

CVPR2011 報告

玉木 徹†

† 広島大学大学院工学研究科情報工学専攻 〒739-8527 広島県東広島市鏡山 1-4-1
E-mail: †tamaki@hiroshima-u.ac.jp

あらまし CVPR2011 の報告 . 修正版 .
キーワード CVPR

A report on CVPR2011

Toru TAMAKI†

† Hiroshima University 1-4-1 Kagamiyama, Higashi hiroshima, Hiroshima, 739-8527 Japan
E-mail: †tamaki@hiroshima-u.ac.jp

Abstract A CVPR2011 report, corrected edition.

Key words CVPR

1. Introduction

これは, Colorado Springs で開催された CVPR2011 (the th IEEE Conference on Computer Vision and Pattern Recognition) の報告である .

- 開催場所: Crown Plaza Colorado Springs, Colorado Springs, USA

- 開催期間: 20-25 June 2011

初日 6/20 と最後の 2 日間 6/24,25 は workshop と tutorial に割り当てられ, main conference は 6/21,22,23 の 3 日間にわたって開催された . セッション構成は 3 日間とも同じ .

- breakfast
- poster A (1 時間半)
- break
- oral A&B の 2 パラレルセッション (100 分)
- lunch
- poster B (1 時間半)
- break
- oral C&D の 2 パラレルセッション (100 分)
- poster C
- reception (main conference の 1 日目と 2 日目のみ)

Oral は 1 セッション 5 件 × 20 分なので, 毎日 20 件の講演 (3 日間の合計 59) . Poster は 1 セッション最大 44 件と demo 数件で, 毎日のポスターは 120 件を超えた (合計 379) . 参加者は事前登録が 1500 人となったら打ち切りで, on-site registration はないと web でアナウンスされていた . 昨年の CVPR2010 では事前登録で 1800 人を超えて急遽登録を締め切ったため, その経験が活かされていたと思われる .

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ら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, その影響が 6/21 午前の CVPR poster には no show が 3 件あったのが残念である .



学会会場の Crown Plaza Colorado Springs

以下では, main conference で聞いた oral session の内容, poster presentation のいくつかを紹介する . ただし, 紹介する oral/poster は単にその内容が聞けたというだけで紹介する

のであって、お勧めする内容だからではないことをご了承願いたい。また実際の論文は一切読まず、スライドやポスターを見聞きした情報だけから内容を書いているため、多分に内容が間違っていたり、偏っていたり、諦めていたりするかもしれないがご容赦願いたい。正しい内容を知りたい方は、CVPR2011の proceedings を入手するか、CVPR on the web^(注1) を利用するか、毎年恒例の CVIM での会議参加報告^(注2) を待つか、コンピュータビジョン勉強会@関東の CVPR 祭り^(注3) もしくは名古屋 CV・PRML 勉強会の CVPR2011 論文紹介^(注4) に参加していただきたい。

2. Registration

今回から CVPR も USB proceedings に変更。Springer のバッグはペラペラですぐに使い物にならなくなりそう。T-shirt は背面に過去 20 年の CVPR 史が掲載。



Conference bag, USB proceedings, T-shirt (front & back), program booklet

3. 21th June, 2011: 1st Day

3.1 Breakfast

毎日提供（おそらく workshop days も？）. 朝 8:30 の poster session が始まる前からあちこちで議論が始まっている。内容はコーヒー・ジュース・ベーグルやドーナツ系の食べ物・バナナやオレンジやリンゴ丸ごと・カップヨーグルト等。



学会会場の breakfast .

3.2 Posters 1A : Statistical Methods and Learning

Breakfast 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。Poster session 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る。さらにそれ以外の 400 人（推定）はどこか観光に行っている？



Poster 会場の 1 つ .

開始前（人がパラパラ）と開始後（黒山の人だかり）.

3. A Scalable Dual Approach to Semidefinite Metric Learning; Chunhua Shen, Junae Kim, Lei Wang

マハラノビス距離を用いた metric learning を Lagrange dual で高速化 .

6. Comparing Data-Dependent and Data-Independent Embeddings for Classification and Ranking of Internet Images; Yunchao Gong, Svetlana Lazebnik

データ依存・非依存の embedding を比較検討。実データでは Manifold の仮定が成り立たないこともあると報告 .

15. Graph Embedding Discriminant Analysis on Grassmannian Manifolds for Improved Image Set Matching; Mehrtash Harandi, Sareh Shirazi, Conrad Sanderson, Brian Lovell

Subspace を Grassman 多様体上の点で表現した判別分析 .

18. Learning invariance through imitation; Graham Taylor, Ian Spiro, Rob Fergus, Christoph Bregler

ユーザにデータと同じポーズをとって webcam で撮影してもらって、様々なデータを生成。Mechanical Turk の映像版？

19. Learning Message-Passing Inference Machines for Structured Prediction; Stephane Ross, Daniel Munoz, J. Andrew Bagnell

(注1): <http://www.cvpapers.com/cvpr2011.html>

(注2): 石川博 他, 「CVPR2011 参加報告」, 情報処理学会 CVIM 研究会, 2011 年 11 月予定.

(注3): <http://sites.google.com/site/cvsaisentan/>

(注4): <https://sites.google.com/site/nagoyacv/>

グラフィカルモデルではなく、パッチの上下左右の関係を反復学習。

23. Max-margin Clustering: Detecting Margins from Projections of Points on Lines; Raghuraman Gopalan, Jagan Sankaranarayanan
サンプル点を直線に投影してマージンを計算する。

26. Modeling the joint density of two images under a variety of transformations; Joshua Susskind, Roland Memisevic, Geoffrey Hinton, Marc Pollefeys

2つの似た画像間の関係を学習し、morphlet というものを作成。

27. Multi-label Learning with Incomplete Class Assignments; Serhat Bucak, Rong Jin, Anil Jain

マルチラベルの各学習サンプルに、正しいラベルのうち幾つかしかつけられていない状態で学習。

28. Multi-layer Group Sparse Coding – for Concurrent Image Classification and – Annotation; Shenghua Gao, Liang-Tien Chia, Ivor W. Tsang

クラスラベルとタグの依存関係をグループスパース学習。

29. Multifactor Analysis Based on Factor- Dependent Geometry; Sung Won Park

pose, light などの submanifold (geometry) を保存するような因子分析。

40. Supervised Local Subspace Learning for Continuous Head Pose Estimation; Dong Huang, Markus Storer, Fernando DelaTorre, Horst Bischof

頭部の manifold を区分的に subspace で表現。

41. TaylorBoost: First and Second-order Boosting Algorithms with Explicit Margin Control; Mohammad Saberian, Hamed Masnadi-Shirazi, Nuno Vasconcelos

通常の boosting のように best update なものを選ぶという間接的な方法ではなく、best weak learner を直接選ぶ。

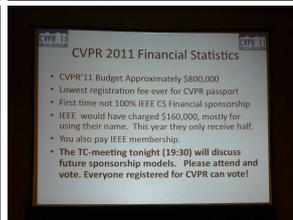
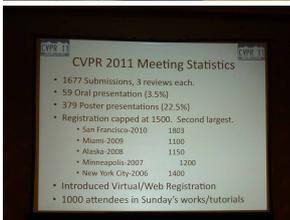
no show

31, 38, 43.

3.3 Break

Oral と poster の間に毎回提供される coffee break . でも内容は breakfast と基本的に同じ .

3.4 Welcome address



General Chair の Terrance Boult (University of Colorado at Colorado Springs, UCCS) , CVPR statistics

3.5 Orals 1A : Image and Video Retrieval

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 . 基本的に前のほうから席が埋まっていく . 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 .



Oral 会場の 1 つ . 開始前 (ガラガラ) と開始後 (満席) .

1. Towards Cross-Category Knowledge Propagation for Learning Visual Concepts; Guo- Jun Qi, Yong Rui, Qi Tian, Thomas Huang
クロスカテゴリの識別器とそのアンサンブルを学習 . 学習サンプルのラベルの出現頻度はバラバラなので、ラベルをあるカテゴリから別のカテゴリへ transfer してしまうというのがモチベーション . 従来は Category-level の相関を用いていたが、これを Instance-level, Inter-Instance-level に拡張 . Intra/Cross-category でラベルを伝播する . Flickr と LSCOM-Lite video dataset で実験 .

早口・スライド多数・時間超過のため、内容はほとんど理解できず .

2. Image Retrieval with Geometry-Preserving Visual Phrases; Yimeng Zhang, Zhaoyin Jia, Tsuhan Chen

Bag-of-Visual-Words(BoW) は空間情報を持っていないので、BoW 同士を組み合わせと位置関係のヒストグラムを表す Bago-of-Phrase(BoP) を提案 . しかし組み合わせの数は指数関数的に増えるため、コストが大きい . そこで、BoW に用いられていた Inverted Index を拡張し、BoP のための Inverted Index である Offset Space を提案、これが $O(M)$ で計算できる . Inveted Index ではなく Min-hash にすると、さらにコストが削減される . Univ. Kentucky Dataset で実験 .

3. Image Ranking and Retrieval Based on Multi- Attribute Queries; Behjat Siddiquie, Rogerio Feris, Larry Davis

Query として「Female White hair」などの複数の attribute を入力するときの検索・ランキング方法の提案 . アイデアは、attribute 間の相関をモデル化すること . Labelled Faces in the Wild (LFW) で実験し、attribute の共起性を解析 .

口ごもった話し方で、ほとんど内容は聞き取れず .

4. Iterative Quantization: A Procrustean Approach to Learning Binary Codes; Yunchao Gong, Svetlana Lazebnik

大量データを扱うときにメモリに乗せるには、画像を binary code にする . そのための特徴量の 2 値化手法の提案 . 特徴には GIST , 学習には PCA + ITQ (Iterative Quantization) , 認識にはハミング距離 . まず特徴量を PCA で圧縮し、その結果を「量子化誤差が最小になるように」 R で回転 . そのための目的関数 $\min_{B,R} \|B - VR\|$ を、 R と B を反復的に推定 . R を求めるときには Procrustes Analysis を使用 . 実験では 16~256bit に圧縮し、評価 . ラベルを用いるために PCA を CCA に取り換えた手法も提案 .

5. Compact Hashing with Joint Optimization of Search Accuracy and Time; Junfeng He, Regunathan Radhakrishnan, Shih-Fu Chang, Claus Bauer

最近傍探索のために ANN や LSH があるが、どれも accuracy だけを最適化している。この手法は、検索時間も同時に最適化する。worst case と average case の検索時間のモデル化を行い、hasing の bucket をバランス化することで最適化できることを示した。

Nvidia Talk

Tegra for Mobile Imaging, Kari Pulli

Tegra チップの紹介。Tegra2 は現在タブレット製品に乗っているが、次世代の Tegra3 にはどんな機能を乗せたらよいか? Computational camera である Fcam (Frankencamera) や、OpenCV の他にも proposal を募集中。

3.6 Lunch

毎日提供 (おそらく workshop days も?)。Oral session が終了すると、3 か所ある lunch buffet にはすぐに長蛇の列ができた。食事の内容は毎日変わるものの、基本的に buffet であることには変わらず、breakfast や break の内容ともそれほど大差ないため (量は違うが)、2 目にはもう飽きてきた。

これまで CVPR では lunch を提供したことはなかったが、会場周辺に 1500 人の食事を提供できる飲食店がないため (結構な田舎)、今回は lunch 提供に踏み切ったのだと予想される。ちなみに breakfast (コーヒー・ドーナツ系・フルーツ丸ごと) は例年提供されていた。



学会 Lunch 会場は 3 か所。どこも満員、長蛇の列。

3.7 Posters 1B : Color and Texture, Document Analysis, Segmentation and Grouping

2. Inertial sensor-aligned visual feature descriptors; Daniel Kurz, Selim Ben Himane

センサでモバイルデバイスの方向を検出し、descriptor の方向を調整。

4. Multi-spectral SIFT for Scene Category Recognition; Matthew Brown, Sabine Susstrunk

カラー SIFT を RGB-NIR (近赤外) 画像用に拡張。

6. Unsupervised Local Color Correction for Coarsely Registered Images; Miguel Riem de Oliveira, Angel Sappa, Vitor Santos

解像度・大きさの異なる画像レジストレーションの色合わせ。

10. Style Transfer Matrix Learning for Unsupervised Writer Adaptation; Xu-Yao Zhang, Cheng-Lin Liu

各個人の特徴ベクトル x を変換するスタイル行列 A を推定し、 $A^T x$ で個人差のない特徴空間に変換。

14. A Global Sampling Method for Alpha Matting; Kaiming He, Christoph Rhemann, Carsten Rother, Xiaoou Tang, Jian Sun

背景と前景をペアにして特徴空間で表現し、サンプリング。

18. Efficient MCMC Sampling with Implicit Shape Representations; Jason Chang, John Fisher III

MCMC でサンプルした分布から、形状などの統計量を計算。

20. Entropy Rate Superpixel Segmentation; Ming-Yu Liu, Oncel Tuzel, Srikumar Ramalingam, Rama Chellappa

目的関数に entropy rate を用いた segmentation 手法。

21. Foreground Segmentation of Live Videos using Locally Competing 1SVMs; Minglun Gong, Li Cheng

各画素が 1-SVM を持つ。学習には周辺画素の情報も利用。

23. From Active Contours to Active Surfaces; Akshaya Mishra, Paul Fieguth, David Clausi

2D 用の Decoupled Active Contour (DCA) を拡張し 3D 用の Decoupled Active Surface (DCS) を提案。

24. From Co-saliency to Co-segmentation: An Efficient and Fully Unsupervised Energy Minimization Model; Kai-Yueh Chang, Tyng-Luh Liu, Shang-Hong Lai

co-saliency を利用して co-segmentation を行う。

25. Graph Connectivity In Sparse Subspace Clustering; Behrooz Nasihatkon, Richard Hartley

subspace の sparse clustering において、グラフの連結性を neighborhood cone で表現。

なぜか Rene Vidal が発表者のように机に座って、逆に Richard Hartley が立って議論していた。

27. Learning to Find Occlusion Regions; Ahmad Humayun, Oisín Mac Aodha, Gabriel Brostow

次フレームで隠れる画素を、学習により判別。

28. Majorization-Minimization mixture model determination in image segmentation; Giorgos Sfikas, Christophoros Nikou, Nikos Galatsanos, Christian Heinrich

Expectation-Maximization がクラスタ数を必要とするのに対し、Majorization-Minimization はクラスタ数も推定する。

31. Nonparametric Density Estimation on A Graph: Learning Framework, Fast Approximation and Application in Image Segmentation; Zhiding Yu, Oscar Au, Ketan Tang

3D 点などを表現するグラフの密度推定 (mode seek)。

33. Object Cosegmentation; Sara Vicente, Carsten Rother, Vladimir Kolmogorov

通常の co-segmentation は同じものしか検出しないが、これは intra-class variation が大きい場合の拡張。

38. Shape Based Pedestrian Parsing; Yihang Bo, Charles Fowlkes

人を検出して、頭部・上半身・下半身に分割。

39. Shape Grammar Parsing via Reinforcement Learning; Olivier Teboul, Iasonas Kokkinos, Panagiotis Koutsourakis, Loic Simon, Nikos Paragios

shape grammar を利用して、facade 画像を構文解析 (領域分割)。

44. Using Global Bag of Features Models in Random Fields for Joint Categorization and Segmentation of Objects; Dheeraj Singaraju, Rene, Vidal

Bag-of-features を用いて、認識だけでなく画素単位の領域分割を行う。

3.8 Orals 1C : Scene Understanding and 3D Structures

1. From 3D Scene Geometry to Human Workspace; Abhinav Gupta, Scott Satkin, Alyosha Efros, Martial Hebert

室内シーンを認識するとき、椅子やソファなどの認識ではなく、3次元形状復元ではなく、どこに座れるか?を認識する。Human Centric Scene Understanding を提案。つまりアフォーダンスで、カテゴリは「chair」ではなく「sitting」。人体モデルと室内の3次元復元を用いて、人体モデルが動けるかという free space constraint, 人体を支えられるかという support constraint などを利用。結果では、画像を sitting/sitting without back support/laying などに領域分割。

2. A generative model for 3D urban scene understanding from movable platforms; Andreas Geiger, Martin Lauer, Raquel Urtasun

車両の自動操縦を目的としている。ゴールは、車載カメラの short video から、シーンの topology (交差点が T 字路か十字路か), geometry (道幅, 交差点の中心点等), semantics (車両が右折しようとしているのか等) を認識。確率モデルには Dirichlet process + Gaussian Process を用い、推定には Reversible Jump MCMC を利用。

3. Single-Image Shadow Detection and Removal using Paired Regions; Ruiqi Guo, Qieyun Dai, Derek Hoiem

単一画像中の影領域を検出、除去。まず多数の領域に分割し、領域を SVM で分類、影領域に hard に分類し、最終的に soft shadow matting で影領域を検出、除去。SVM で領域を分類するとき、単一の領域を対象とせず、領域のペアに対して輝度値の差があるかどうかを分類の特徴に用いている。

4. Discrete-Continuous Optimization for Largescale Structure from Motion; David Crandall, Andrew Owens, Noah Snavely, Daniel Huttenlocher

photo tourism と bundle を生み出した Noah Snavely のチーム。SfM は $\min f(R, T, P)$ の最適化を必要とし、計算量が大きい。一般的な手順は、feature detection, feature matching, initialization, bundle adjustment (BA) となる。initialization には Incremental BA が用いられるが、scalable ではない。そこで camera pose を一度に推定する方法として、カメラをノードとするグラフを生成し、離散最適化である MRF で解く方法を提案。カメラ間の対応点はリンクである binary term に、カメラの GPS 位置情報や消失点はノードである unary term に設定。2枚からの SfM で R_{ij} と T_{ij} を求めた後、最適な R_i, T_i と R_j, T_j を推定する。この大規模 MRF の計算には離散 loopy BP を用いている。

5. Active Learning for Piecewise Planar 3D Reconstruction; Adarsh Kowdle, Yao-Jen Chang, Andrew Gallagher, Tsuhan Chen

Multi-view Stereo (MVS) において対応点があいまいなテクスチャがない領域で、高レベルの知識を利用する。つまり、あるあいまいなエッジが平面上にあるのか、物体の端にあたるのか、遮蔽輪郭なのかを、ユーザに入力を求める。これは active learning の枠組みであり、それを 3次元復元に利用。

3.8.1 Video overflow

昨年から Video Overflow という予備の部屋が用意されている。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 の映像を中継していた。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、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ひ、試しにこのセッションで行ってみた。



Oral の Video Overflow .

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良かった)。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だったので、その後は Oral 会場で席を確保するように努めた。

3.9 Posters 1C : Object Recognition, Image based Modeling, Human ID

なぜか Fei-Fei Li と Silvio Savarese が手をつないで poster 会場をルンルンと歩いていた。きっと何かのジョークに違いない。

3. Classification with Invariant Scattering; Joan Bruna, Stephane Mallat

同じクラスの画像 f, g があるときに $\min \|\Phi(f) - \Phi(g)\|$ を最小にする Φ を求める。かつ、連続で discriminative であることを要求する。

4. Clues from the Beaten Path: Location Estimation with Bursty Sequences of Tourist Photos; Chao-Yeh Chen, Kristen Grauman

写真撮影する旅行者が同じ経路をたどりやすいという情報を位置認識に利用。

5. Combining Randomization and Discrimination for Fine-Grained Image Categorization; Bangpeng Yao, Aditya Khosla, Li Fei-Fei

discriminative な領域を見つけるために、dense な特徴量 + ランダムフォレスト + discriminative 決定木を利用。

7. Deformation and Illumination Invariant Feature Point Descriptor; Francesc Moreno

HKS, Laplace-Bertalmio 作用素を用いて、非剛体変形と照明

変化に不変な detector を開発 .

12. Exploring Knowledge Transfer and Zero-Shot Learning in a Large-Scale Setting; Marcus Rohrbach, Bernt Schiele, Michael Stark

knowledge transfer のいろいろな手法を評価 .

13. Exploring Relations of Visual Codes for Image Classification; yongzhen Huang, Kaiqi Huang, Tieniu Tan

Visual words のグラフを作り, そのグラフのエッジに沿った visual word のペアに対して voting して特徴ヒストグラムを作成 .

24. Rank-SIFT: Learning to Rank Local Interest Points; Bing Li, Rong Xiao, Zhiwei Li, Rui Cai, Bao-Liang Lu, Lei Zhang

SIFT 特徴量を, 何枚の画像に登場するのかでランキング .

29. Where's Waldo: Matching People in Images of Crowds; Rahul Garg, Deva Ramanan, Steve Seitz, Noah Snavely

MVS の手法を利用して, 人物の同定 .

30. A Complete Statistical Inverse Ray Tracing Approach to Multi-view Stereo; Shubao Liu, David Cooper

MVS において, 多方向からの visibility の同時確率を計算 . PMVS2 と比較 .

31. A general method for the Point of Regard estimation in 3D space; Fiora Pirri, Matia Pizzoli, Alessandro Rudi

視線方向推定において眼球の dynamics を利用 .

3.10 Banquet / reception

学会には付き物の懇親会 . これも Breakfast/break/lunch と同じ場所で, 同じ形式 . やっぱり長蛇の列で, どこも満席 . 1500 人に食糧を提供 (食事ではない) しなければならないのだから, 仕方がないと諦め .

3.11 PAMI TC meeting

CVPR に参加する楽しみの一つが PAMI TC meeting . 楽しみにしている参加者はそれほど多くないのかもしれないが, ここで top researcher 達が研究以外の議論を交わしているのを見るのは, 結構楽しい . また, conference や journal の情報が手に入るのでも有益なこともある . banquet でアルコールが入った後の夕方 7:30 から 9:30 まで, 眠たいのによく議論するなど感心する . それだけ, top researchers は top conferences を自分たちのことだと感じ, よいものであり続けようとしている証拠だろう .

以前に参加した meeting は, CVPR2009 マイアミ (やはり banquet 後 10 時まで), ICCV2009 京都 (夕食がでなかったのに空腹を我慢しながら夜 8 時まで), CVPR2010 サンフランシスコ (昼食時にサンドイッチを配布, これはよかった) . これが 4 度目となるので, 大体どんな議題が出るのかと, 過去の経緯も把握できていた .

毎回の議題 (agenda) は web に掲載され^(注5), また mail list^(注6) でも流れている . この mail list は誰でも登録できる . PAMI TC meeting は, その学会参加者ならだれでも参加してよい . 議題はその場の参加者の挙手で決定される . オープンでかつ透明性が高い, 日本ではありえない意思決定の場である .

(注5): <http://www.computer.org/portal/web/tcpami/june11>

(注6): <http://mailman.vast.uccs.edu/listinfo/pamitc>



今回の agenda .

議題をセレクトして説明する .

3.11.1 CVPR2014 bids

今回も立候補なし .

説明しよう . CVPR は organize したい人達がチームを結成して PAMI TC meeting で立候補し (a bid), その場でプレゼンし, 挙手で決定する . CVPR2011 も, CVPR2009 マイアミでの PAMI TC で挙手により開催が決定された . ちなみにその時の対抗馬は Chicago . コロラドチームは, ホテル代がシカゴの半分で済むことをアピールして勝利を収めた . 同時に CVPR2012 の立候補アピールもあり, Philadelphia と Rhode Island が競い合い, Rhode Island に決定 .

しかし CVPR2010 での PAMI TC では, CVPR2013 の立候補がなかった . おそらく過去に例がなかったため, 「立候補がない場合には PAMI TC がチームを結成して信任投票を行う」という議題を立てて, 通過した .^(注7)

そして今回の CVPR2011 では, 少なくとも CVPR2014 の立候補がなし . PAMI TC がチームを結成することに .

自由でオープンを標榜する PAMI TC としては, 立候補がなく投票なしで CVPR 開催地・チームを決定しなければならない事態が 2 年連続したことに, かなり危機感を抱いたようである . そこで, 次の議題が登場する .

3.11.2 Submission by general chairs and program chairs

CVPR2009 マイアミでの PAMI TC で決定した agenda の一つに, general chairs (GCs) と program chairs (PCs) の submission 禁止があった . それ以前に問題になっていたらしく, organizers の論文がバンバン accept されてしまうことに fair ではないとの声があがっていたらしく (これは単なる推測), PAMI TC の agenda の一つになっていた . そして会場の挙手の結果, GC と PC は submission 禁止が決まった . 当然その場で反論があり, CVPR2012 general chair になったばかりの Terrance Boult が「なぜ general chair まで?」と意義を唱えたが, 決まったことは仕方がない . これが CVPR2009 でのお話 .

そして翌年の CVPR2010 では, CVPR2013 の立候補がなかった . さらに CVPR2011 でも CVPR2014 の立候補がなかっ

(注7): <http://www.computer.org/portal/web/tcpami/july10>

た．これはやはり，CVPR を organize する general chair が論文を投稿すらできないというデメリットが大きすぎるからであろう，という意見が再燃し，CVPR2011 の PAMI TC では「general chair は投稿してよい」という規程に戻す motion (動議) が Marc Pollefeys から提出された．

司会^(注8)：A second? (賛成の人は?)

会場：(ばらばらと手が挙がる)

司会：Any discussion?



議論の様子 (左) 反対派の Jitendra Malik (右) 賛成派の Luc Van Gool, Marc Pollefeys. 壇上には司会の Sudeep Sarkar.

マイクの争奪戦になるほど議論が白熱．General chair は submission system にアクセスできないから意味がないとか，fair であるだけでなく fair に見えなければならぬとか，PAMI TC 主催のどの会議に適用されるのかとか (答：CVPR と ICCV のみ)．

motion の文言の修正の後 (「can submit だ」とか「may にしろ」など英語の微妙な表現からフレーズの言い直しまで含む)，挙手で決議をとった結果，否決．つまり，General chair は投稿できないままである．これで来年の CVPR 立候補も望みが薄いだらう．

どちらでもいいんじゃないか，というのが個人的な意見．それよりも review に直接関わる area chairs がなぜ投稿禁止にならないのか，という CVPR2009 PAMI TC での発言に賛成．とはいえ，こういう議論がされること自体，長尾賞が領域チェアが実行委員からしか選出されていない MIRU よりも健全であろう．

3.11.3 IEEE CS Sponsorship

CVPR は IEEE Computer Society (CS) が 100% スポンサーであったが，今回は 50% であった．今後 IEEE CS のスポンサーをどうするか，が議題．

IEEE CS のスポンサー規程を今回の CVPR のような大規模学会に適用すると，参加者 1 人当たり \$100US も IEEE CS に支払わなければならなくなる．しかも，事前にローンや予算を組むことも禁止という無茶な話のため，今回は別のスポンサーを探し (CVPR2011 General chair の Terrance Boult 所属の UCCS)，IEEE CS と UCCS が半々のスポンサーとなった．

motion は，次回以降どうするかを，PAMI TC が議論し，PAMI TC member にアナウンスする (PAMI TC meetings でか mail list^(注9)でか不明)．あっさり承認．ただし，Terrance Boult ただ一人否決に挙手．理由は「no difficulty to make a

decision」とのこと．察するに，議論してアナウンスして決議なんてめんどくさいことをしなくても，答えは簡単，ということらしい．これが今回の general chair の意見．

続いての motion は，これを議論する committee の承認．女性研究者代表 (?) の Linda Shapiro が「なんで committee member に女性がいないの」という反対意見をだし (すでにこれは彼女の持ちネタ的な意見になっている^(注10))，彼女を加えた committee で決議を取って，承認．

3.11.4 e-Voting

これが大議論を巻き起こした今回の問題議題．

上記のような議論や決議は，年に数回の PAMI TC meeting でしか行えない．そこで e-Voting を導入する，というのが趣旨．この投票権を持つのが誰かとか，今までの投票のやり方を継続すべきだなど意見続出．

投票権は，元の motion では「IEEE CS member であること」が条件になっていて，これが議論的．CVPR に参加するには CS member でなくてもよく (参加費が高いだけ)，だれでも CVPR や ICCV に参加するだけで PAMI TC member になれた．これを変更することはおかしいという意見が多数．また実際に誰が投票権を持つのかをどうやって調べるのかについては，PAMI TC がやるしかないが，どうやるかは全く不明．フロアからは PAMI TC meeting はこれまでもオープンで，これからもオープンであるべき，という意見が多く，motion を提出した側は e-Voting しないとなかなか議論が進まない，という意見と対立．

議論の末，挙手で決議をとっても賛否拮抗して，一目で判断できず．

会場：「こんなときに挙手をカウントできる vision システムないのか (笑)」

Terrance Boult：(iPhone で会場の写真を撮りながら)「あとで解析できるぞ (爆笑)」

それはともかく，賛成派反対派 1 名ずつ壇上に立って挙手しなおしカウントした結果，否決．



for (left) and against (right) the motion. かなり僅差．

e-Voting は当然なさそうである．次回の PAMI TC meeting はおそらく ICCV2011．ICCV に参加する方はぜひ清き一票を．個人的意見では，e-Voting にした途端に誰も投票に興味を持たなくなると思われる．会議中に集まって挙手するから成り立っているのだろう．

(注8)：Sudeep Sarkar

(注9)：前出．

(注10)：実際本日 2 回目の意見で，これより前の CVPR2012 進捗報告で organizers になんて女性がいないの，と 1 回目の意見を述べている．Area chair には女性がいるはず，との回答を得た．おそらく Fei-Fei Li のことか．

4. 22th June, 2011: 2nd Day

4.1 Posters 2A : Video Analysis and Event Recognition, Computational Photography, Vision for Graphics, Sensors

1. A Large-scale Benchmark Dataset for Event Recognition in Surveillance Video; Sangmin Oh, Anthony Hoogs, A.G.Amitha Perera, Chia-Chih Chen, Jong Taek Lee, Jake Aggarwal, Hyungtae Lee, Larry Davis, Xiaoyang Wang, Eran Swears, Qiang Ji, Kishore Reddy, Mubarak Shah, Carl Vondrick, Hamed Pirsiavash, Deva Ramanan, Jenny Yuen, Antonio Torralba, Bi Song, Anesco Fong, Amit Roy-Chowdhury, Mita Desai

大規模データセット公開 . TRECVID の 2 倍程度 . ACCV2010 の懇親会で「新しいデータセットを作っているんだ」と話をしてくれた kiware の A. G. Amitha Perera がポスター前で説明していた .

3. Action recognition by dense trajectories; Heng Wang, Alexander Klaser, Cordelia Schmid, Cheng-Lin Liu

SIFT と KLT を追跡 , 軌跡上で HOG, HOF, MBF を計算し認識に利用 .

6. Activity Recognition using Dynamic Subspace Angles; Octavia Camps, Mario Sznajder, Binlong Li, Teresa Mao, Mustafa Ayazoglu

動画画像を subspace で表現 , m フレームを表現する $m \times m$ のハンケル行列を利用 . subspace 同士の類似度を用いて認識 .

28. Blur kernel estimation using the Radon Transform; Taeg Sang Cho, Sylvain Paris, Bill Freeman, Berthold Horn

Blur を , エッジに垂直方向のラドン射影で表現 . 提案する RadonMAP を用いて debluer 画像を推定 .

31. Estimating Motion and Size of Moving Non-Line-of-Sight Objects in Cluttered Environments; Rohit Pandharkar, Andreas Velten, Andrew Bardagjy, Ramesh Raskar, Mounsi Bawendi, Ahmed Kirmani, Everett Lawson

Femto Photography と命名 . 1 フレームあたり 0.5mm の距離を光が進む超高速の画像撮像デバイスを作成 . 光が反射 , 散乱している様子が目に見える動画を撮影できる .

32. Exploring Aligned Complementary Image Pair for Blind Motion Deblurring; Wen Li, Jun Zhang, Qionghai Dai

ぼけた画像と , その画像の変換後にぼけた画像の 2 枚から , 変換された blur kernel を導出 . 変換と畳み込みが複数得られた場合の解析 .

42. Camera Calibration with Lens Distortion from Low-rank Textures; Zhengdong Zhang, Yasuyuki Matsushita, Yi Ma

射影で歪んだテクスチャを sparseness を利用して基に戻す . このパラメータを推定することで構成を行う .

43. High-resolution Hyperspectral Imaging via Matrix Factorization; Rei Kawakami, John Wright, Yu-Wing Tai, Yasuyuki Matsushita, Moshe Ben-Ezra, Katsushi Ikeuchi

高解像度 RGB 画像と低解像度スペクトル画像を用いて , 高解像度スペクトル画像を (超解像) 復元 .

4.2 Orals 2B : Optimization Methods

1. A Non-convex Relaxation Approach to Sparse Dictionary Learning; Jianping Shi, Xiang Ren, Jingdong Wang, Guang Dai, Zhihua Zhang

L2 ノルムで dictionary learning をするとき , sparse 正則化項として L0 ノルムや L1 ノルム (Lasso) をよく用いるが , ここでは non-convex な MC penalty 項を用いる . sparse coding と dictionary learning の 2 ステップを反復するアルゴリズムを提案 .

2. A Study of Nesterov's Scheme for Lagrangian Decomposition and MAP Labeling; Bogdan Savchynskyy, Jorg Kappes, Stefan Schmidt, Christoph Schnorr

MRF におけるグラフのエネルギーが convex かつ non-smooth であるような問題を解く Nesterov の手法を改良 .

3. Scale Invariant cosegmentation for image groups; Lopamudra Mukherjee, Vikas Singh, Jiming Peng

co-segmentation はヒストグラム類似度を利用するので , MRF のエネルギー項にヒストグラム類似度を追加する . しかし , n 枚の co-segmentation においては類似度の組み合わせが増大し n^2 個の項が必要となる . これを n に抑えるアイデアは , 画像のヒストグラムベクトルを列に持つ行列を作ること . 同じ対象物体のヒストグラムであれば , この行列の列は同じはずであり , 行列のランクは 1 になる . したがって MRF エネルギーにランク 1 制約を加える . このアイデアは , 対象物体の大きさの変化にも対応できる . 大きさの変化はヒストグラムの定数倍に対応するが , rank1 には変わらないから .

4. Submodularity beyond submodular energies: coupling edges in graph cuts; Stefanie Jegelka, Jeff Bilmes

MRF エネルギーの binary 項は , エッジについてのエネルギーの和である . これはコントラストが弱くなる画像ではうまく働かないため , エッジの対 (coupling) についてのエネルギーの和にする . こうすると submodular ではなくなるが , 近似アルゴリズムを提案 , Cooperative Cut と命名 .

5. Scale and Rotation Invariant Matching Using Linearly Augmented Trees; Hao Jiang, Tai-Peng Tian, Stan Sclaroff

領域 (superpixel) をノードとするグラフで対象物体を表し , スケール変化と回転にロバストなマッチングを実現 . そのために , グラフを修正する Linearly Augmented Trees (LAT) を提案 . 目的関数を導出して mixed integer optimization として定式化 . しかしこれを解くことは困難であるので , linear relaxation を用いて近似的に解く .

4.3 Posters 2B : Stereo and Structure from Motion, Performance Evaluation, Object Detection

5. An Analysis of Using High-Frequency Sinusoidal Illumination to Measure the 3D Shape of Translucent Objects; Michael Holroyd, Jason Lawrence

半透明物質の形状測定のために , sinusoidal descattering を利用して , triangulation を行う .

10. L1-rotation averaging using the Weiszfeld algorithm; Richard Hartley

L1 ノルムを用いたカメラ姿勢の推定 . 単一カメラ姿勢の averaging と , 複数カメラの姿勢の consistency を利用 .

22. Structure from motion for scenes with large duplicate structures; Richard Roberts, Sudipta

180°対称のテクスチャがあると、対応をとるときにあいまいになる。そこで majority consistency を計算して形状を復元。

23. The Light-Path Less Traveled; Srikumar Ramalingam, Sofien Bouaziz, Peter Sturm, Philip Torr

カメラと鏡がすべて既知で構成されている場合に、3D-2D の対応からカメラ姿勢を求める。光線をすべて追跡するより効率的。

27. What Makes a Chair a Chair?; Helmut Grabner, Juergen Gall, Luc VanGool

3D シーンに 3D 人体が座れるかどうかという affordance detector を計算。

33. Efficient Subwindow Search with Submodular Score Functions; Senjian An, Patrick Peursum, Wanquan Liu, Svetha Venkatesh

ウィンドウ探索において劣モジユラの目的関数を利用。

40. PClines - Line Detection Using Parallel Coordinates; Marjeta Dubska, Adam Herout, Jiri Havel

parallel coordinates を用いて、直線検出。Hough 変換よりも高精度。

4.4 Orals 2D : Segmentation and Grouping

1. Occlusion Boundary Detection and Figure/Ground Assignment from Optical Flow; Patrik Sundberg, Jitendra Malik, Michael Maire, Pablo Arbelaez, Thomas Brox

遮蔽輪郭を検出するために motion を利用。gPb contour (エッジ) と Optical flow の境界を検出し、前景・背景のラベルを付ける。

2. Nonlinear Shape Manifolds as Shape Priors in Level Set Segmentation and Tracking; Victor Prisacariu, Ian Reid

shape prior を用いて遮蔽にロバストな物体抽出を実現。deformable shape を Elliptic Fourier descriptor で記述し、Gaussian Process で次元削減 (GP-LVM)。PCA や kPCA よりも低次元で高精度を実現。segmentation のためには、fitting の形状パラメータで微分する。

3. Kernelized Structural SVM Learning for Supervised Object Segmentation; Luca Bertelli, Tianli Yu, Diem Vu, Salih Gokturk

Structural SVM で segmentation を行うというアイデアを提案。画像 X とマスク Y の対を学習する。Structural SVM の目的関数に、学習マスクからの距離の項を追加。Graph Cut で解ける。カーネルは object kernel と mask kernel の積で表す。

4. Contour Based Joint Clustering of Multiple Segmentations; Daniel Glasner, Shiv Vitaladevuni, Ronen Basri

Joint clustering (joint segmentation) を提案。複数の画像において、形状が似ていて、かつ内部が coherent であるような領域分割を見つけるが、これが時間的にも coherent であることを要請する。まず over-segmentation の状態から開始し、併合後の領域が各フレームで似ており、かつ領域内が coherent であるような bounding counter をを見つける。新しい boundary descriptor を提案、これにより、領域の併合が descriptor の和で表せる。これを convex optimization で解く。

5. Real-time Human Pose Recognition in Parts from Single Depth Images; Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Andrew Blake

Microsoft Kinect の技術のうち、inferring body parts & clustering を説明：joint localization, 1 million training images, fast & accurate。学習には合成画像を用い、体系と姿

勢を変えて、depth とパーツ毎に色をつけた画像をレンダリング。特徴量に Fast Depth Image Feature を考案、各点での depth とその周辺の depth の差を利用。これを Decision forest classifier で分類、mean shift で peak を見つけて、領域を検出。

この発表の一つ前の発表が終わるころになって、会場がものすごく混雑してきた。一番前の席に座っていたのに、それよりも前の床に座り込む人も出てきたほど。広い会場の後ろから横までびっしり立ち見で埋め尽くされる。何だろう? と思っていたが、発表が始まって「ああ Kinect だったんだ」と納得。かなり注目度の高い発表だった。

4.5 Posters 2C : Optimization, Vision for Robotics, Scene Understanding, Image and Video Retrieval

9. Submodular Decomposition Framework for Inference in MRF with Global Constraints; Dmitry Vetrov, Anton Osokin, Vladimir Kolmogorov

MRF のエネルギーを分解する submodular 分解を提案。

19. Image analysis by counting on a grid; Alessandro Perina, Nebojsa Jojic

Bag-of-visual-words をグリッド状で計算して、sliding window で画像全体を認識・分割。

24. Semantic structure from motion; Sid Ying-Ze Bao, Silvio Savarese

object detection と SfM の同時実行。

26. City-Scale Landmark Identification on Mobile Devices; David Chen, Georges Baatz, Kevin Koeser, Sam Tsai, Ramakrishna Vedantam, Timo Pylvanainen, Kimmo Roimela, Xin Chen, Jeff Bach, Marc Pollefeys, Bernd Girod, Radek Grzeszczuk

GPS も利用して landmark 検出。1.7 million 枚を学習。

30. Hello neighbor: accurate object retrieval with k-reciprocal nearest neighbors; Qin Danfeng, Stephan Gammeter, Lukas Bossard, Till Quack, Luc VanGool

k-NN ではなく、k-reciprocal NN (互いに k-NN であるペア) を用いる。

4.6 Paper awards

Poster session が終わって 2 日目 banquet が始まる前に、best paper が発表された。プレゼンターは Terrance Boulton (この人、本当に大活躍)。

Best Student Paper Honorable Mention (注11) : Separating Reflective and Fluorescent Components of An Image, Cherry Zhang, Imari Sato

Best Student Paper : Recognition Using Visual Phrases, Ali Farhadi, Mohammad Amin Sadeghi

Best Paper Honorable Mention : Discrete-Continuous Optimization for Large-scale Structure from Motion, David Crandall, Andrew Owens, Noah Snavely, Daniel Huttenlocher

Best Paper : Real-time Human Pose Recognition in Parts from Single Depth Images, Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake

プレゼンター代わって General co-chair の Shmuel Peleg . Longuet-Higgins Prize は、長い年月を経て評価される論文に与えられる賞で、10 年前の CVPR で発表された論文が対象 . 今回は CVPR2001 に発表された「あの」論文で、今やどの教科書にも登場するアルゴリズムであり、他のどの CVPR 論文よりも 15 倍^(注12)以上引用されている、最も有名な手法に贈られる、とのこと .

Longuet-Higgins Prize : Rapid Object Detection using a Boosted Cascade of Simple Features. Paul A. Viola, Michael J. Jones

賞状を受け取りに来たのは Paul A. Viola で、この受賞のためだけに来たそうである .

最後は T-shirt contest . 過去の CVPR で配布された T シャツを表彰しようというもの . web でもアナウンスされていた . この学会は多くのボランティアに支えられている、ということで Standing ovation で拍手喝采 .



T-shirt コンテスト表彰結果 .

4.7 Banquet / reception

前日と同じ . かなり食事内容には飽きている . しかし議論には飽きたりず、夜遅くまで話し込んでいる人が多数 .

5. 23th June, 2011: 3rd Day

5.1 Posters 3A : Motion and Tracking; Shape Representation and Matching; Illumination

3. A Two-Stage Reconstruction Approach for Seeing Through Water; Omar Oreifej, Guang Shu, Teresa Pace, Mubarak Shah

averaging と non-linear registration (B-spline) の 2 段階を繰り返して、揺れる水面から底の模様を復元 .

26. Aggregating Gradient Distributions into Intensity Orders: A Novel Local Image Descriptor; Bin Fan, Fuchao Wu, Zhanyi Hu

SIFT のような orientation の計算が不要な、回転不変な局所特徴の提案 .

29. Global Optimization for Optimal Generalized Procrustes Analysis; Daniel Pizarro, Adrien Bartoli

一般化 Procrustes Analysis (複数剛体形状同時位置合わせ) を、SOSP と SDP の反復で解く .

36. Robust Point Set Registration Using EM-ICP with Information-Theoretically Optimal Outlier Handling; Jeroen Hermans, Dirk Smeets, Paul Suetens, Dirk Vandermeulen

EM-ICP はモデルを GMM で表現するが、これはシーンモデルも GMM で表現した、ロバストな EM-ICP .

40. Interreflections removal for photometric stereo by using spectrum-dependent albedo; Miao Liao, Xinyu Huang, Ruigang Yang
反射をアルベドの多項式で表現し、m 回の相互反射を分離 .

5.2 Orals 3A: Object Recognition

1. Describing Images: Understanding and Generating Image Descriptions; Girish Kulkarni, Visruth Premraj, Sagnik Dhar, Siming Li, Alexander Berg, Yejin Choi, Tamara Berg

タイトルが「Describing Images」から「Baby Talk」に変更されていた (baby が話す程度の言葉が出力できるから) . 画像を入力し、それを説明する詳細な英語の文章 (descriptive text) を生成することが目的 . 従来は summarization や retrieval の手法を流用していたが、ここでは descriptive text 自体を多数学習する . 流れ: 入力画像から物体を検出し、attribute を生成 . attribute と、それらの間の空間配置 (on, over, ... などの前置詞) を、vision&text potentials を作成して CRF で学習 . 最後に英語文章を template ベースと language model ベースの 2 手法で生成 (自然言語処理) . 評価は BLUE score を用いて、人間が生成した英語文章と比較 . 時々めっちゃくちゃな文章を生成する language model よりも、template ベースのほうが成績がよい . Tamara Berg の上手なプレゼンテーションが印象的 .

2. Recognition Using Visual Phrases; Ali Farhadi, Mohammad Amin Sadeghi

パーツ (小)、物体 (BoW) (中)、シーン (大) という対象のスパンのうち、ギャップが大きい「物体とシーン」の間に位置する表現を提案 . ここでは visual word の組を visual phrase と呼んでいるが、全ての組み合わせは意味がないため、「シーンの一部で意味のある chunk」を visual phrase と定義している . 従来のように、検出物体同士の直接的な関係を使わない . なぜなら、「人」と「馬」のそれぞれ記述を足しても、「馬に乗った人」の記述にはならないから . この「馬に乗った人」の記述を検出することがゴール (phrase detector) で、visual word 同士の少数の関係を mining して学習 . phrase detector で検出した「馬に乗った人」の検出結果から逆に「馬」と「人」に分解し、「馬」「人」単体の検出よりも高性能を示す .

3. Uncovering Vein Patterns from Color Skin Images for Forensic Analysis; Chaoying Tang, Adams Wai Kin Kong, Noah Craft

発端は、カンボジアの性的暴行事件で、写真に写った犯人の腕だけから人物を特定できないか、というもの . 腕の静脈はよい biometrics だが、従来のように近赤外 (NIR) 画像を使うのではなく、RGB 画像を使う . このために画像生成過程を、RGB 光 (標準光 3 種類) とカメラモデル (Nikon と Canon の 2 種類)、肌反射と肌内部散乱 (3 層の Kubelka-Munk) の 3 段階でモデル化 . これの逆モデルのために 3 層 FF-Neural net を使用 .

4. What You Saw is Not What You Get: Domain Adaptation Using Asymmetric Kernel Transforms; Brian Kulis, Kate Saenko, Trevor Darrell

Train と test の画像サンプルの撮像条件が異なるという、visual domain shift の問題を扱う . ここでは、何が変化するの

(注12): 50 倍だったかも

かを学習するアプローチをとる（これが domain adaptation）．従来はサンプル画像の変化だけを扱っていたが，train と test で特徴ベクトルの種類（SIFT と SURF）や次元が異なっても adaptation できる（ECCV2010 に発表済み）．この特徴ベクトルの変換を線形変換で表し，カーネル化して学習．

5. Automatic Photo-to-Terrain Alignment for the Annotation of Mountain Pictures; Lionel Baboud, Martin Cadik, Elmar Eisemann, Hans- Peter Seidel

写真の写った山に「これは* * 山です」という annotation を加える研究．3次元地形データとカメラ映像マッチングによるカメラ姿勢推定問題だが，天候の変化や雲の有無で見え方が全く異なり，難しい問題．しかもロッキー山脈の写真では水平線情報も使えない．特徴はシルエットエッジ，カメラ位置は GPS かユーザー入力とし，問題を回転の推定だけに限定．spherical edge map のマッチング問題にして，SO(3) の FFT を利用して高速に相関を計算．最後のスライドで CVPR2011 webpage のロゴになっている山の写真に annotation^(注13)をするという余興に歓声と拍手．



CVPR2011 ロゴに annotation．講演後は質問者に取り囲まれる．

5.3 Posters 3B: Shape from X, Video Surveillance, Face and Gesture

3. Adequate Reconstruction of Transparent Objects on a Shoestring Budget; Sai-Kit Yeung, Tai- Pang Wu, Chi-keung Tang, Tony F. Chan, Stanley Osher

透明物体（ガラスなど）の形状復元．

39. Real Time Head Pose Estimation with Random Regression Forests; Gabriele Fanelli, Juergen Gall, Luc VanGool

顔パッチの各部分を random forest で回帰．1日目のデモでやっていた（Real Time Head Pose Estimation From Consumer Depth Cameras）．

41. A RankOrder-based clustering algorithm in face annotation; Chunhui Zhu, Fang Wen, Jian Sun

データセット依存の Rank-order に基づく距離を提案．

42. Support Tucker Machines; Irene Kotsia, Ioannis Patras

Tensor Tucker decomposition に基づく 2 クラス識別器の提案．

5.4 Orals 3D: Statistical Methods and Learning

1. Principal Regression Analysis; Jason Saragih

回帰問題 $y = f(x) + \epsilon$ という問題で， y の位置を subspace で拘束するという geometry 制約を入れた回帰．これを顔画像 x を顔モデル y に適用．最適化には実用的な SD-Wiberg を利用．

2. Accelerated Low-Rank Visual Recovery by Random Projection; Yadong Mu, Jian Dong, Xiaotong Yuan, Shuicheng Yan

(注13): 論文の冒頭に同じ図がある．

ランク制約問題 $\min f(A) + \text{rank}(A)$ を緩和した，nuclear ノルムで制約を付けた最小化問題 $\min f(A) + \|A\|_*$ を解く．アイデアは， $\|A\|_*$ を random projection P で射影した A' を用いて $\|A'\|_*$ ， $A' = PA$ と分解すること．

3. Online Domain-Adaptation of a Pre-Trained Cascade of Classifiers; Vidit Jain, Eric Learned- Miller

Viola-Jones の登場 10 年で Face detection はもう解決済みなのか？という問題提起をして，検出できない例を示して No という．例えば太陽がまぶしい日中で撮影した写真の顔は，陰影が付きすぎて（日の当たった顔の右側は明るく，日陰になった顔の左側は暗い），学習した Harr-like 特徴は使い物にならない．そこで，写真中に検出された複数の顔の情報を利用して，その写真中の未検出画像を検出してやろう（同時に誤検出も減らしてしまおう）というのがアイデア．

4. Coupled Information-Theoretic Encoding for Face Photo-Sketch Recognition; Wei Zhang, Xiaogang Wang, Xiaoou Tang

顔写真と顔写真を照合するのではなく，顔写真とスケッチ（イラスト）の顔を照合するのが目的．そのために，写真とイラストのペアから codeword を作り，coupled projection tree を学習．このために CVHK database を作成．

5.5 Posters 3C: Early & Biological Vision, Medical, Applications

24. sLLE: Spherical Locally Linear Embedding with Applications to Tomography; Yi Fang, S.V.N. Vishwanathan, Mengtian Sun, Karthik Ramani

球面上データの LLE．

25. 3D Motion Reconstruction for Real-World Camera Motion; Yingying Zhu, Mark Cox, Simon Lucey

実際にありそうなカメラ運動で制約を付けた 3次元復元（人体姿勢）．

37. Predicting Image Matching using Affine Distortion Models; Daniel Fleck, Zoran Duric

小数画像のパッチを affine 変換して合成し，マッチングしそうな画像を予測．

6. Summary

6.1 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，top conference 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 PhD 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 人程度だろうか．アジア系の顔を見かけたら，圧倒的にそれは中国人が韓国人．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，存在感としてはシンガポールの方が上である（というのが個人的感想）．昨年 CVPR や一昨年の ICCV などでも NUS や NTU などの発表が多く，来年の CVPR2012 の area chair にはシンガポールから 2 名入っている（中国人の area chair はまだまだ少ない）．日本人の area chair は，今年は石川博先生（早大），

来年は佐藤洋一先生（東大）と、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。

ここまでは中国やシンガポール、韓国に押されて、日本の存在感がなくなってしまうだろう。しかしそれを心配しているのは日本人だけだろう。欧米人から見れば中国人も日本人も見分けがつかず（たぶん）、振興が激しいアジア勢の一員としかみなされていないのかもしれない。日本だろうかシンガポールだろうか香港だろうか、中国と何が違うの？という程度の認識かもしれない。そうすると、「日本が日本が」と言っても、単にアジアの中で張り合っているに過ぎない。日本の存在感がなくなると何が問題なのか、根底から議論して共通認識を作るべきだろう。そうでなければ戦略も立てられないのだから。

6.2 Trend

自分の興味に従って眺めた、CVPRにおける trend をいくつか紹介する。

- submodular が流行り。つまり MRF/graphical model 全盛。
- “everything is graphical model” “Oh, yes” (reception のときに誰かが誰かとそう話していた)
- vision の研究で何かを推定するときには確率モデルを導入して事前確率と尤度を使う。geometry では当然。photometry ではまだまだ。
- Geometry で Bayes は必須。認識では他の手段は多数あるが、Bayes ができないと今後は Geometry で生きていけない。
- dual decomposition が流行り？
- Bag-of-words はすでに古い。次の特徴量を探している。
- kernel は完全に sparse に取って代わられた。何でも L1 ノルム最適化。多様体学習や kernel の発表は皆無、SVM はほんの一握り。
- poster の最初は”Goal & Motivation” でダイレクトに訴えかける。
- 所属と国境を越えて共同研究・発表しているグループ多い。
- poster に「author1*, author2*, author3 (*: equal contribution)」と書いてあるものがいくつもあった。昨年は見かけなかった、新しい傾向。こう書いたからって意味があるとは思えないのだが、何かアメリカで新しい動きがあるのだろうか。
- oral でも最初のスライドで (*: equal contribution) と明示しているものがあった。さらに、speaker が途中で交代するという発表が何件もあった。

6.3 Videos

昨年の CVPR は videolectures.net に oral/poster spotlight の様子が公開されているが、^(注14)今年も video overflow のために撮影をしていたので、公開はされるだろう。ただし web^(注15)には “After CVPR2012, all CVPR2011 video proceedings will be free for anyone to access.” と書いてあるので、来年までお預けのようである。

6.4 Virtual Attendee Registration

今回の registration は 1500 人で打ち切ったため、新しい試みとして Virtual Attendee Registration を導入していた。Oral の生中継と録画が見られて、USB proceedings もついて、US\$300 という破格の値段。しかし、web^(注16)には “Virtual attendees can even ask questions after the live talk.” と書いてあるものの、実際の oral で座長がそのような端末操作や質問画面を眺めている様子は全くなかったため、上記売り文句は空振りに終わったようである。同じホテル内の Video Overflow room で中継映像を視聴した限りでは、生中継の映像の質はお世辞にも良いとは言えなかった（録画の方の質は不明）ため、virtual attendees もストレスの溜まる中継映像を見ていたに違いない。

6.4.1 予算は 30 万

ぜひ多くの人に CVPR に実際に参加して、あの熱気を感じてほしい。予算は 30 万円 = 航空券 20 万円 + 宿泊費 5 万円 + 参加費 5 万円。論文が通ったら参加する？日本人の論文が accept される事前確率が 1% なのだから^(注17)、実際問題としてそれは無理な話だろう。こうして、一握りの研究室以外は top conference に参加しないことが当然となりつつある日本において、若い世代は「そんな top conference に行かないし投稿もしない」ことが当然となっている（研究人生のスタートからしてそうなのだから）。ぜひそんな悪循環を断ち切って、世界の top を体感してみてほしい。

(注14): http://videolectures.net/cvpr2010_san_francisco/

(注15): <http://www.cvpr2011.org/videos>

(注16): <http://www.cvpr2011.org/registration>

(注17): この確率はヤマ勘です。MIRU の投稿数が 200 件、CVPR での日本人発表件数が 2 件、という概算から。