

2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2006) 報告

井尻 善久 (オムロン)	波部 斉 (京都大学)
阪野 貴彦 (東京大学)	藤吉 弘亘 (中部大学)
榎原 靖 (大阪大学)	増田 健 (産業技術総合研究所)
斎藤 英雄 (慶應義塾大学)	佐藤 洋一 (東京大学)

joyport@ari.ncl.omron.co.jp	habe@media.kyoto-u.ac.jp
vanno@cvt.iis.u-tokyo.ac.jp	hf@cs.chubu.ac.jp
makihara@am.sanken.osaka-u.ac.jp	t.masuda@aist.go.jp
saito@ozawa.ics.keio.ac.jp	ysato@iis.u-tokyo.ac.jp

あらまし: 2006 年 6 月 19 日~21 日に米国ニューヨーク州ニューヨーク市で開催された国際会議 CVPR2006 の概要を報告する。

Report on 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2006)

Yoshinisa Ijiri (Omron Corp.)	Hitoshi Habe (Kyoto Univ.)
Atsuhiko Banno (Univ. of Tokyo)	Hironobu Fujiyoshi (Chubu Univ.)
Yasushi Makihara (Osaka Univ.)	Takeshi Masuda (AIST)
Hideo Saito (Keio Univ.)	Yoichi Sato (Univ. of Tokyo)

Abstract: This report gives an overview of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2006), which was held in New York City, NY, USA, from June 19th to 21st, 2006.

1 はじめに

2006 年 6 月 19 日から 21 日までの 3 日間にわたり、国際会議 IEEE Computer Society Conference on Pattern Recognition (CVPR2006) が New York University のキャンパス (米国ニューヨーク州ニューヨーク市) で開催された。General Chairs は Daniel Huttenlocher (Cornell University) と David Forsyth (UC Berkeley), Program Chairs は Andrew Fitzgibbon (Microsoft Research), Camillo J. Taylor (University of Pennsylvania), Yann LeCun (New York University) がそれぞれ務めた。

今回の CVPR は第 24 回目にあたり、最近の ICCV, ECCV, ACCV などコンピュータビジョン分野における他の主要会議と同様、非常に多くの論文が投稿され、会議にも多数の人が参加した。具体的には 1131 件の論文が投稿され、その中から 54 件がオーラル発表に、264 件がポスター発表にそれぞれ採択された。これを採択率でみるとオーラルが 4.8%、ポスターを加えた全体で 28.1% と、ICCV などよりはやや採択率が高いとはい

え、かなりの狭き門となっていることがわかる。

CVPR2006 における Best Paper Prize は Derek Hoiem, Alexei Efros, Martial Hebert(CMU) による"Putting Objects in Perspective"に贈られた。また、Honorable mention としては Andreas Opelt, Axel Pinz, Andrew Zisserman(Univ. of Oxford) による"Incremental Learning for Object Detection Using a Visual Shape Alphabet"が選ばれた。また、今回の CVPR では新たに Longuet-Higgins Prize が設けられた。これは CVPR'96 に発表された論文の中で、10 年間にわたり高い評価を得た論文に対して与えられるものであり、Cordelia Schmid と Roger Mohr(INRIA) による"Combining Greyvalue Invariants with Local Constraints for Object Recognition"と、Henry Rowley, Shumeet Baluja, Takeo Kanade(CMU) による"Neural Network based Face Detection"が選ばれている。

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と発表件数を表 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。

また、本会議前後には、3 日間にわたり、14 件のサテライトワークショップ、6 件のチュートリアル、デモ展示が開催され、本会議同様、非常に多くの人々がさまざまなトピックの会議に参加した。

(以上、斎藤、佐藤)

2 オーラルセッション

O1& O11: Recognition I/Recognition III

近年、画像認識の分野では多クラスの認識を問題とする一般物体認識アルゴリズムが特に脚光を浴びている。今までに単一のクラスの認識問題に対するアルゴリズムは数多く提案されており(例えば顔検出等において)かなりの成功を収めている。しかしこれらの方法を直接多クラス問題に適用しようとすると、それに伴い処理時間やリソースサイズの増大といった問題が生じる。このために全てのクラスに対して共通の処理や判別モデルの一部を共有化することを目的とした研究が目立つ。またこれらの問題に対して統計的モデルに基づく Generative Model と統計的モデルを機械学習法により学習する Discriminative Model の両方の観点から研究が進んでいる。

Discriminative Model に関しては、A. Opelt(Graz univ. of Tech) らが、CVPR2004 で発表された A. Torralba(MIT) らの JointBoosting の考え方を基に改良を

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	33	145	178
イギリス	5	24	29
中国	4	23	27
イスラエル	2	14	16
ドイツ	2	8	10
フランス	2	7	9
カナダ	2	5	7
ベルギー	1	5	6
日本	0	6	6
オーストラリア	1	4	5
オーストリア	1	2	3
ギリシャ	1	2	3
イタリア	0	3	3
シンガポール	0	3	3
スウェーデン	0	2	2
スロベニア	0	2	2
スイス	0	2	2
韓国	0	2	2
その他 (計 5 カ国)	0	5	5
合計	54	264	318

表 2: セッション一覧

Oral Sessions
O1: Recognition I
O2: Segmentation
O3: Geometry
O4: Stereo and 3D Reconstruction
O5: Low Level Analysis
O6: Recognition II
O7: Matching and Registration I
O8: Graphics and Image Synthesis
O9: Learning
O10: Matching and Registration II
O11: Recognition III
O12: Tracking and Video Analysis
Poster Sessions
P1: Learning
P2: Graphics and Image Synthesis
P3: Tracking and Video Analysis
P4: Matching and Recognition
P5: Segmentation
P6: Geometry
P7: Recognition I
P8: Recognition II
P9: Low Level and Photometric Analysis
P10: Recognition III
P11: Stereo and 3D Reconstruction
P12: Matching and Registration II

加えた方法を提案している。JointBoosting は単一クラス認識問題において成功した Adaboost を多クラス問題に適用する方法でありクラス間で弱識別器を共有する方法である。しかし JointBoosting では予め扱おうとするクラス全てに対して大量のデータを収集し、事前にそれら全てのデータを用いて学習させておく必要があった。したがって新たにクラスを追加したいときには全てのデータを用いて再度学習する必要があったが、彼らの提案によれば事前に学習されたクラスの情報を基に新規クラスを任意に追加可能である。共有化および追加学習は特徴量および弱識別器の二段階で行われる。尚この発表は CVPR2006 の Honorable Mention を受賞している。

一方、Generative Model を使った多クラス認識に関して、Mikolajczyk(Univ. of Surrey) らは appearance および geometry に関する確率分布を基に階層的な特徴量表現をすることにより、より効果的に Bayes 識別器を構成する方法を提案し、この方法により単一のクラスに特化した識別器と同等の性能を実現できると報告している。

CVPR2006 の Best Paper Prize 受賞論文である Hoiem(CMU) らの発表では、おおよそのカメラ位置、物体の種類、空間的配置の確率的因果関係を求めることにより性能向上を図る方法が提案されている。この手法は、検出器の部分は任意に変更可能であり、応用範囲が広く将来的拡張性の高い柔軟なモデルである。

ところでこれら一般物体認識問題の基本となる特徴量抽出やその表現方法またそれらのマッチング方法に関しても様々な方法が提案されている。前述の Mikolajczyk らは局所特徴量の Descriptor として PCA により次元削減した SIFT(Scale Invariant Feature Transform) を用いている。またこれを物体中心に関する極座標系で記述することにより回転不変性を与えている。Lazebnik(UIUC) らも、SIFT をいくつかの離散的なタイプに分割し領域毎にマッチングを行う方法を提案している。この方法では特徴量の離散化によりよりロバストでありながら空間的情報も考慮したマッチングが可能となる。なおこれは Grauman(MIT) らが提案している SIFT 特徴量を複数の解像度において抽出し適当な順序で配列したベクトルをマッチングさせる手法の拡張でもある。

物体認識全体の傾向として以前は Bag of words モデルに見られるように空間的な配置を考慮せずに実現しようとする物が多かったが、空間的情報を上手に利用する方法が近年盛んに提案されるようになってきていると言える。

(以上、井尻)

O2: Segmentation

A. Criminisi ら (Microsoft Research, Cambridge) の "Bilayer Segmentation of Live Video" では、CVPR2005 で発表した Stereo Bilayer Segmentation に続く研究として、リアルタイムで1台のカメラ映像から全景と背景の二つのレイヤーに分割する手法を提案している。本手法では、時間的な事前確率、空間的な事前確率、色の尤度、モーションの尤度からエネルギーを定義し、グラフカットを用いてエネルギーの最小化を行い、全景と背景にセグメンテーションする。実験では、従来のステレオセグメンテーションアルゴリズムと比較した結果、同等の検出精度で検出できることが示されている。

J. Winn ら (Microsoft Research, Cambridge) は、対象物のカテゴリが既知である物体の一部にオクルージョンが発生していても、物体を検出しセグメンテーションする手法を提案している。物体領域をグリッド状にパッチをレイアウトし、各パッチを Part Detector によりどの部位に属するかをラベリングする。Layout Consistent Random Field は、パッチの部位検出結果とそのレイアウトの整合性から局所における空間的な制約を出力する。実験では、オクルージョンの影響を受けずに、車や顔の検出とそのセグメンテーションが可能であることが示されている。

(以上、藤吉)

O3: Geometry

このセッションでは、幾何に関する発表が5件おこなわれた。

Frahm(North Carolina at Chapel Hill) らは、縮退したようにみえるデータから、精度良く F 行列や射影行列を求めるための手法 (QDEGSAC) を提案した。例えば、1枚の平面が大部分を占めているような対象シーン (Quasi-Degenerate data) では、RANSAC を使って F 行列を求めようとしても、特徴点の大半がその平面から採られてしまい、縮退してしまっているのと同じ状況となって、正確な F 行列推定等ができなくなる。そこで著者らは、もともとの RANSAC でアウトライアとしてはじかれてしまったインライアを復活させ、正しい行列を推定している。

次に、Zheng(上海交通大) らは画像の口径食 (Vignetting) 補正を単一画像のみからおこなった。この手法では、Vignetting を考慮した画像のセグメンテーションをおこない、各々のセグメントに対しての重みを計算して、Kang-Weiss モデルによる Vignetting のパラメータ群を求めている。

Carceroni(Pennsylvania)らは、もしカメラの位置が既知で姿勢のみが未知であった場合に、どのような制約が導き出されるのかを、3枚の画像の基本行列(E行列)を解析することで、数理的な考察をおこなった。従来は、カメラの位置・姿勢両方を未知として3次元復元をおこなうものが殆どであったが、GPS等によってカメラ位置が既知であるようなケースが今後増えるだろうことを見越しての問題設定である。

Sim(ANU)らは、 L_∞ ノルムを用いたアウトライア除去を提案した。CVにおいては、とりわけ幾何的な問題に関して、コスト関数最小化のフレームワークで最適解を導き出す手法が一般的である。これまで、 L_2 ノルムを最小化する方法が多く提案されてきたが、 L_∞ ノルムを採用することで、準凸最適化問題に帰着され、2次錐計画問題(Second Order Cone Programming)として解が得られることから、最近 L_∞ ノルムに関する発表が多くなっている。ただし、 L_∞ ノルムはアウトライアに弱いことが知られているため、著者らは L_∞ ノルムを用いて、アウトライアを順次取り除いていった。なお、同一著者らによる、 L_∞ ノルムを用いたカメラ運動推定が、ポスターセッション6で発表されていた。

(以上、阪野)

O4: Stereo and 3D Reconstruction

このセッションでは、複数枚の画像からの3次元形状復元について、4件の発表がなされた。

Atkinson(York)らは、偏光を用いて2枚の画像からの3次元形状復元をおこなっている。テクスチャのない物体を20度回転させた2枚の偏光画像から、それぞれ局所パッチを生成し、それらに対し偏光度と位相を基にしたコストを設定し、マッチングをとることで距離画像を得ている。

Hornung(RWTH Aachen)らは、階層的Volumetricな3次元復元について発表をおこなった。彼らはボクセルの代わりに、それら6面を結ぶ8面体グラフ構造を利用し、全体のグラフから最小カットを結ぶことで、多面体表面として対象物体の形状が得られることを示した。

Drouin(Montreal)らは、視差画像におけるデプス値のラベリングで発生する不連続面を検出し(Border-Cut)、1次元のDPによるロバストな滑らかさ拘束を導入することによって、より正確な視差画像を推定した。

Seitz(Washington大)らは多眼ステレオによる3次元形状復元の既存アルゴリズムに関して6つの項目をたてて、比較・評価をおこなった。

なお、このセッションに限らず、提案したアルゴリズムの評価については、<http://cat.middlebury.edu/stereo/>

でのランキングを用いるのが一般化しつつあり、もし、ステレオアルゴリズムを考案したら、ここに参加してみるのも面白いと思う。

(以上、阪野)

O5: Low Level Analysis

M. Elad(Israel Institute of Technology)らは、ローカルな画像パッチに対する冗長辞書(基底画像)とそれに対する疎な重み係数による線形表現を利用したノイズ除去手法を提案している。従来はDCT等の固定の冗長辞書が用いられていたが、提案手法では入力画像からK-SVDアルゴリズムによって事例ベースで辞書を学習している。

C. Liu(MIT)らは、輝度値に対するノイズの大きさ(分散)を表すノイズレベル関数(NLF)を、単一画像から推定する手法を提案している。入力画像の画素値を一次関数で区分的に当てはめ、各区分(領域)の平均輝度値と当てはめ誤差の分散の組を取得し、その組に対する下限包絡線をNLFとして考える。但し、実際には外れ値の影響があるため、代表的なカメラレスポンスとノイズの組み合わせから学習した基底関数の重み付き線形和に限定してNLFを推定している。

S. Lucey(CMU)らは、方向の異なる顔画像同士の照合を、辞書画像を分割した各パッチ画像と入力画像全体との間の見えに関する同時生起確率を用いて行う手法を提案している。従来の照合の評価基準には、画像の位置合わせ誤差に対して頑健な全体画像に関する同時確率や、方向変化などのローカルな変化に対応し易い画像パッチの差分が用いられていたが、提案する評価基準はその両者の利点を持ち合わせており、顔画像データベースFERETに対する実験で性能が改善されることを報告している。

F. Champagnat(Office National d'Etudes et Recherches Aerospatiales)らは、バーコードのような一次元画像特徴の超解像手法と性能評価のモデリング手法を提案している。対象の一次元信号は、サンプリング間隔で分割した端点で滑らかに接続する二次関数で近似され、その係数をTikhonov正規化の枠組みで最適化している。また、対象物体までの観測距離・超解像に用いる画像枚数に対する解像度の上昇率の関係を、実験的及び理論的アプローチで求め、その比較を行っている。

(以上、横原)

O6: Recognition II

S. Todorovic(Univ. of Illinois)らは、未知クラス物体を撮影した画像セットから、各クラスの部分画像を抽出

する手法を提案している。最初に、画像を段階的に領域分割することで、各領域をノードとする木を作成し、各ノードに輝度の平均値や分散等の光学的特徴及び面積や重心位置等の幾何的特徴を与えておく。次に、画像セットから特徴の類似度を最大化する共通部分木を抽出することで、各クラスの部分画像を領域の集合として学習する。

O.C. Ozcanli(Brown Univ.)らは、軽トラックやSUV等の車種クラス認識において、シルエット形状に加え、二つの形状間のシルエット内部の密な対応点における見えを利用した照合手法を提案している。内部の対応点は、輪郭上の対応点から法線方向に沿って内側に一定距離だけ離れた点や、輪郭上の四つの対応点に対する二本の対角線の交点として取得する。照合時には、対応点の輝度と輝度勾配の各々についてヒストグラム(同時確率分布)を作成し、その相互情報量を最大化するクラスとして認識する。

H. Chen(UCLA)らは、服装に見られるような襟・肩口・袖口等の要素を様々に組み合わせた構造を表現するためのグラフを提案している。従来のAndノードグラフによる固定構造では組み合わせ選択を許容する構造への対応が困難であったため、構文解析において用いられていたAnd-Orグラフを適用して対応している。実験では、スケッチ画やエッジ画像から服用要素や要素間関係の抽出を行い、その有効性を報告している。

B. Wu(Univ. of Southern California)らは、シーンや人物同士による隠蔽条件下における、人体の各部位(全体・頭部・胴体・足)の検出結果に基づく人物追跡手法を提案している。各部位の候補領域検出は、学習データを基に作成したカスケード型識別器によって行う。追跡時には、前フレームの結果に対して、各部位の候補領域間の位置・大きさ・見え(領域内の色ヒストグラム)の類似度を計算し、隠蔽を考慮した可視度による重み付けをすることで、各対応を求める。

A. Veeraraghavan(Univ. of Maryland)らは、速度変化による時間軸上のワープを考慮した行動認識の手法を提案している。認識対象となる各行動は、基準速度における行動の特徴ベクトルの時間関数と時間ワープ関数の取り得る範囲の関数集合の組で表現され、関数集合はワープ関数の性質を利用して上限・下限関数のみを用いて定義される。認識時には、DTW(Dynamic Time Warping)によって範囲内から最適なワープ関数を選択し、特徴量の誤差を最小にする行動として認識する。

(以上、横原)

O7: Matching and Registration I

L. Gu(CMU)らの”3D Alignment of Face in a Single Image”では、一枚の入力顔画像から3次元顔部位(両目、鼻、口、顎のライン)のアライメント手法を提案している。3次元顔モデルの作成に、3次元顔画像データベースから顔の各部位の83点のランドマークを抽出し、その集合から固有形状(Eigenshape)を求める。同様に、ランドマーク点を中心としたパッチの集合から、固有パッチ(Eigenpatches)を求める。入力画像から顔検出器により得られた顔領域を初期値として、繰り返し法によりフィッティングを行う。しかし、照明等の影響により顔領域に陰影が存在すると、正確なランドマーク点を探索することが難しい。そこで、部分空間において平均シェイプへのShrinkageと呼ばれる正則化を行い、例外値であるランドマーク点の座標を修正しつつ、繰り返しフィッティングを行う。これにより、正しい顔部位のランドマーク点を得る信頼性を上げることが可能となることが示されている。

L. Liangら(Microsoft Research, Asia)は、上記の発表と同様に、正確な顔部位の位置同定の手法を提案している。顔形状サンプルから得られた顔形状モデルを組み込んだMarkov Networkによりラインセグメントで構成される最適な顔部位の形状を求め、正則化を行う。正則化の際に、画像情報を考慮することで、より正確な顔部位の位置同定を実現している。

(以上、藤吉)

O8: Graphics and Image Synthesis

H.W. Kang(USTC, China)らは、ビデオデータから時空間領域で有用な情報をもつ領域を抽出・要約する手法を提案した。従来のフレーム単位(時間方向)のビデオ要約に比べて、時空間双方でコンパクトに要約可能であるとしている。まず、時空間領域で高いsaliencyをもつ領域(saliency blobs)に分割したあとで、出力ビデオへのpacking問題を解いている。論文の中では局所的な画像特徴を用いてsaliencyを評価しているが、発表の際には顔検出器の出力をsaliencyとしたときの結果も紹介された。

S.J. Koppal(CMU, USA)らは、固定カメラ・固定物体に対して光源を滑らかに移動させたときの画素値変化に現れる極値から、対象物体の幾何形状(法線方向)をクラスタリングする手法を提案した。対象物体の反射特性に関する事前知識や参照物体などを用いなくても、十分長い時間画像を撮影すれば、幾何形状に依存して生じる極値が安定に抽出できることが示されている。

(以上、波部)

O9: Learning

F. Kang らの"Correlated Label Propagation with Application to Multi-label Learning":では、一つのデータに対して複数のラベルが対応するような学習を行うために、ラベル間の相関性を考慮した Correlated Label Propagation を提案し、画像データの自動ラベル付けの実験を行い性能評価している。

D. Lin らによる"Pursuing Informative Projection on Grassmann Manifold"では、特徴空間での計量とクラス判別の間での情報量を Effective Information とし、それを最大にするように射影行列 A を求める。 A をグラスマン多様体上の点として扱うことにより、直交性と回転不変性を満たすことができる。顔認識で実験したところ、従来の PCA や LDA などより性能の向上がみられたとしている。

R. Hadsell らによる"Dimensionality Reduction by Learning an Invariant Mapping"では、近接データ間には引張スプリング、非近接データ間には圧縮スプリングを作用させることにより、低次元空間への次元圧縮を行う研究。手書き数字の 4 と 9、飛行機の画像データなどで実験している。

C. Sminchisescu らの"Learning Joint Top-Down and Bottom-up Processes for 3D Visual Interface"では、単眼の画像系列から SIFT オペレータによって抽出された特徴から関節角度で表現された人間の姿勢を推定するための学習を行っている。

(以上、増田)

O10: Matching and Registration II

A. Vedaldi らによる"Local Features, All Grown Up"では、局所アフィン不変特徴量によって得た局所的な対応領域を拡大することによって画像間で共通する領域を抽出している。例えば、同一の瓶を含む異なった複雑画像間に適用し、瓶を共通領域として抽出できるか実験している。

S. K. Zhou らの"BoostMotion: Boosting a discriminative similarity function for motion estimation"では、複雑な状況での運動追跡や変化抽出を可能にするために、変化を評価する関数を boosting の手法によって学習する手法を提案し、心臓の超音波画像系列での運動推定について実験している。

A. Rajwade らによる"A New Method of Probability Density Estimation with Application to Mutual Information Based Image Registration"では、単に輝度値の組み合わせを joint histogram として単に画素数を数えることにより確率密度関数を得るのではなく、画像の輝度

値の微係数の関数を積分することによっても確率密度関数が推定できることを示し、相互情報量による画像の位置あわせに適用した場合、従来手法より正確な結果が得られたとしている。

D. Cremers らによる"Nonparametric Priors on the Space of Joint Intensity Distributions for Non-Rigid Multi-Modal Image Registration"では、位置あわせされた状態で画像に様々な変化を加えたときの相互情報量の変化を事前知識として学習しておくことにより、相互情報量による画像の位置あわせを、隠蔽やノイズなどの劣化に対して頑強にすることができるという研究が紹介されている。

L. M. Lui らによる"Automatic Landmark Tracking and its Application to the Optimization of Brain Conformal Mapping"では、大脳表面のメッシュモデルを大域的な等角写像によって平面や球に射影することにより、大脳溝を自動的に検出する手法が提案されている。

(以上、増田)

O12: Tracking and Video Analysis

P. Sand (MIT, USA) らは、空間的に密に配置した particle の長期間の挙動を追跡する手法を提案した。まずオブティカルフローによって密な速度場を求めたあと、前フレームで得られた particle との結合、空間的な配置の最適化、不要な particle の削除、新規 particle の追加を繰り返して長期間にわたる挙動をトラッキングしている。

V. Parameswaran (Siemens, USA) らは、Mean-Shift を用いた対象追跡の拡張として、バンド幅が調整可能なカーネル関数を用いた手法を提案した。そこでは、人物の追跡を行うためにモーションキャプチャデータから人の各部位の動きを求め、それぞれに適応したバンド幅を学習している。頭部に比べると足の動きが大きい、といった情報を表現することができ、従来に比べてオクルージョンの影響を受けにくく、安定に追跡ができることが示されている。

(以上、波部)

3 ポスターセッション

P1: Learning

機械学習における性能向上の方法として、新しい入力が得られた際にそれらを用いてオンライン学習をして性能向上を図ることが考えられる。この問題に対して、Grabner(Graz Univ. of Tech.) らは Boosting をオンラインで行う方法を提案している。この方法では、予

め定義された複数の特徴量から最小エラーを与える最適特徴量を選択する Selector を構成しておき、Selector が選択した複数の最適特徴量セットの各々に対し最適な重みを算出する。このオンライン学習過程は Haar-like 特徴や方向性ヒストグラムといった簡単な特徴量を使えばかなり短時間で実行、実際に実時間トラッキングへ応用している。

(以上、井尻)

P2: Graphics and Image Synthesis

A. Zomet (Columbia Univ., USA) らは、レンズを用いず、光の減衰率が時間的・空間的に変化できる絞りをを用いた光学系で、様々な画像を撮影できることを示した。機械的な駆動系を持たないパン・チルトの実現や、異なった視野の画像の同時撮影、さらには、光学的畳み込み演算を利用したテンプレートマッチングなどに応用できることがプロトタイプシステムを用いて示された。

P.H. Huang (NTHU, Taiwan) らは、対象物体を囲むように設置したミラーに映った物体の像を 1 台のカメラで撮影して、カメラキャリブレーションと 3 次元形状復元を行う手法を提案した。像の輪郭同士で共有する接線が(仮想的な)カメラ間のエッジ曲線に相当することを利用してキャリブレーション問題を解いている。

N. Ramanathan (Univ. of Maryland, USA) らは、幼年期の人の顔の成長のモデル化法を提案し、顔認識の精度が向上することを示した。人体測定学の知見を利用し、顔のパーツ相互の位置関係の変化を取り入れたモデルを提案している。

(以上、波部)

P3: Tracking and Video Analysis

G. Brostow (Cambridge) らは "Unsupervised Bayesian Detection of Independent Motion in Crowds" において、人混みの中から歩行している各個人をベイズ識別により検出する手法を提案している。従来、人の検出にはアピアランスや形状の特徴を用いるが、本手法では点特徴の時系列データ群のみを入力とする。一人の人に対して複数のクラスタとなるように点特徴群をオーバークラスタリングしておく。各点の動きが同一人物内であれば類似した動きであることに着目し、ある 2 クラスタ間が同一クラスタである事前確率、各クラスタに属する二つの特徴点間の動きの整合性と、各クラスタが同一クラスタであるかの信頼性から、クラスタの結合と切り離しを実現する。実験では、アピアランス情報等を用いなくても関わらず、従来の手法とほぼ同程度の精度で検出できることが示さ

れている。

(以上、藤吉)

P4: Matching and Recognition

J. Ham らによる "Learning a Manifold-Constrained Map between Image Sets: Applications to Matching and Pose Estimation" では、姿勢・視点・照明などの条件を変えて撮影した多くの画像から生成された特徴空間の多様体間の対応付けをおこなうことにより、視点や姿勢などの推定を可能にしようとする研究が紹介されている。

J. Yang らの "Non-Rigid Image Registration Using Geometric Features and Local Salient Region Features" では、板拘束された変形を伴う画像間の位置あわせを行う RHDM (Robust Hybrid Deformable Matching) という手法を提案している。Local saliency 特徴を用い、外れ値除去を含んだエネルギー最小化により定式化している。

H. Liu らによる "Neighborhood Aided Implicit Active Contours" では、Level set の積分を含んだ汎関数によって定義した Active contour によってノイズを含んだり分断されたりした輪郭も復元しようという研究が報告されている。

A. Mittal らによる "An Intensity-Augmented Ordinal Measure for Visual Correspondence" では、小領域内の画素対の輝度値の大小関係の変化に基づく特徴量を用いることによって従来手法より輝度変化に頑強な変化抽出や対応付けを行えたとされている。

Z. Sun らによる "Image Comparison by Compound Disjoint Information" では、Disjoint Information (DI) は結合エントロピーから Mutual Information (MI) を引いた部分であるが、評価基準として DI を用いたほうが MI より安定に位置あわせが行えるという内容の研究が紹介されている。

H. Q. Dinh らによる "Multi-Resolution Spin Images" では、Spin Images の bin size や support length を変えることによって多重解像度化すると速度や性能の向上が見込めるかが検討されている。

J. Zhang らによる "Robust Traching and Stereo Matching under Variable Illumination" では、Markov Network による Stereo Matching や Optical Flow の定式化であるが、Illumination Ratio Map と組み合わせることによって輝度変化にも頑強になった報告されている。

O. Chum らによる "Geometric Hashing with Local Affine Frames" では、Maximally Stable Extremal Re-

gionsから求めた3点の組であるLocal Affine Framesを特徴量としてGeometric Hashingすることにより視点と輝度の変化に頑強な位置あわせ手法を提案している。(以上、増田)

P5: Segmentation

T. Tyoda(東工大)らは、画像の全体から得られるTop-down情報と、画像の局所的な特徴から得られるBottom-up情報を統合してシーン理解の新しいフレームワークを提案している。Top-down情報には、画像全体から得られる構図を利用し、Bottom-up情報には、テクスチャと色から得られる特徴を利用する。これらのTop-down情報とBottom-up情報を、Isingモデルにより統合しラベルの整合性の検査と更新を繰り返し、最終的には局所的にも大域的にも整合した認識結果を得る。道路画像のラベリングに適用した結果、統合することで91.8%まで改善したことが報告されている。

(以上、藤吉)

P6: Geometry

このセッションでは、主に幾何に関するテーマ(カメラキャリブレーションを含む)についての発表がおこなわれ、そのなかから興味深いと感じたものを数例紹介する。

Ke(CMU)らは、 L_∞ ノルムを用いた画像からの形状復元をおこなった。通常、 L_∞ ノルムの最小化が意味を持つのは、画像から抽出される特徴点の検出位置の誤差が、画像中の位置に関して独立で、等方的な場合に限られる。しかし、実際の場合にそのような仮定は成り立たない。そこで著者らは、準凸最適化のフレームワークに、誤差の方向性を考慮したものを提案した。

Gurdjos(IRIT-TCI)らは、Confocal Conicsによる平面パターンを利用したカメラキャリブレーションについて、理論と実用面について、さまざまな2次曲線について考察している。キャリブレーションに用いる平面パターンでは、チェッカーボード等が広く使われているが、彼らの研究によれば、同一焦点を持つ2次曲線群の方が精度が良いことを示している。

Yu(香港中華大)らは、対象シーンの3次元形状を復元することなく、ステレオ視に使用しているカメラの位置・姿勢を推定した。通常、SFMでは、対象となる物体(表面上の点)の3次元形状や座標の復元を経由してカメラ位置等を検出するが、ここでは画像列からカメラ位置・姿勢を直接に求めている。この発表ではステレオ画像列から、拡張型カルマンフィルタとtrifocal tensorを用いて、姿勢の追跡をおこなっており、積算される誤差を抑えている。(以上、阪野)

P7: Recognition I

多クラス認識問題に関して、Wang(UIUC)らは前述のLazebnikらと異なり画像から抽出された特徴間の確率的依存性を考慮した方法を提案している。Li(UIUC)らの提案手法であるHierarchical Dirichlet Processに、あるクラスに特有の特徴量間の依存性を学習して、それを考慮したモデルを作ることにより精度向上ができることを報告している。性能比較においてはLazebnikらの手法が良いようであるが、将来的な拡張の余地を残した手法であると考えられる。

(以上、井尻)

P8: Recognition II

本セッションでは、イベントや行動の検出・認識・解析に関する13件の発表があった。認識対象には顔の表情変化や歩容等の様々なものが含まれていたが、ここでは歩容認識に関する1件の発表を紹介する。F. Cuz-zolin(UCLA)は、辞書に無い未知方向から観測した歩容に対して、双線形モデルを用いて個人を認識する手法を提案している。最初に、いくつかの既知方向における歩容特徴ベクトルを学習データとして、各方向への変換行列と方向に依存しない個人の特徴ベクトルを取得する。認識時に未知方向の特徴が入力として与えられると、EMアルゴリズムによって観測に最も合う個人特徴ベクトルの選択(個人認証)と未知方向への変換行列を求める。

(以上、楢原)

P9: Low Level and Photometric Analysis

G. Vogiatzis(Univ. of Cambridge, UK)らは、ハイト成分をもつ物体を回転させて撮影した画像群から、その形状と照明条件を推定する手法を提案した。まず、視体積交差法でvisual hullを求める。次いで、contour generatorが物体の真の形状に相当することを利用してRANSACによって光源方向を推定する。具体的には、visual hull上で任意に選んだ点の法線方向と輝度からの光源方向推定を十分良い精度が得られるまで繰り返す。次いで、得られた3次元形状と光源方向を用いて形状の高精度化を行っている。

Q. Zhu(Univ. of Pennsylvania, USA)らは、Shape from Shadingにおける法線方向決定問題において、グラフ構造を用いて大域的な整合性をとる方法を提案した。カメラの視線方向と法線方向が一致するsingular pointsをグラフのノートとし、その間の奥行き差をグラフのエッジに重みとして与え、Max-cut問題に帰着されることを示した。

T. Parag(Rutgers Univ., USA)らは、動的環境での

背景差分に有効な特徴量の選択手法を示した。様々なバンド幅をもつカーネル関数を用いたカーネル密度推定で背景画素値が従う確率密度関数を推定し、さらに、それらによる識別結果を RealBoost に入力して最良の特徴（カーネル密度関数）の組み合わせを選んでいる。

(以上、波部)

P10: Recognition III

本セッションでは、物体クラス認識や画像検索等を含む合計 16 件の発表があった。ここでは、三次元形状の検索手法に関する発表を紹介する。Y. Liu(北京大)らは、物体の部分的な三次元形状検索のためのコンパクトなモデル表現とそれによる効率的な照合手法を提案している。最初に三次元形状の表面上の各点に対して、回転や並進に対して不変な Spin Image と呼ばれる形状特徴を計算し、それらのクラス代表を、低次元の Word Histogram(WH) と呼ばれる特徴として保持する。更に、文書解析の分野で用いられる bag-of-words モデルを WH に適用し、Shape Topics と呼ばれる階層的な特徴を作成することで、認識率の低下を最小限に抑えつつ、更なる低次元化と照合の効率化を行っている。

(以上、横原)

P11: Stereo and 3D Reconstruction

ここでは、ステレオ(広義に、校正カメラによる複数枚画像からの形状復元)や SFM に関する研究が発表された。

Liu(CUNY)らは、大規模物体(例えば建築物)に対し、レンジセンサによって取得した密な 3 次元形状と、SFM を利用して復元した疎な 3 次元形状を位置合わせすることによって、3 次元形状モデルにテクスチャマッピングをおこなっている。

Hirschmüller(German Aerospace Center)は、テクスチャの少ない平面や不連続線の多い構造的環境(Structured Environments)下において、本人が昨年の CVPR で提案した Semi-Global Matching を改良した手法によって、3 次元形状復元をおこなった。なお、プロシーディングスには含まれていないが、彼が発表時にノート PC で出していた、航空写真からベルリン市街の 3 次元形状を復元したビデオは秀逸であった。

Strecha(K.U.Leuven)らは、画像中に突発的なアウトライア(例えば建造物の画像に映り込んだ通行人など)があった場合の多眼ステレオで、EM アルゴリズムを用いて、3 次元形状復元をおこなった。この研究では、E-ステップのエネルギ関数について、Bethe 近似と Mean Field 近似と両方を実装し、前者の性能が良いことを示した。

Goessele(Washington)らは、多眼ステレオにおいて、ごく単純に、ウィンドウ・マッチングからデプスマップを求め、それらをマージングすることで、かなり精度の良い 3 次元形状復元を実現した。基本的な既存手法のみによる形状復元であるが、極めればレンジセンサ顔負けの 3 次元形状が得られる、との例を示したものと見える。

このセッションの全体的な印象として、画像からの特徴点抽出に殆どの場合 SIFT が用いられるようになったこと、CV の分野全体においてステレオ特に多眼視(数十〜数百枚程度)に関する研究が多くなってきたこと、等が挙げられる。

(以上、阪野)

P12: Matching and Registration II

J. Xiao らによる"Simultaneous Registration and Modeling of Deformable Shapes"では、2 次元および 3 次元の変形を伴う複数形状の位置あわせと形状統合を factolization の枠組みで定式化している。2D と 3D 両方の実験結果が示されている。

L. Rai による"Real-time Image-based Guidance Method for Lung-Cancer Assessment"では、気管支鏡から得られる映像を CT データに画像ベースでリアルタイムに位置あわせすることによって肺がん検査の支援を行うという内容が報告されている。

S. Wang らによる"3D Surface Matching and Recognition Using Conformal Geometry"では、3 次元曲面形状を等角写像によって 2 次元平面に射影すれば 3 次元の形状マッチングを 2 次元画像の問題として扱えると報告されているが、実際には spin image など多くの処理を必要としていると思われる。

H. Wu らによる"Coupled Bayesian Framework for Dual Energy Image Registration"では、X 線の強度を変えて撮影された二枚の dual energy images を使用して、位置あわせ問題と骨と軟組織の分離問題を同時に解こうとする研究が報告されている。

J. Liebelt らによる"Robust AAM Fitting by Fusion of Images and Disparity Data"では、3 次元形状データを Active Appearance Model と組み合わせることによって安定性と性能を向上できたとしている。実験では、ステレオカメラとビデオカメラで撮影した会議の参加者の姿勢や表情の分析を行っている。

P. Bhat らによる"Piecewise Image Registration in the Presence of Multiple Large Motions"では、局所的に剛体運動している物体を含む大きく離れた視点からの二つの画像を入力として、運動による物体の分離と形

状復元を行っている。使用している主要な技術は SIFT オペレータと RANSAC である。

A. Eden らによる "Seamless Image Stitching of Scenes with Large Motions and Exposure Differences" では、複数の露出で多方向を撮影した画像を用いて High Dynamic Range パノラマ画像を生成する研究が紹介されている。幾何補正の部分は既存の研究であり、graph-cut による移動物体検出と HDR な輝度補正の部分に重点が置かれている。

M. Kolomenkin らによる "Image Matching Using Photometric Information" では、異なった環境で計測された色の拘束条件を利用して対応関係を確率によってモデル化した Photometric Image Matching により視点や照明条件の異なる画像間の対応を行うとともに、SIFT と組み合わせることによりより正しい点間の対応付けが行えたとしている。

(以上、増田)

4 おわりに

本報告では、2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2006) について、その概要をまとめた。会議の参加者有志により CVPR2006 で発表された研究の動向を紹介したが、300 件以上の論文の全てを紹介することは不可能であり、注目すべき研究はまだ他にも多く発表されていると思われる。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 CVPR2007 は 2007 年 6 月 17 日から 22 日にミネソタ州ミネアポリス市で開催されることになっている。論文投稿締切りは 2006 年 12 月 3 日（エントリ締切りは 11 月 27 日）となっており、日本からも多くの論文が投稿されることを期待したい。最後に、CVPR2006 会場において次々回の CVPR、すなわち CVPR2008 の開催地が参加者の投票により選ばれ、アラスカ州アンカレッジ市に決まったことを報告する。

(以上、斎藤、佐藤)