Vol.2011-CG-145 No.18
Vol.2011-CVIM-179 No.18
2011/11/18

CVPR2011 報告

安倍 満[†1] 石川 博[†2] 岩村雅一[†3]
坂上文彦[†4] 佐藤いまり[†5] 佐藤真一[†5]
杉本茂樹[†6] 玉木徹[†7]
西山正志[†8] 阮翔[†9]

2011 年 6 月 21 日〜23 日に米国コロラド州コロラドスプリングスで開催された国際会議 CVPR2011 の概要を報告する．

CVPR2011 Report

MITSURU AMBAI,[†1] HIROSHI ISHIKAWA,[†2]
MASAKAZU IWAMURA,[†3] FUMIHIKO SAKAUE,[†4]
IMARI SATO,[†5] SHIN'ICHI SATOH,[†5] SHIGEKI SUGIMOTO,[†6]
TORU TAMAKI,[†7] MASASHI NISHIYAMA[†8] and XIANG RUAN[†9]

We give an overview of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2011), which was held in Colorado Springs, Colorado, USA, from June 21st to 23rd, 2011.

†1 デンソーアイティーラボラトリ (Denso IT Laboratory)
†2 早稲田大学 (Waseda University)
†3 大阪府立大学 (Osaka Prefecture University)
†4 名古屋工業大学 (Nagoya Institute of Technology)
†5 国立情報学研究所 (National Institute of Informatics)
†6 東京工業大学 (Tokyo Institute of Technology)
†7 広島大学 (Hiroshima University)
†8 株式会社 東芝 (Toshiba Corporation)
†9 オムロン株式会社 (OMRON Corporation)

はじめに

CVPR[⋆1] は，ICCV とともに CVIM 関連分野のトップコンファレンスと位置づけられる国際会議である．毎年 6 月に開催され，2011 年は 6 月 21 日から 23 日のあいだ，米国コロラド州コロラドスプリングスの Crowne Plaza ホテルで行われた．実行委員長はコロラド大学コロラドスプリングス校の Terrance Boult 教授，カーネギー・メロン大学の金出武雄教授，エルサレム・ヘブライ大学の Shmuel Peleg 教授の 3 人，プログラム委員長はシカゴ大学（その後ブラウン大学に移籍）の Pedro Felzenszwalb 准教授，イリノイ大学アーバナ・シャンペーン校の David Forsyth 教授，スイス連邦工科大学ローザンヌ校の Pascal Fua 教授の 3 人である．実行・プログラム委員長それぞれ一人ずつは北米以外の人であることからもわかる通り，CVPR は北米の地域的な会議とは言い難く，ECCV も同様だが少なくとも欧米をまたいだ会議といえる．これはエリアチェア (AC) の顔ぶれを見ても同様である．CVPR は近年巨大化しており，サンフランシスコで開催された昨年は 1900 人以上の参加者があったという．今回は事前参加登録が 1500 人で打ち切られ，現地登録は不可能であった．そのため今年は，口頭発表のネットビデオが見られ，質問もできるヴァーチャル参加も可能になった．

今回の有効論文投稿本数は 1677 本，その中から 59 本 (3.5%) の口頭発表論文と 379 本 (22.5%) のポスター論文が採択された．投稿時に著者が選択する分野毎の採択率と占有率が公表されているので表 1 に示す．これらの統計は論文の採否が決定した後でとられたもので，採否決定の上では分野の比率等は全く考慮されない．

論文はダブルブラインドで投稿，査読，採否決定され，査読者はもちろん採否を決定する AC にも各論文の著者はわからない．3 人のプログラム委員長は著者を含め全てのデータを見ることができるが，彼らは採否について意見を述べないことになっている．各論文に 3 人つく査読者は，査読後も担当 AC および他の査読者との匿名のオンライン討論に参加を求められる．査読結果が著者に開示されると著者は反論の機会を与えられ，それをうけて担当 AC と査読者はさらに討論をする．それらの後の 2 月，全 AC がシカゴ空港近くのホテルに

⋆1 CVPR の正式名称は以前は “IEEE Computer Society Conference on Computer Vision and Pattern Recognition” だったが，前回 2010 年から “IEEE Conference on Computer Vision and Pattern Recognition” と “Computer Society” がとれたようである．これは IEEE Computer Society が単独スポンサーでなくなったからのようで，今回などは上納金をめぐって一旦は IEEE Computer Society を完全にスポンサーから外すなどの交渉が行われたようである．最終的には IEEE Computer Society とコロラド大学コロラドスプリングス校がスポンサーになり，そのおかげで参加費が安くなったという話も聞かれた．今後の方針は決着しておらず，ICCV2011 の PAMI TC Meeting でも話し合われる予定である．

© 2011 Information Processing Society of Japan

Vol.2011-CG-145 No.18
Vol.2011-CVIM-179 No.18
2011/11/18

表 1 分野による採択率と占有率

分野	採択率	全投稿中	口頭発表中	ポスター中
Vision for Robotics	16%	1.13%	0.00%	0.79%
Human Identification	17%	1.07%	1.69%	0.52%
Face and Gesture Analysis	21%	5.84%	3.39%	4.99%
Performance Evaluation	22%	0.54%	0.00%	0.52%
Applications of Computer Vision	23%	5.01%	3.39%	4.46%
Motion and Tracking	23%	7.10%	10.17%	5.51%
Sensors	23%	0.78%	0.00%	0.79%
Early and Biologically-inspired Vision	24%	2.44%	1.69%	2.36%
Object Detection	25%	5.49%	8.47%	4.72%
Scene Understanding	26%	2.80%	5.08%	2.36%
Medical Image Analysis	26%	3.40%	0.00%	3.94%
Shape Representation and Matching	28%	3.82%	3.39%	4.20%
Vision for Graphics	29%	0.42%	0.00%	0.52%
Image and Video Retrieval	29%	4.95%	8.47%	4.99%
Color and Texture	29%	1.43%	1.69%	1.57%
Stereo and Structure from Motion	30%	4.89%	3.39%	6.04%
Video Surveillance	31%	2.15%	1.69%	2.62%
Segmentation and Grouping	31%	6.74%	5.08%	8.40%
Shape-from-X	32%	2.21%	1.69%	2.89%
Document Analysis	33%	0.72%	1.69%	0.79%
Illumination and Reflectance Modeling	33%	1.25%	1.69%	1.57%
Computational Photography and Video	34%	3.46%	8.47%	3.94%
Object Recognition	35%	5.66%	6.78%	7.61%
Statistical Methods and Learning	35%	7.87%	3.39%	11.55%
Video Analysis and Event Recognition	35%	4.89%	8.47%	6.30%
Optimization Methods	36%	2.62%	6.78%	3.15%
Image-Based Modeling	38%	2.03%	3.39%	2.89%
Withdrawn/Admin Rejects	0%	9.30%	0.00%	0.00%

2 日間集まって最終的な採否を決定した．そこでは AC は複数のパネルと呼ばれるグループに分けられ，各パネルに論文が割り当てられる．これは，多くの AC は自身論文を投稿しているので，著者のいるパネルに論文が行かないようにして，各パネルでは自由に討論できるようにするためである．そのため AC は他パネルの AC とは論文の話をしないように求められる．さらに AC は二人ずつペアになり，それぞれの担当論文の採否を互いに説明しあった上で最終的に採否を決定する．総じて極めて入念に採否決定過程の公平性を保とうとしているようである．

以下，口頭発表された論文を中心に，その概要を報告し，また末尾には会場の雰囲気などを伝える．
（文責：石川 博）

Oral Session 1A: Image and Video Retrieval

Gong らは，高次元の特徴量を短いバイナリコードに変換する問題設定において，PCA および CCA を用いた教示無し／有り学習に基づく変換手法を提案している．著者らは，PCA によって次元圧縮された特徴量をランダムな直交行列によって回転させた後に，各軸における符号列をバイナリコードとして用いるという単純な方法でも，ビット数が短いときは従来法よりも性能が良いという興味深い実験結果を示している．また，ランダムな直交行列の代わりに，バイナリコード化前後における量子化誤差を最小化するような回転行列を用いることで，さらなる性能改善を達成している．提案手法は PCA を CCA に置き換えることで教示有り学習に拡張が可能である．

Zhang らは，Visual word 間の相対位置関係を表すオフセット空間を用いて，局所的・大局的な Visual word の組み合わせから成る Geometry-preserving Visual Phrases(GVP) を定義することで，Visual word の位置情報を活用した画像検索を実現している．オフセット空間において，相対位置関係が共通する Visual word の組み合わせは同一のビンに投票される．この性質を利用し，転置インデックスにオフセット空間における投票の概念を組み込むことで，Visual word の位置情報を考慮した画像検索を実現している．また同様に，GVP は MinHash による索引構築にも適用可能であることを指摘している．
(担当：安倍　満)

He らは，大量画像に対する類似検索を高精度かつ高速に行うための新たなハッシュ法を提案している．本論文では，ハッシュ関数による変換のゆがみを最小化する Spectral Hashing (SH) を元に，さらにハッシュバケット中のデータ数を均等にするため，ハッシュ値の相互情報量を最小化することで，探索速度の最適化も実現するよう拡張した，Similarity Preserving Independent Component Analysis (SPICA) 法を提案している．100 万 Web 画像などを用いた SH, LSH, Kernel LSH との比較実験により，Precision-Recall を用いた評価でも高い性能を達成しており，精度と速度のトレードオフの評価のための Recall-Time 評価でも，同等の計算時間でより高い精度を達成している．実験では画像を用いたが，論文本体は画像とは無関係であり，このような論文が CVPR のオーラルに採択されていることは興味深い．

ⓒ 2011 Information Processing Society of Japan

Vol.2011-CG-145 No.18
Vol.2011-CVIM-179 No.18
2011/11/18

Posters

Weinzaepfel ら[6] は，画像から得た SIFT 等の局所特徴群から，元の画像をそれなりに再構成する手法を提案している．方法は単純であり，対象画像の局所特徴群に対し，一定規模 (1000 強程度) の画像データベース中の最近接局所特徴量に対応する画像パッチを当該位置にはめ込む．これをすべての局所特徴について行い，重なったパッチは Poisson image editing 法でブレンディングするのみである．局所特徴量のみからとは思えないほどの再構成結果が示されている．画像データベースのインデックスにはこうした特徴量が用いられており，一般に特徴量のみからは元の画像は再構成できないことが暗黙の了解だが，この論文ではこの仮定に警鐘をならしている．問題設定が大変スマートである．

Chum ら[2] の研究では，画像を問い合わせとした物体レベルの画像検索において，問い合わせ中の物体の写りが悪い（一部しか写っていない）などの場合に対応するため，検索結果の上位の一部を仮想的に問い合わせ画像として使う問い合わせ拡張 (query expansion) を画像検索向けに実現している．当該ポスターでは，Jiří Matas 教授には多くの質問者が群がっていたので，横にいた主担当と思しき学生（多分，Andrej Mikulík?）に話を聞いた．初めは恥ずかしそうにとつとつと説明していたが，かなり突っ込んだ質問をしてみると，目が輝き出して発表とは関係のないことまで色々と話してくれた．ずいぶんと広く，深く検討をしているようで，感心した次第である．

Joly ら[3] は，高次元ベクトルデータの最近接探索を高速に行うための新たなハッシュ法を提案している．こうした手法の中では，LSH が有名であり，ランダムに発生させた方向への投影をハッシュ関数としているが，データ非依存のため性能が悪い．前出の SH はデータ分布に基づくハッシュ関数を生成可能だが，複数のハッシュ関数間の独立性が悪く性能低下の要因となる．そこで考案された方法は「目から鱗」的な手法であり，端的に言うとデータからランダムに抽出したサンプルにランダムに二値ラベルを振って SVM を学習させ，それをハッシュ関数とするものである．データ依存性と独立性を両立させたハッシュ関数が実現でき，高性能を達成している．これも画像とは無関係の論文である．

(担当：佐藤真一)

Oral Session 1B: Computational Photography

Lin らは，Affine Stiching と呼ばれる新たなパノラマ画像作成技術を提案している．ホモグラフィー変形に基づく従来のパノラマ技術が画像間の視差に対応できなかったのに対し，本技術は，画像間のグローバルな affine 変換となめらかに変化するローカルな affine 変換をうまく組み合わせることにより，異なる視点で撮影された画像間のパノラマを実現している．さらに，本技術は，物体遮蔽が観察されるシーンにも頑健に働くという特徴を持つ．

Rouf らは，写真などに輝きの印象を加えるため（クロススクリーン効果）に用いられる特殊なフィルタを用いて，高輝度情報を低ダイナミックレンジ画像に符号化し (glare encoding と呼ぶ)，複合することで高ダイナミックレンジ画像を生成する新たな技術を提案している．シャッタースピード変えて撮像された複数枚の画像を利用する従来手法に比べ，光学フィルタの特性を利用することにより，一枚の低ダイナミックレンジ画像から高ダイナミックレンジ画像を生成できる点が大きな特徴となっている．

Reshetouski らは，平面鏡により構成される Kaleidoscopic mirror system を考案し，この mirror system 内に物体を配置して 1 台のカメラを用いて撮像するという新たな multi-view 画像撮像システムを提案している．提案される Kaleidoscopic imaging システムを用いて撮像された単画像から，物体を取り囲む形で半球状に広がる視点から観察される画像を高画質で獲得することができる．単画像を用いるため，得られる multi-view 画像は完全に同期するという利点を持つ．

Posters

高解像度の multispectral 画像を獲得するための技術が２つ紹介されていた．Cao ら[1] は，低解像度の multispectral 映像と高解像度の RGB 映像を用いたハイブリッドカメラシステムにより高解像度の multispectral 映像を生成する技術を提案している．Cao らの手法では，まず入力 multispectral 画像と RGB 画像間で画素間の対応付けを行い，さらに色と空間的な距離の近さに基づき multispectral データをその他の画素に伝搬させていくアプローチを用いている．一方，Kawakami[4] らは，低解像度の hyperspectral 画像からシーンの分光分布を近似する基底を求め，この基底と高解像度の RGB 画像を用いて，高解像度の hyperspectral 画像を生成する新たな技術を提案している．

Liao ら[5] は，光源色を変更しながら撮像された画像列に基づき，シーンの直接反射と相互反射を分離する技術を提案した．光源色を変化させることがシーンのアルベドを変更した場合と同じ変化が観察されるという事実に基づき，シーンのアルベドを変化させて観察することが相互反射の分離に有効であることを示した．さらに，相互反射を除いた直接成分のみを用いることにより，照度差ステレオに基づき精度の高い形状推定が実現できることを示した．

(担当：佐藤いまり)

ⓒ 2011 Information Processing Society of Japan

Vol.2011-CG-145 No.18
Vol.2011-CVIM-179 No.18
2011/11/18

Oral Session 1C: Scene Understanding and 3D Structures

Gupta らは，従来の画像認識のように画像中のどこに何が存在するかを認識するのではなく，シーンの中において人が座れる場所や横になって寝られる場所を認識するという，人の行動に着目した画像認識手法を提案している．ここでは，まず，既存手法を用いて，1 枚の画像から屋内環境 (Manhattan world を仮定) の 3D 構造のボクセル表現 (occupancy map) を取得する．そして，事前にモーションキャプチャから得た人の特定の姿勢を表すボクセル表現について，その姿勢が入るフリースペースとその姿勢を支えるサポートがあるという拘束に基づき，シーン中で人がその姿勢が採れる場所を探している．

Geiger らは，道路環境認識において，道路を鳥瞰したときの道路のトポロジー (道路幅や分岐の形) を，Generative model を作成して動画像から推定する手法を提案している．観測データは，ステレオ画像から得られる道路上の 2D occupancy map と，移動体の特徴点 3D フローであり，これらはステレオ画像を用いた visual odometry の技術によって，フレーム間が関連づけられている．道路のトポロジーに関するパラメータと観測データを関連づける尤度関数をモデル化し，その関数に含まれる分布パラメータをデータベースを用いて学習したうえで，このモデルに基づいた MAP 推定の問題を，MCMC をベースにしたアルゴリズムを用いて解いている．

Guo らは，シーン中の影を検出する方法として，Mean shift によって過分割された領域をペア毎にチェックして，一方が影になっている同一素材か，同じ属性を持つ同一素材かを判断し，その情報に基づいて定式化したコスト関数をグラフカットによって解く方法を提案している．また，後段の matting によって，影および非影の属性をバイナリ値ではなく連続値で表現し，これに基づいて影を除去することにより，効果的な影の除去を実現している．

Crandall らは，大規模な SfM(Structure from Motion) の問題を効率的に解く方法として，バンドル調整を行う前に，カメラ位置に関する離散値推定問題を解くことでその初期値を求める方法を提案している．この初期値推定では，予め 2 枚の画像間のカメラ運動（回転と並進）が推定されており，かつ全画像の隣接関係がグラフ化されている画像群において，全カメラの位置を推定する問題を MRF(Markov random field) の最適化問題と考え，各カメラについて，隣接する画像間のカメラ運動が推定されたものと近くなるように，離散化された運動パラメータを BF(belief propagation) で解く．BF が並列処理に適していることや，コストにロバスト関数を利用していること，および，離散問題を解いた後に改めてカメラ運動のみを連続値推定するなどの工夫により，効率的でありながらロバストかつ高精度な初期推定が実現されており，その結果を用いたバンドル調整による最適化も短時間で実現できている．この論文は，Best Paper Honorable Mention の賞を獲得した．

Kowdle らは，平面で構成されたシーンを複数のカメラで再構成する際に，テクスチャレスであったり観測したカメラが少なかったことによって推定がうまく推定できなかった部分に対し，Active learning の枠組みを利用したユーザインタラクティブな手法によって適切に復元する手法を提案している．ここでは，過分割によって得られた superpixel を，グラフカットを用いて多数の平面候補の 1 つに割り当てるという再構成手法において，上述の理由により，どの平面ラベルにも属さなかった領域や，複数のラベルに属する可能性のある領域を検出する．そして，その領域に関連するエッジをユーザに与え，そのエッジの状態をユーザが入力することで，グラフのエッジの重みを変更して再推定を行う．
(担当：杉本　茂樹)

Oral Session 1D: Video Analysis

このセッションでは動画像解析について 5 件の発表がなされた．

Le らは，行動認識のために，従来の SIFT や HOG のように人が設計した特徴量ではなく，訓練サンプルから教師なしで学習する特徴量を提案している．独立成分分析を拡張した手法であるが，階層構造をもつ識別器へ組み込むことで，その単純さにも関わらず，公開データベース (Hollywood2, UCF, KTH, and YouTube) で高い認識性能を得ることができている．

Shahar らは，動画像の超解像に向けて，与えられた動画像内で繰り返される特徴量のみを利用する手法を提案している．この特徴量は，例えば，走る人や回転する羽など同じ動きを繰り返すものを撮影した動画から獲得される．動画像を時空間でパッチに区切り，類似するパッチを推定し，それら類似するパッチを特徴量として動きぼけを除去した鮮鋭な画像を生成している．

Ricci らは，動画像から類似性をもつ時空間パターンを，凸最適化問題で効率的に推定する手法を提案している．この論文は，道路の監視映像やスポーツ映像のように複雑な動画像を対象としている．公開データベース (APIDIS, Basket, and London's traffic) でクラスタリング性能が向上していることが確認されている．

Liu らは，行動認識に向けて，複数の属性を特徴量として識別時に利用する手法を提案している．人が決めた属性と，訓練サンプルから抽出した属性とを統合し行動を識別する．属性の自動抽出には latent SVM を適用している．公開データベース (Olympic Sports, UIUC, KTH, and Weizmman) で認識性能が改善していることが確認されている．

ⓒ 2011 Information Processing Society of Japan

Vol.2011-CG-145 No.18
Vol.2011-CVIM-179 No.18
2011/11/18

続いて Liu らは，複数カメラを用いた行動認識に向けて，カメラ間で関連する特徴量を教師無し学習で抽出する手法を提案している．従来手法は単一カメラの特徴量を Bag-of-Visual-Word で表していたが，提案手法はカメラ間の特徴量を二部グラフによる Bag-of-Bilingual-Words でバイリンガルに表している．公開データベース (IXMAS) で認識性能が改善していることが確認されている．
（担当：西山　正志）

Oral Session 2A: Object Detection

従来の能動学習手法は，研究者が事前に用意したデータセットにおいてしかテストを行わないため，実環境の応用では，精度，計算速度など多くの問題が残る．Vijayanarasimhan らは，Large-scale live active learning を提案し，実環境の応用問題に対しては，part-based 検出器と linear SVM，hash table の組み合わせで，実環境に適用している．

Ma らは，輪郭情報を用いた物体検出の手法を提案している．従来手法は，検出したい物体のエッジと背景のエッジが連接されている場合，対象を検出しにくい欠点がある．この問題に対して，改良した SC(Shape Context) 特徴を用いて，画像のエッジと検索対象物体の輪郭モデルの間に部分的な輪郭マッチングを行い，得られたマッチング結果から重み付けグラフを構築し，そのグラフの Maximum clique inference によって物体輪郭を検出するアルゴリズムを提案している．提案手法は，物体の位置と輪郭を同時に検出することを可能にしている．

Yang らは，姿勢検出に良く使われる Picturial structure を改善し，柔軟性が高い検出アルゴリズムを提案している．提案法は，Picturial structure の spring model にパーツ間の co-occurrence 関係を表現するパラメータを導入することによって，パーツ間のロカール剛性をモデル化でき，より高い精度で姿勢検出できる．
（担当：阮 翔）

Oral Session 2B: Optimization Methods

Shi らは，Sparse coding を利用した dictionary learning において，広く使われている convex な L_1 ノルム拘束の代わりに，より L_0 ノルム拘束に近い MCP(minimax concave penalty) を利用した方法を提案している．また，そのオンライン学習アルゴリズムを提示するとともに，画像ノイズ除去や inpainting 等のアプリケーションに適用し，提案手法の優位性を示している．

Savchynskyy らは，一般的な 2 階エネルギーの MRF を用いた多値ラベリングの大域的最適化問題において，LPR(Linear programming relaxation) を用いたアルゴリズムに着目し，近年提案された Nesterov の枠組みを用いて解く手法の改良版を提案している．この提案手法では，従来法における繰返しごとの解のステップサイズ，および，コストを平滑化して解く際の平滑化パラメータをを適応的に算出することで大幅な高速化を実現している．また，解の lower bound だけでなく，upper bound を算出する方法も示し，これにより従来法にはなかったアルゴリズムの停止条件を与えている．

Mukherjee らは，複数の画像から同一の前景物体を抽出する cosegmentation の問題において，既存手法を拡張した場合よりも，画像中の物体サイズの違いに影響されにくく (scale invaliant)，かつ計算コストの小さい方法を提案している．著者らは，scale invaliant 性を考慮したとき，各画像の前景のヒストグラムが正しい領域分割後には互いに線形結合で表されることに着目し，グラフカットを用いたセグメンテーションのためのエネルギー関数に，その線形性を示すランク 1 の行列を推定するためのコストを加える．この場合，そのエネルギー関数は non-convex となるが，これを単調減少性を補償するアルゴリズムによって最小化している．

Jegelka と Bilmes は，MRF を用いた二値のラベル付け問題において，特定エッジのカットが，他のエッジの重みに影響を与えるような非劣モジュラなエネルギー関数に対し，グラフカットをベースにした最適化手法を提案している．ここでは，ノードのラベル集合について劣モジュラ性を持つ平滑化項の代わりに，エッジのクラス集合について劣モジュラな拘束項を利用する．この結果，全体のエネルギー関数はノードのラベル集合について必ずしも劣モジュラ性を持たないが，内在する劣モジュラ性を利用して既存のグラフカットアルゴリズムをサブルーチンとして用い，その近似最適解を得ている．

Jiang らは，画像中の特徴点とその配置がグラフ構造で表現されている画像特徴どうしの回転およびスケール変化に不変なマッチングについて，LAT(lineary augmented tree) を利用した手法を提案している．提案手法では，回転とスケールの不変性および全体のグラフ整合性を保つための線形な拘束項を用いてこの問題を高階のエネルギー関数で表現し（このような拘束を LAT constraints と呼んでいる），その最小化問題を MILP(Mixed integer linear programming) の問題に変換する．そして，これを部分的に DP(Dynamic programming) で解くことにより，効率的に解を得ている．
（担当：杉本　茂樹）

ⓒ 2011 Information Processing Society of Japan

Vol.2011-CG-145 No.18
Vol.2011-CVIM-179 No.18
2011/11/18

Oral Session 2C: Segmentation and Grouping

このセッションでは領域分割について５件の発表がなされた．

Sundberg らは，オプティカルフローによる動き特徴を用いることで，隠れ境界と輪郭とを区別しながら動画像中の物体輪郭を検出する手法を提案している．この論文では，隠れ検出のアルゴリズムが新しく，従来手法と比較して高い精度で輪郭を抽出できている．

Prisacariu らは，従来のレベルセットに基づく物体領域分割に向けて，形状情報を加えた手法を提案している．領域分割のために，形状を楕円フーリエ記述子で表し，ガウス過程潜在変数モデルで次元削減をした上で，エネルギー最小化を解く．追跡処理も行うことで，物体の運動と形状変化に合わせた領域分割ができていることを実験で示している．

Bertelli らは，教師ありの領域分割に向けて，物体という高次の情報と画像特徴という低次の情報を統合する手法 (kernelized structural SVM) を提案している．この統合のためのカーネルとして，物体の見え方，物体の形状，画像間の色分布，全体の色分布の４つを用いる．潜在的な分割誤りを避けるように学習できる利点がある．

Glasner らは，動画像の連続したフレームのように，相関のある２枚以上の画像を同時に満たすよう領域を分割するため，輪郭形状に基いた教師無しの手法を提案している．形状に基づく同時クラスタリングを quadratic semi-assignment problem として解いている．領域分割の性能を Stein らの隠れ輪郭データベースで評価している．

Shotton らは，Best Paper を受賞しており，一枚の奥行き画像から個々の身体部品を推定し，それらの部品の組み合わせから姿勢を決定する手法を提案している．姿勢推定に用いる識別器の訓練サンプルをコンピュータグラフィックスで合成しており，生成された大量の訓練サンプルで識別器を学習している．既に実用化されており，家庭用ゲーム筐体の Xbox と Kinect でリアルタイムに動作する．

（担当：西山　正志）

Oral Session 2D: Motion and Tracking

Liu らは追跡対象物体を細かなパッチの集合により表現し，これらのパッチ集合を辞書中の少数の基底を用いて表現（sparse coding）したモデルを用いて頑健な追跡を実現している．ただし，ローカルなパッチのみでは対象の大域的構造を表現できないため，基底パッチが対象物体の中でどのように分布しているかを表現した sparse coding histogram を利用することにより，対象の大域的構造を考慮した追跡を実現している．また，物体追跡においては時間経過とともに物体の見えが変動するため，時刻に合わせてモデルの更新を行う必要があるが，これはパッチの表現に用いる辞書データを更新することにより実現されている．この更新には蓄積された画像列を効率的に表現可能な K 枚の画像を選択する K-selection と呼ぶ方法が用いられている．

Brendel らはビデオ画像上の複数物体の同時追跡の問題が，グラフ構造における Maximum-weight independent set（MWIS） を探索する問題と等価であることを示し，それを用いた物体追跡法を提案している．複数物体の追跡問題は，各フレームでの独立な検出結果をどのように対応づけるかととらえることができる．ここでは，連続するフレームにおける検出結果のペア（tracklet）をノードとし，各ノードにペアの類似度に基づく重み付けがなされている．また，同一の検出結果を含む tracklet が接続されグラフ構造により表現される．さらに，時刻 t と $t+1$ のグラフ，$t+1$ と $t+2$ のグラフは互いに独立したサブグラフとされる．この場合，物体の追跡を行うことは隣接するサブグラフにおいて，tracklet を矛盾なく対応付けることと考えられる．これが MWIS と等価であることが証明され， MWIS のアルゴリズムを用いた物体追跡が実現されている．

（担当：坂上 文彦）

Oral Session 3A: Object Recognition

Kulkarni らは，1 枚の画像からその画像を説明する文を自動生成する手法を提案した．従来手法の限界を越えて，複数の物体，その修飾語句ならびにそれらの関係を記述でき，より詳細な画像の説明文が得られる．文の生成には Conditional Random Field を用いる．

従来,「馬に乗る人」のように単一の物体よりも複雑なものを表すためには,「人」と「馬」といった個々の物体に分解し，独立にモデル化していた．物体単位のモデル化はサンプル収集を容易にするため，学習に有利な方策であった．しかし,「人」と「馬」の見た目を組み合わせても「馬に乗る人」にはならないため，Sadeghi らは「馬に乗る人」を直接表す visual phrase を導入した．機械翻訳の知見を取り入れることで，少数のサンプルで学習した visual phrase 検出器は大量のサンプルで学習した物体検出器より高い精度を実現した．この論文は Best Student Paper に選ばれた．

静脈パターンは生体認証に適した特徴であるが，静脈は皮膚に覆われているため，取得するのに赤外線やレーザーを使用する必要があった．Tang らは犯罪捜査を目的として，通常のカラー画像から静脈パターンを推定する方法を提案した．犯罪捜査目的とは，照明の種類とカメラの応答関数が既知であることを意味すると思われる．提案手法は静脈パターンの推

ⓒ 2011 Information Processing Society of Japan

Vol.2011-CG-145 No.18
Vol.2011-CVIM-179 No.18
2011/11/18

定に光学的ならびに生物物理学的な知見を利用するため，従来手法のように写真撮影時に白色光を用いる必要がない．

物体認識において，学習時と認識時で物体の向きや撮影機材などが異なることがある．このような場合に認識を可能にするドメイン適応問題において，Kulis らは学習時と認識時で特徴量の次元数が異なる場合に適用可能な特徴量の非線形変換を求めた．

Baboud らは山岳写真に山の名前や標高などの注釈を付けるために，写真の撮影位置を推定する手法を提案した．GPS 搭載カメラの使用を想定し，撮影位置はわかるが撮影方向はわからないという問題設定である．撮影方向（3 次元の回転）は，写真中の山の形状と地形データを照合することで求める．山の形状は雲や雪，霧などの影響を受けて変化するが，それらに頑健なエッジ照合手法を提案している．応用として，地形データから合成した稜線画像に山岳写真をはめ込むことができる．
（担当：岩村 雅一）

Oral Session 3B: Image Modeling

Chandraker らは等方性の BRDF により表現可能な物体における不変量を示し，これを用いた形状復元法を提案している．ここでは，光源がカメラの光軸まわりの円上を移動する場合，画像の時間微分が BRDF の特性に関わらず位置（法線）情報のみに依存することが示され，これを用いた形状復元法が提案されている．

Tian と Narasimhan はカメラを用いた書面スキャナのための画像補正法を提案している．ここではまず，文字どうしの類似性に着目することにより書面の水平方向検出している．また，多くの文字には縦方向のストロークが多く含まれることに着目し，書面の垂直方向を検出している．最後にこれらの方法により得られる書面上のグリッド構造から書面の 3 次元形状を推定し，推定された形状を用いて幾何学的，光学的な補正を行い目的の画像を取得している．
（担当：坂上 文彦）

Oral Session 3C: Statistical Methods and Learning

Jain らは，顔検出を例とし，Online domain adaptation のアルゴリズムを提案している．検出信頼度が低いのは，その検出が誤検出である可能性が高いと仮定し，信頼度が低い検出に対して，他の信頼度が高い検出結果との類似度を用いて Gaussian Process Regression で信頼度を更新することによって検出精度を改善している．

Zhang らは，顔の写真-スケッチ (Photo-sketch) 認識のアルゴリズムを提案している．画像とスケッチを同じ modality 空間に変換してマッチングするのは，殆どの従来手法の考え方である．それと違って，本研究の基本アイディアは，特徴抽出段階に modality gap を埋めることにある．coupled information-theoretic encoding を用いた新たな顔表現を提案し，写真-スケッチ技術を実用レベルの 98.7% まで改善している．また，著者らは，FERET データベースにおける計 1194 人のスケッチを含む世界最大の顔スケッチデータベースを作成している．
（担当：阮 翔）

Oral Session 3D: Applications

Zhang と佐藤は，蛍光について分析し，蛍光の色の見えは照明に影響されないことと，反射および蛍光成分をもつ物体の色の見えはそれらの線形結合で表されることを示した．また，未知の 2 光源下で撮影された画像において独立成分分析を使うことで，蛍光成分と反射光成分を分離する方法を提案している．この論文は Best Student Paper Honorable Mention を受賞した．

例えば同じ物体を様々な方向から写した画像など，いくつかの画像の集合で対象を表し分類に利用する画像集合分類 (image set classification) において，Hu らは画像を高次元ベクトル空間中の点と考え，一つの対象を表す全ての点のアフィン包によって対象を表すことを提案している．これは例えば異なる 2 点ならそれらを通る直線，同一直線上にない 3 点ならそれらを通る平面を意味する．その上で著者らはアフィン包間の距離として Sparse Approximated Nearest Point (SANP) というものを定義することで画像集合分類への応用を実現し，顔認識の実験により有効性を示している．
（担当：石川 博）

会場の雰囲気など

気　　候

今回の CVPR 会場となったコロラドスプリングスはアメリカ中西部に位置し，日本からの乗り継ぎはサンフランシスコかロサンゼルス経由になる．しかしそのアメリカ国内乗継便も 4 時間程度かかる．日中は晴れれば 30 ℃，夜間は 15 ℃程度に冷え込むが，学会会場はいつものごとく冷房が効きすぎて寒いほどである．おおむね晴れて天候には問題なかったものの，学会初日 6/20 朝はあいにくの雨であった．これはサンフランシスコからコロラド付近までアメリカ西部には thunder storm が発生した影響であるが，そのせいでアメリカ国内乗

ⓒ 2011 Information Processing Society of Japan

Vol.2011-CG-145 No.18
Vol.2011-CVIM-179 No.18
2011/11/18

継便が欠航し，コロラドスプリングスまでたどり着けない参加者が多数いた模様である．多くは 6/21 中には到着したようであるが，その影響か 6/21 午前の CVPR poster には no show が 3 件あったのが残念である．

ポスター会場

休憩が終わる前に Poster 会場に直行しないと，見たいポスター発表に人が集まってしまう．poster 開始後は，どの poster にも発表者に近寄れないくらい人が集まってしまう状況であった．毎回の poster session には最大 44 件発表があり，全ての poster にそれぞれ常時 10 人は張り付いているため，400 人以上が poster 会場に集まっていることになる．それ以外の 400 人（推定）は，poster と関係なくあちこちで議論し，くつろぎ，ノート PC で仕事をし，論文を読み，コーヒーを飲んでいた．

オーラル会場

メインの oral セッションの 300 席程度ある会場は，oral が始まる前に満席になった．基本的に前のほうから席が埋まっていくため，早めに行って前列の席を確保しなければ，最後尾から豆粒程度の発表者を見るしかないという状況であった．

Video overflow

2010 年と同様に，video overflow という予備の部屋が用意されていた．参加者があまりに多いため，oral の 2 会場では収容しきれない（数字の上では）．そこで別の部屋を用意して，そこに speaker と slide の映像を中継するというものである．Oral 会場では席がシアター形式（机なし）のためメモが取りにくいが，video overflow 会場には机もあり人も少ないため，実は快適だろうと思い，試しにこのセッションで行ってみた．

残念なことに，スライド中継の映像コーデックが悪いらしく，画面の切り替わりやスライド中のアニメーションなど，映像中に motion があるとブロックノイズが発生し，全然スライドの内容が分からなかった（静止スライドだけならまだましだったが）．また speaker の映像と slide の映像にタイムラグがあり，どこを説明しているのかわからないこともあった．ということで，快適なはずの video overflow 会場はストレスがたまる一方であった．

雑　　感

CVPR の発表件数が 400 件程度に対して，参加者は 1500 人．発表者以外の見学が多数いるだろうこの状況は，トップコンファレンスとして見に行く価値がある会議であることを意味している．空港に着いてホテルまでタクシーを相乗りした韓国人の Ph.D. 学生も，教授の代わりに見学しに来たと言っていた．

日本人参加者は本当に少ない．多く見積もっても 50 人程度で，アジア系の顔を見かけたら，圧倒的にそれは中国人か韓国人である．アメリカ在住だけでなく，中国本土の大学からも多数来ていた．参加者数は中国人が多いようであるが，シンガポールも存在感を増してきている．CVPR2010 や ICCV2009 などでも NUS や NTU などの発表が多く，CVPR2012 の area chair にはシンガポールから 2 名入っている．日本人の area chair は，今年も来年も 1 名ずつ．この数字だけならまだ他のアジア諸国には負けていない．

（担当：玉木　徹）

おわりに

以上 CVPR2011 について，出席者有志により概要をまとめた．紙面の都合上，基本的に口頭発表論文の報告となったが，他にも多くの注目すべき発表があった．プロシーディングスを参照されたい．

次回 CVPR2012 は，Rama Chellappa，Benjamin Kimia，Song Chun Zhu が General Chair となって，2012 年 6 月 18 日〜20 日の日程で，ロードアイランド州プロビデンスにて開催の予定である．

（文責：石川 博）

参　考　文　献

1) X. Cao, X. Tong, Q. Dai, and S. Lin, “High Resolution Multispectral Video Capture with a Hybrid Camera System,” Posters 2A.
2) O. Chum, A. Mikulik, M. Perdoch, and J. Matas, “Total Recall II: Query Expansion Revisited,” Posters 2C.
3) A. Joly and O. Buisson, “Random Maximum Margin Hashing,” Posters 2C.
4) R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, “High-resolution Hyperspectral Imaging via Matrix Factorization,” Posters 2A.
5) M. Liao, X. Huang, and R. Yang, “Interreflections removal for photometric stereo by using spectrum-dependent albedo,” Posters 3A.
6) P. Weinzaepfel, H. Jegou, and P. Perez, “Reconstructing an image from its local descriptors,’ Posters 2A.

ⓒ 2011 Information Processing Society of Japan