステップを繰り返すことにより、交差点の抽出をおこなっている。

(担当：高松 淳 )

### 4. O1P-1: Motion Analysis for Structure, Shape, and Pose

Furukawa (Univ. of Illinois) らは多視点映像を用いた形状と運動の同時復元手法として、初期フレームの形状復元とその頂点移動による変形という枠組みに分類されるアプローチを提案した。提案手法の特徴は、(1) 各頂点とその近傍を剛体と仮定した場合の局所的な運動 (頂点移動) 推定と、(2) 全体を非剛体運動と仮定した最適化という 2 段階の最適化プロセスにある。今回の Furukawa らの手法は、ポスター発表の別件と同様に、昨年度の CVPR で発表された “oriented-point” を活用したものであり、1 つの基礎技術から一連の研究が展開されてい

る点に興味深い。

Zhang (Zhejiang Univ.) らは時系列画像を用いたステレオによる depth 推定手法として、bundle-adjustment によって隣接フレーム間の coherency を確保する手法を提案した。すなわち (1) まず各フレーム独立に depth-map を推定し、(2) 次に各フレームの depth-map を、その前後の depth-map を固定したと見なした上で、フレーム間の depth-map 変化に関する滑らかさ制約を満たすように更新する、という手法である。

Lim (Australian National Univ.) らは全方位カメラのようなデバイスによって得られる球面画像を用いた自己運動推定問題に対して、球面座標上のある 1 点と、球の中心に対してその反対側の点という “antipodal points” において optical flow を観測すれば、自己運動としてあり得る解は一定の範囲に絞られることを利用した手法を提案している。端的には、1 組の “antipodal points” における optical flow が同じ方向であれば並進が、逆方向であれば回転が中心の運動であるということであるが、著者らは 1 組の “antipodal points” から得られる解の範囲を数理的に示した上で、これを複数組用いて解の範囲の積をとることによって自己運動推定問題を従来の手法よりもより頑健に解くことができることを示している。

Lager (EPFL) らは鏡面反射成分を用いることで、3 次元モデルと撮影映像の間のレジストレーション精度を向上させる手法を提案している。一般的にはフレーム間のテンプレートマッチングによる対応点探索を用いるところであるが、この手法のポイントは、(1) 粗く 3 次元モデルをレジストレーションする、(2) レジストレーション結果を用いて各フレームにおける光源環境マップを推定する、(3) これらが固定光源下では一定であるとみなして再度レジストレーションを行うという手順を踏む点である。

Salzmann (Ecole Polytechnique) らは textureless な対象の変形運動を単視点映像から推定するための deformable mesh model を提案している。提案手法のポイントは、deformable mesh model の変形パラメータ (運動モデルパラメータ) を学習によって獲得する際に、モデル全体ではなく、微小なパッチについてのみ学習を行い、これを組み合わせて全体を表現する点である。これによって学習に必要なサンプル数が大幅に削減され、またモデルの形状・トポロジーが変化したとしても同じ材質であれば学習データをそのまま使用することが可能となっている。(担当：延原章平 )

### 5. O1P-2: Object Detection, Categorization, and Recognition (1)

Zass らは 2 枚の画像間の特徴点群対応問題において、Probabilistic Graph を用いる手法を提案している。ここでは、各画像における特徴点の相対位置関係を hypergraph で表現することで、2 画像間の対応問題を hypergraph matching の対応問題を解くことで解決している。

Wu らは、Viewpoint-Invariant Patches (VIP) を用いて、シーンのマッチングを行う手法を提案している。VIP はテクスチャマップされた 3 次元シーンを、Rectification, SHIFT-

feature による特徴点抽出を行ったあと、その各点の3次元位置、パッチのスケール、法線方向などをエンコードすることで得られる。

Kim らは、画像群からオブジェクトのカテゴリを自動的に求め、分類する方法を提案している。本手法では画像群の局所特徴をすべて対応付けし、その対応付け関係の相関係数から Similarity Matrix を用いる。オブジェクトのカテゴリははこの Similarity Matrix を Normalized Clustering 法で解析することで求めることができる。

Shotton らは、画像をオブジェクトの領域ごとに分割する問題を、Semantic Texton Forest(STF)を用いることで解く手法を提案している。画像はパッチに分割され、パッチごと(あるいは相対的な)輝度等の特徴量として用いる。画像中の部分領域に対しこの特徴量のヒストグラムを求め、あらかじめ学習したヒストグラムと比較することで、部分領域ごとのオブジェクトカテゴリの認識を行っている。

Yang らは一般化物体認識問題において近年盛んに用いられている Bag of Feature(BoF) モデルの問題点を指摘し、新たな手法を提案している。BoF モデルでは画像群から得られる局所特徴をクラスタリングしコードブックとすることで認識を行うが、彼らの指摘はこのコードブック化におけるクラスタリングによって認識性能が低下するという提案であり、コードブックを用いずカテゴリ毎に最適化された"visual bits"を用いることで従来法よりも良い性能が出ることを示している。

(担当: 中澤寛志)

## O2A-1: Motion and Tracking

モーションとトラッキングのセッションでは5本の論文があった。その中で、3本は複雑環境での人体の検出トラッキングに関連するもので、残りは人の動作認識と画像のブレの復元に関するものであった。近年、多くの研究者の関心が顔より難しい人体の検出とトラッキングのほうに移ってきている。人体の画像は顔よりも見え方のパラエティが富んでいるため、複雑環境でロバストに人体検出を行うことは難しい。解決の手がかりの一つは静止画からの人体検出と動画からの動き検出、トラッキングによる融合である。本セッションの中で、3本の論文ともこの考え方を取り入れている。ドイツの TU Darmstadt の Andriluka らの "People-Tracking-by-Detection and People-Detection-by-Tracking" [1] はまずタイトルが分かりやすく、印象的であった。複雑背景と隠れのある環境での人体検出とトラッキングのどちらも難しいが、本論文では(1)人体検出の部分に人体のパーツを連結させた人体モデルを用いて、歩行する人を横から見た場合の下肢の位置をも検出する。(2)複数フレームから検出した下肢の動きをダイナミックモデルで人体検出の性能を高める。(3)動画の中から検出された一連の人体検出の結果からオフラインでその人の見え方を学習するという方法を使って隠れを含む複雑環境の中での人体検出トラッキングができるようにした。このセッションのもうひとつ面白い論文は米国 Northwestern Univ の Dai らによる "Motion from Blur" [2]。一枚の画像の中のブレ領域からオブ

ティカルフローと同じようなローカルな線形計算でその部分の動きを推定できることを提案した。

(担当: 芳世紅)

## O2A-2: Object Detection, Categorization, and Recognition (II)

K. Ni (Georgia Tech.) らは、Torralba らの混合ガウスモデルを用いた Location Recognition を拡張し、Epitome を用いる手法を提案している。同一クラスの画像は Epitome 上のある特定の領域に射影されるという拘束を、通常の Epitome の取得方法に加えた上で作成し、それを用いてベイズ推定に基づき認識をおこなっている。

C. H. Lampert (Max Planck Institute for Biological Cybernetics) らは、Branch-and-bound による枝がかりにより Sliding Window による物体位置推定 (Object Localization) の手法の高速化を提案している。探索する範囲の最小、最大を表す領域集合  $\mathcal{R}$  を定義し、その領域を逐次絞っていくことで位置推定をおこなう。その際、

$$(1) \hat{f}(\mathcal{R}) \geq \max_{R \in \mathcal{R}} f(R).$$

$$(2) \hat{f}(\mathcal{R}) = f(R), (\mathcal{R} = \{R\} \text{ のとき}).$$

を満たす  $\hat{f}$  を定義・利用することにより Branch-and-bound による枝がかりをおこなうことができる。なお彼らはこの研究でベストペーパーを受賞した。

D. Hoiem (CMU) らは、彼らが提案してきた Surface Orientation, Occlusion, Viewpoint-and-object 検出器を複合して用いることにより、それぞれの検出器を単独で用いるよりも良い結果を得る方法を提案している。推論の際には、3つの検出器を交互に繰り返し使い、各繰り返しのステップでは、他の2つの検出器の現在の推定結果を利用して残りの検出器の推定をおこなう。

M. Brand (Mitsubishi Electric Research Labs) らは、循環を持つグラフ構造からなる Conditional Random Field (CRF) を、効率よく学習する方法を提案している。通常 CRF を学習するためには、項が和と積の形で混合された評価関数の最小化をおこなう必要があるが、粗く見積もった下限値は和のみで表現できることに着目し、それを最小化することにより学習をおこなった。これにより、学習の全工程が単なる線形方程式を解く程度でおこなえるようになる。

J. Cui (Tsinghua Univ.) らは、複数枚の同系統画像向けの Transductive な画像の前景背景分離手法を提案している。具体的には、画像間のマルコフ確率場 (Markov Random Field: MRF) における Gibbs エネルギー関数の Likelihood の項に Local Color Pattern (注目領域を囲む領域の色の関係) や Edge Profile Feature (注目領域中のエッジの周辺色の関係) を用いることを提案している。これらの特徴は並進・回転・スケールに不変な特徴がある。

(担当: 高松 淳)

## 6. O2P-1: Statistical Methods and Visual Learning

Carreira-Perpinan らは、従来法に比べ局所解に陥りにくい、次元圧縮による regression の方法を提案している。本手法では Dimensionality Reduction by Unsupervised Regression (DRUR) 法を提案し、サンプル点以外の周辺においても良いマッピングが行える方法であることを示している。

Zhang らは、独立成分分析 (ICA) において、サンプル数が少なかつた場合に正しく推定ができない問題を、入力画像をテンソル化することで解決する手法を提案している。

Tappen らは、MRF を用いた画像領域のラベリング問題に対し、その領域同士の関係をグラフモデルで学習する Logistic Random Field 法を提案している。このモデルにより、物体領域の相互関係性をコンパクトに学習でき、路面-路肩分割などの問題に適応できることを示している。

Talwalkar らは、数百万オーダーの巨大な顔サンプルからその形成空間 (manifold) を学習する方法を提案している。元となる手法は Isomap および Laplacian Eigenmaps という既知のものであるが、巨大なノードおよびエッジ (18 million nodes and 65 million edges) により空間を形成している。このため、2種類の近似的な解法を試している。実験結果から、Nystrom 近似の方が他方 (Column-sampling) より優れていることを確認している。

(担当: 中澤篤志)

## 7. O2P-2: Stereo

このセッションでは、以下の6件 (ステレオアルゴリズム2件, 特徴記述子1件, SfM1件, キャリブレーション1件) の発表があった。なお、Woodford らの “Global Stereo Reconstruction under Second Order Smoothness Priors” は、Best Paper Award を受賞した。Gallup (University of North Carolina) らは、距離推定誤差が一定になるように、ベースライン長と画像解像度を変えつつマルチベースラインを用いて3次元形状を再構成するアルゴリズムを提案した。三角測量を行うときの、カメラへの2つの光線間角度が一定になるという特徴がある。移動するカメラで計測対象を撮影した複数画像を入力として想定し、マルチベースラインアルゴリズムでは Plane-sweeping を用いて画素ごとの対応度を求めている。Woodford (University of Oxford) らは、視差マップの滑らかさ拘束に2次の事前分布を組み入れたグローバル最適化アルゴリズムを提案した。入力とするステレオ画像は2枚以上の複数枚、画像間の類似度評価にはオクルージョン画素を除くロバスト評価関数を用いる。滑らかさ拘束には、水平・垂直各方向に連続する3画素位置の視差の2階微分の総和を利用する。グローバル最適化には、何らかの提案視差と現在の推定視差を用いた、QPBO (Quadratic Pseudo-Boolean Optimization) と呼ばれるアルゴリズムを用いる。視差はサブピクセルレベルで推定できる。Tola (EPFL) らは、ベースライン長が大きなステレオ画像間での対応を求めるための、高速に計算できる局所領域の特徴記述子を提案した。

提案手法では、画像を8方向に微分し、微分値が負の場合には0で置き換える。さらにそれぞれ複数回、ガウシアンカーネルを畳み込む。ある注目位置に対して決められている位置にある処理結果を取り出すことで、方向と解像度に対する特徴が得られる。提案する特徴記述子は、SIFT や GLOH と似ているが計算は高速である。Byröd (Lund University) らは、魚眼レンズのような歪曲収差が大きな画像間での、基本行列や基礎行列とレンズの収差パラメータを同時に推定する手法を提案した。歪曲収差は、1個のパラメータで表す。画像ごとに異なる歪曲収差パラメータと基礎行列を9個の対応位置から求める方法と、共通の歪曲収差パラメータと基本行列を6個の対応位置から求める方法を示した。Aguar (Instituto Superior Técnico) らは、SfM における観測行列が不完全な場合にも、因子分解法のグローバル最適解を得るアルゴリズムを提案した。観測行列の欠損領域は、ヤング図形形状になることを仮定する。残差行列のスペクトルノルムを最小化することで、繰り返し計算なしで欠損要素を順次求めることができる。Kumar (University of North Carolina) らは、視野に重ならないカメラ間で外部パラメータを推定するために、鏡の反射像を利用するキャリブレーション手法を提案した。カメラとキャリブレーションパターン (平面) は固定し、キャリブレーションパターンがいろいろな姿勢で撮影されるように鏡の向きを変える。まず、通常のキャリブレーション手法で内部パラメータと鏡に反射するカメラの姿勢を推定する。次に、推定結果と実際のカメラとの関係から、実際のカメラの姿勢を推定する。(担当: 清水雅夫)

## 8. O3A-1: Face, Gesture, and Action

このセッションでは同じく5本の発表があり、顔を対象としたのが2本で、手の動作認識が1本、残りの2本は人体の動作理解である。従来の顔器官検出アルゴリズムの多くは、特徴点を特定するために、予め特徴点の位置であれば1を、そうでない場合は0を出力する識別器を学習させることで実現しているものが多いが、Univ. of Maryland の Hu らは顔の器官輪郭抽出に Boosted Ranking Model を導入し、モデルと特徴点の位置ずれの大きさに応じてスコアが変わるため、降下法でモデルを正しい方向にフィッティングさせていくことができる。CMU の Ashraf らの “Learning Patch Correspondences for Improved Viewpoint Invariant Face Recognition” のアイデアは顔認識を行うときの顔の向きによる影響を取り除くためのシンプルな手法で、異なる向きの顔画像の中のパッチの関係を統計学習で実現したもの。このセッションの最後に発表されたのは INRIA の Laptev らによる “Learning realistic human actions from movies”。なんと一般の動画の中のシーンから人の行動を自動的に認識するという大胆な課題設定。極めてチャレンジな課題設定だと思うが、下記の方法で実現できることをデモした。やり方は (1) まず映画脚本を用いて映像の中の人の動作の自動ラベリングを行い、学習データベースを作る。(2) 時間空間に拡張した “bag-of-features” と非線形 SVM を使った識別器で動作を認識。映画のシーンから8種類の人の動作を認識できるデモを実演できた。

(担当: 芳 世紅 )

## 9. O3A-2: Correspondence and Registration

このセッションでは、以下の5件の発表があった。Olsson (Lund University) らは、アウトライアーを含む点群間の対応と変換に関する大域的最適解を、多項式時間で求めるアルゴリズムを提案した。まず、RANSACによって解の候補を見つける。次に、見つけた解の候補とは異なる対応と変換が可能かどうか、アウトライアーをチェックする。提案したアルゴリズムを使うと、50%以上のインライアーによる解がないことも検出できる。Ahmed (MPI Informatik) らは、動作する人物などを複数の方向から撮影した同期ビデオを使って、アニメーションのための3次元サーフェースモデルを再構築する。グラフィックスハードウェアも利用したアルゴリズムを提案した。画像中の特徴点を抽出して時間方向の疎な対応を取る。フレーム内の特徴点を元にサーフェースメッシュを作り、時間方向でパーテックスの対応を取る。Hedau (University of Illinois) らは、画像をセグメンテーションしたセグメント間での対応を得るために、partial matching cost (PMC) と region adjacency graph (RAG) を利用したアルゴリズムを提案した。照明変化やカメラの姿勢変化などのわずかな影響によって、画像のセグメンテーションは大きく異なる可能性がある。しかし、コントラストが高い主要領域については安定に対応が求まることから、PMCを利用する。また、RAGの対応を得るために、多対一のラベリングを利用する。Hansen (Technical University of Denmark) らは、非剛体レジストレーションで用いる多変量Bスプラインの制御点を適応的に設定するアルゴリズムを提案した。画像間で非剛体レジストレーションするために、初期状態では画像の画素値勾配の大きさに応じた密度で配置した制御点を使って、シンプレックススプラインを基底関数とする多変量Bスプラインを用いて片方の画像を変形する。画像間のSSDを評価値として、制御点数の増加が評価値を小さくすれば制御点を増やす。パラメータ(制御点の移動)最適化には、inverse compositional 繰り返し最適化手法(ICLA)を用いている。Zhu (University of Kentucky) らは、TOF(Time-of-flight)距離センサとステレオ距離計測結果を統合する手法を提案した。TOFセンサとカメラは、内部及び外部パラメータをキャリブレーションする。ステレオ画像から距離マップを得るときにMAP-MRFを用いるが、事前確率としてTOFセンサからの距離情報を利用する。距離情報の統合結果は、いずれか一方だけからの結果よりも高精度である。

(担当: 清水雅夫)

## 10. O3P-1: Video Analysis and Image and Video Retrieval

このセッションでは、動画解析や、動画および静止画像の検索に関して5件の発表が行われた。

Prati らは、監視カメラなどから得られた人の動きの軌跡を解析し、それらの軌跡を分類する手法を提案した。提案手法で

は、軌跡を、運動方向の確率分布で表現する。確率分布の数式モデルとしては、混合 Von Mises(フォンミーゼス)モデル(円周関数における混合ガウス分布)が利用されており、EMアルゴリズムによってパラメータを決定する。これにより方向ベクトルの類似性が Bhattacharyya 距離(確率分布同士の類似性尺度の一種)として定義され、この類似性に基づいたDPマッチングとクラスタリングによって、動きの軌跡の分類が行なわれる。

Peter らは、2次元図形同士の類似性を表現するために、まず図形を2次元確率分布で近似し、分布同士の距離を類似度とする方法を提案した。分布を表すモデルとしては、wavelet変換を用いる。ここで、分布そのものでなく、分布の平方根をwavelet変換し、これによってwavelet係数が超球面上に乗ることを利用して、分布同士の類似性を球面上の距離で表した。また、2次元図形同士の類似性を表現する上で、平行移動による影響を無視するために、wavelet係数の順序の変更を許す方法を提案した。これは、図形領域では、一部の矩形領域が「スライド」することを許す類似性で、筆者らはこれをsliding waveletと呼んでいる。

Li らは、画素の特徴のヒストグラムに画素同士の近接性の情報を付加する方法を提案した。この方法では、画素の近接情報を、ラベル同士の共起性を記述した「共起行列」によって表現する。続いて、この共起行列を、各ラベルを遷移状態とするマルコフ連鎖の遷移確率行列として解釈し、マルコフ連鎖の「定常分布」を求め、これと従来のヒストグラムとをあわせたベクトルを、マルコフ静的特徴(MSF)として提案した。MSFは、従来のヒストグラムの2倍程度のサイズで、画素同士の空間的近接性の特徴を表すことができる。

近年、画像検索の手法としてBag of featuresパラダイムが利用されるようになったが、このパラダイムでは、keypoint同士の空間的位置関係が無視されている。Behmo らは、keypoint同士の空間的距離を考慮した新しい画像表現を提案した。具体的には、keypointを頂点とし、keypoint同士の空間的距離を辺の重みとしたグラフを作り、これからkeypointのクラス間の距離を定義して画像の特徴とする。クラス間の距離は、グラフ理論における距離概念の一種である、「Graph commute times」によって定義される。直観的には、この数値は、海、空、ビルなど、ある特徴を持ったkeypoint同士の、画像上での大まかな位置関係を表現している。

大量の画像から、ユーザが定義した画像間距離による最近傍探索を行なう場合に、Locally Sensitive Hash function(LSH関数)(特徴ベクトルから整数値へのハッシュ関数で、特徴ベクトルの類似性がハッシュの衝突確率に一致するもの)を利用して探索を高速化することができる。Jain らは、ユーザによって与えられた一部の画像対についての類似度から、その類似度を反映したマハラノビス距離を学習し、その距離によるLSH関数を構成する手法を提案した。提案手法を高次元の特徴空間に効率的に適法する方法も同時に提案されている。本手法を人体姿勢推定や画像分類に適用した場合、全体の1%以下の画像の探索でも、全探索に近い分類性能が得られたことが報告された。(担当: 古川 亮 )

## 11. O3P-2: Selected Topics

Simakov (Weizmann Inst. of Sci.) らは “bidirectional similarity” という基準を用いた画像・映像の要約手法を提案した。ここで bidirectional とは、(1) 要約前→要約後：要約には最大限元画像・映像の情報が含まれる、(2) 要約後→要約前：要約して得られたものには元画像・映像に無かった情報が加えられてはならない、という 2 つの基準を表しており、それぞれ completeness, coherence と呼んでいる。著者らはこの制約に基づく評価関数 (similarity measure) を導入し、これを最適化することによって要約を実現している。

Porikli (MERL) は実行コストが定数時間となる bilateral filtering を提案した。ここで定数時間とは、フィルタのウィンドウサイズに対して実行コストが不変という意味である。定数時間を実現するポイントは、同じ著者が過去に提案した “integral histogram” を用いて冗長な計算を省略する点にある。

Treibitz (Technion-Israel Inst. of Technol) らは、水中カメラのように撮像系と対象物体の間に屈折率の異なる媒質が存在する場合のカメラモデルとその幾何学的校正について提案している。具体的には水中カメラのように、空気中のカメラがガラス面を通して水中の対象物を撮影する状況を想定しており、このような場合に対象から撮像面に至る光路上に存在する水-ガラス面およびガラス-空気面による屈折をモデル化し、そのキャリブレーション法を提案している。

Liu (MIT) らは optical flow 推定に関する研究のベンチマークのために、実環境データセットと、これを作成するためのソフトウェアを発表した<sup>(注1)</sup>。

Chandraker (UCSD) らはコンピュータビジョン分野の問題のうち双線形計画問題となっているものを分枝限定法によって効率的に大局最適解を求める手法を提案した。提案された手法では、コンピュータビジョン分野における具体的な問題は、しばしば線形計画問題の一方が他方よりも変数の数の上で極端に少ないことを利用して、変数の数が少ない側でのみ分枝操作を行うことで効率化を達成している。

(担当：延原章平)

## 12. Panel

CVPR2008 では、キーノート講演がなく、パネルディスカッションが行われた。Computer vision education, Future directions in computer vision research, Computer vision applications という三つのテーマ設定がなされ、それぞれのテーマについて、5, 6 名のパネラーが各 1 時間ほど、ディスカッションが行われた。全般的に議論が発散してしまい、少しまとまりのない議論になる傾向があった。以下にそれぞれのテーマについて、キーワードを中心に簡単に報告する。

Computer vision education のテーマでは、まず、各パネラーが、それぞれの大学におけるコンピュータビジョン関連のコースの紹介、テキストブック、プロジェクト研究を、順番に紹介

した。おもしろいプロジェクト研究でなくては、学生を惹きつけられない。一方、おもしろいプロジェクト研究の本質を理解するためには、数学、光学など幅広い知識が要求される。逆に、本質を理解しなくても、Matlab や OpenCV を利用することで、研究らしきことはできてしまう。つまりは、コンピュータビジョンに必要な知識の広さが、教育の困難さの一因であると議論されていた。また、産業界からも、既存の方法は良く勉強しているが、新しいことに取り組むことができない学生が多いという意見も挙がっていた。

Future directions in computer visions のテーマでは、今後の方向性として、(1) 膨大な量のデータ、(2) 評価方法の確立、(3) システムとしての統合、(4) 他分野 (AI, CG, 数学, センサ) との融合の 4 項目が重要になってくると議論がされていた。膨大な量のデータについては、そもそもどうやって集めるか、仮に集められたとしてどのように取り扱うかが問題である。また、コンピュータビジョンだけで何がができるわけではないので、コンピュータビジョンの技術を含むシステムを統合していくことが重要である。一時期、AI などとは距離があったが、最近では関連する研究も多くなり、他分野との融合も今後期待されているなどといった議論がされていた。

Computer vision applications のテーマでは、学術的な成功とビジネス的な成功には大きなギャップがあるというところから議論がはじまった。コンピュータビジョンを利用したキラーアプリケーションの必要性が昔から言われているが、コンピュータビジョンはシステムを構成する要素技術であり、システムではない。そのため、キラーアプリケーションという考え方自身が少し違うのではないかという意見も出された。一方、現在では、数多くの防犯カメラの設置やカメラが非常に普及している。そのため、今後は、多くのカメラを利用したアプリケーションや、Computational photography の分野がコンピュータビジョンの応用として期待されるという意見が出ていた。

(担当：田中正行)

## 13. あとがき

9 名の報告者により CVPR2008 の会議報告を行なった。次回 CVPR は、2009 年 6 月 20-26 日に米国フロリダ州マイアミビーチで開催される (<http://www.cvpr2009.org/>)。多くの優秀な論文が日本から投稿されることを期待する。

(注1) : <http://people.csail.mit.edu/ceiliu/motionAnnotation/>