CVPR2012 報告

玉木　　徹†

† 広島大学
E-mail: †tamaki@hiroshima-u.ac.jp

あらまし　CVPR2012 報告．修正版．

A report on CVPR2012

Toru TAMAKI†

† Hiroshima University
E-mail: †tamaki@hiroshima-u.ac.jp

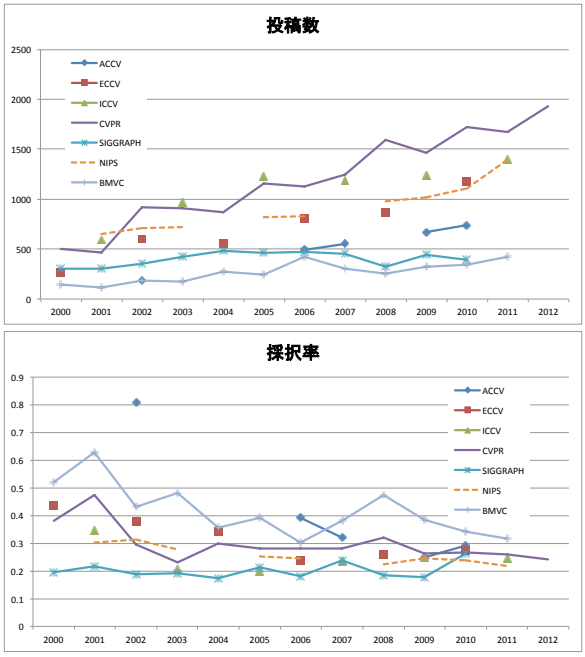
Abstract　A report on CVPR2012. A revised version.

1. はじめに

これは Rhode Island, Province, USA で開催された CVPR2012 (the 25th IEEE Conference on Computer Vision and Pattern Recognition) の報告である．これまでにも ACCV2010 [1]、CVPR2011 [2] 等の報告を独自に行っている．また CVIM 研究会主催の報告も行われているので [3]，別途参照してほしい．

CVPR とはコンピュータビジョンとパターン認識の分野において，世界中の研究者が一堂に会する年に 1 回の国際会議である．発表される論文のレベルの高さと，投稿された論文数に対する採択される論文の数の少なさ（採択率の低さ）の為に，トップカンファレンスの一つと呼ばれている(注1)（グラフを参照）．

ポケットガイド [4] に掲載されている General Chair からのメッセージによると，投稿された 1933 本の論文を 45 人のエリアチェアに割り当てる際に，bag-of-words を用いた割当アルゴリズム [5] が利用されたそうである．これは NIPS でも用いられているそうである．



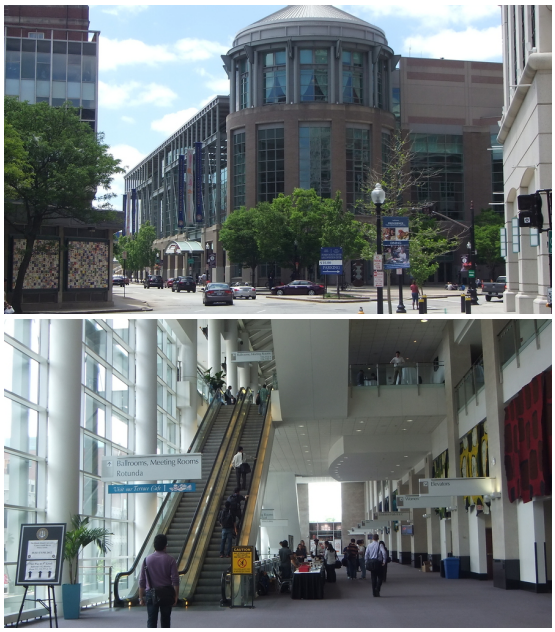
メジャーな top conference の（上）投稿数と（下）採択率の推移．

今回の開催地は Providence．ニューヨークのやや右上に位置しており，ボストンとニューヨークの中間あたりの場所．日本からのフライトは，シカゴ経由やデトロイト経由という人が多かった．その中でも最も最短経路と思われるのは，日本からボストンに直行便で飛び，ボストンからはプロビデンスまでは電車で移動，というもの．気候は日本の 5 月の気温に近く（最高気温 23 度程度），雨も全く降らないため，かなり快適である，と最初は思っていた．しかし CVPR の 3 日目から東海岸は extreme hot wave（熱波）に襲われ，日中の最高気温が華

(注1)：この分野の他のトップカンファレンスには ICCV と ECCV が，他の分野では例えば NIPS や SIGGRAPH 等がある

氏 98 度〜105 度という日中の外出が危険な程の猛暑になってしまった．街自体は閑静で，近くに幾つか大学があるらしく，大きなショッピングモールが一つある他は住宅街が多い．比較的安全なだろうと思っていたが，学生が一人で夜（バンケットの後）に中華料理屋へ行こうと人気の少ない道を歩いていたら，「金を出せ！」と大柄な黒人と白人に取囲まれたという．やはりアメリカでの夜道の一人歩きはやめた方がよい．

会場は Providence convention center．過去数年はホテルでの開催が多かったが，近年の参加者の急増のために大きな会場を使ったと思われる．その結果参加者数は 1800 人を超えた模様（過去 2 年は 1500 人で事前登録を打ち切っていた）．しかしホテルとは違って，かなり大きな会場であり，パシフィコ横浜や幕張メッセのように展示会場として利用するような場所である．そのため，音が反響して聞こえにくかったり，巨大なファンの音がうるさかったり，隣の会場の設営準備の音が響いてきたりして，あまり集中できないことも多かった．また会場が巨大すぎるのにプロジェクターが通常サイズで，1 つのオーラル会場に大小のスクリーンを 4 つ並べていた（遠くからではかなり見づらい）．その代わり、いつも座る椅子を探すのに苦労するオーラルセッションでも十分な数の席が用意されていて，いつでも空席を見つけることができた．逆に CVPR で満席になっていない（しかも閑散として見える）オーラル会場は初めて．おそらく設営された椅子の数は，大きなオーラル会場では 1500 席程度はあったと思われる．



Providence convention center の外観と registration デスク

Disclaimer

以下では，main conference/workshop の 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．また実際の論文はほとんど読んでいない．スライドやポスターを見聞きした情報や，他の人からの伝聞だけで書いているため，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．また一言コメントをお願いしたところ何人かの方から簡単な説明文を頂いた．ここに感謝する次第である．

正しい内容を知りたい方は，CVPR2012 の proceedings を入手するか，CVPR on the web (`http://www.cvpapers.com/cvpr2012.html`) を利用するか，第 20 回 コンピュータビジョン勉強会@関東「CVPR2012 読み会」(`http://atnd.org/events/29608`) や第 21/22 回 関西 CV・PRML 勉強会（CVPR2012 輪講）(`http://groups.google.com/group/cvprml?hl=ja`) に参加していただきたい．

2. 16th/17th June, 2012: Workshop / Tutorial days

CVPR 本会議の 2 日前から Workshop や Tutorial が多数開催されている．

- Point Cloud Processing：大盛況．小さい部屋に立ち見がいっぱい．
- Egocentric Vision：午後の Keynote talk は Google の Hartmut Neven．タイトルは TBA だったが，ワークショップの内容と Google ということと時期的にも，Project Glass[(注2)]のことではないだろうかと期待して参加．しかし全くの期待はずれ… Google glass に関する Technical な話も，将来的なビジョンの話でもなかった．[(注3)]
- Vision Industry & Entrepreneur Workshop：多数のベンチャー企業が製品やサービスを発表で紹介していた．日本では大企業メーカーが製品発表するような内容を，ベンチャーがポスター発表．ちなみに，新しくなって復活した Computer Vision Central (`http://cvisioncentral.com/`) も登場．非常にいいサイトだったが，開設 1000 日の節目を迎えたということで昨年に一旦更新停止状態になっていた．
- Projector–Camera Systems：（山形大の天野先生より）このワークショップはもうすぐ 10 年になり，投稿数も参加者も伸び悩んでいるので，今後どうするのかを話し合っているらしい．
- Biometrics：部屋はガラガラ，参加者が少ない…
- Deep Learning Methods for Vision：このチュートリアルは中ぐらいの部屋で行っていたが，午前中はあまりに客が多かったので，午後からは広い部屋に移動していた．

人であふれている Deep Learning Methods for Vision

(注2)：`http://www.youtube.com/watch?v=9c6W4CCU9M4`
(注3)：post conference workshop でも同じプレゼンをしていた．

3. 18th June, 2012: 1st Day

3.1 Registration

今年も予稿集は USB メモリ．お土産 T シャツはシンプルなもの．プログラム冊子 [4] は例年通り CD ケースサイズのコンパクトなもの．


USB メモリの予稿集と T シャツ

3.2 Demos

FaceHugger: The ALIEN Tracker Applied to Faces, Univ. of Florence

（中部大の後藤さんより）ネーミングが面白い．映画「エイリアン」のように，顔に取り付いて離れないから face hagger.


相変わらずポスター会場は人でいっぱい

3.3 Posters 1A: Computational Photography, Shape Representation & Matching, Illumination & Reflectance, Shape from X

5. Laser Speckle Photography for Surface Tampering Detection, YiChang Shih, Abe Davis, Sam W. Hasinoff, Fredo Durand, William T. Freeman

（匿名）すっと触るだけで微細な speckle の変化を検出する．NCC で差を取って可視化するだけでも，変化は分かる．

（九大の長原先生より）誰かが触ったかどうかも，差分で分かる．laser speckle という他の分野の技術を CV に持ち込んだ点が評価されたか．

11. Enhancing Underwater Images and Videos by Fusion, Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, Philippe Bekaert

濁った水中写真をクリアにする．物理的な仮定を何も使わずに，1 枚の入力写真を 2 枚にして（ホワイトバランス・コントラスト調整），それぞれ 4 枚の重み画像（ラプラシアンマップなど）を計算し，多重解像度で合成し直す．エッジ強調のようなもの．

13. Compressive Depth Map Acquisition Using a Single Photon-Counting Detector: Parametric Signal Processing Meets Sparsity, Andrea Colaco, Ahmed Kirmani, Gregory A. Howland, John C. Howell, Vivek K. Goyal

エリア（イメージ）TOF センサは高価なので，1 個のフォトンセンサだけを使い，空間解像度は安価な DMD を使う．これで奥行きが計測できてしまう．

19. Scale Resilient, Rotation Invariant Articulated Object Matching, Hao Jiang, Taipeng Tian, Kun He, Stan Sclaroff

初期フレームで人物をストロークとして入力し，後続フレームではそのストロークを検出．古典的な pictorial structure では回転に弱いが，これはロバスト（体操選手の平均台上の回転も追跡）．

The Schrodinger Distance Transform (SDT) for Point-sets and Curves, Manu Sethi, Anand Rangarajan, Karthik Gurumoorthy

距離変換はコストが高いので，近似計算．数値微分を使わないので早いし，パラメータで滑らかさを制御できる．

32. Depth from Optical Turbulence, Yuandong Tian, Srinivasa G. Narasimhan, Alan J. Vannevel

暑い日に遠くの風景は揺らぐ．これを利用して距離計測．ステレオでやると高精度．Ray Optics ではなく Wave optics を考慮，精度が距離に依存しない！110m と 160m で実験．

42. 2.5D Building Modeling by Discovering Global Regularities, Qian-Yi Zhou, Ulrich Neumann

上空から得た建物の point cloud から建物モデルを復元．屋根や壁は垂直だったり傾きが 45 度であったり平行であるという regularity を仮定．

44. Robust Stereo with Flash and No-flash Image Pairs, Changyin Zhou, Alejandro Troccoli, Kari Pulli

（東北大の伊藤先生より）フラッシュ有りと無しの画像で，届く光量が違うため，距離計測ができる．

45. Detection by Detections: Non-parametric Detector Adaptation for a Video, Xiaoyu Wang, Gang Hua, Tony X. Han

どんな静止画の detector も動画用の detector にしてしまう．

3.4 Messages from Chairs

Co-general chair の Benjamin Kimia の welcom talk

Chair からの挨拶では，30 年前の第 1 回 CVPR83，20 年前の CVPR93，10 年前の CVPR2003 が紹介され，CVPR2012 の紹介へと続く．

プログラム冊子には「Ballroom A」と書いてあるのに，当日に会場変更（Hall C に）されていた．部屋の広さをみたら，明らかに Ballroom A は狭いし，前日まで昼食会場に使われてい

たので，明らかにおかしいと思って受付に聞いてみた．でも受付の人も知らないようだった（張り紙を見つけて初めて知ったらしい）．

その後も会場の変更が相次ぎ，プログラム冊子に書かれている部屋は全く信用できず，毎日確認することに．会場変更だけでなく，前日にオーラル会場だった部屋が次の日には別のイベントで使用されるなど，混乱しやすい状態であった．

3.5 Orals 1A: Computational Photography

発表予定の Antonio Torralba と座長の James Hays が，発表の 1 時間前に誰もいない会場に入って，プロジェクターの写り具合を入念にチェックしていた．確かに，照明が明るすぎてプロジェクターは暗すぎる．

ものすごく広いオーラル会場

1. From Pixels to Physics: Probabilistic Color De-rendering, Ying Xiong, Kate Saenko, Trevor Darrell, Todd Zickler

デジカメで撮った写真のカラー調整．カメラ毎に色が変わってしまうので，確率モデルを用いて JPEG 画像から P(raw RGB 画像) を推定（だから逆レンダリング (de-rendering)）．アプローチは，色 y が色 x になる確率 $P(x|y)$ を局所 GP 回帰で推定．応用には，確率的多重露出（HDR 作成），確率的照度差ステレオ．

2. Decomposing Global Light Transport using Time of Flight Imaging, Di Wu, Matthew O'Toole, Andreas Velten, Amit Agrawal, Ramesh Raskar

表面下散乱と相互反射は，ps（ピコ秒）単位で光の反射が異なるので，それを利用して分離．デバイスには femto camera を利用 (2 ps/frame)（これは昨年の CVPR ポスターでも発表していた）．その後 Nature で紹介されたビデオを上映... 結局，直接反射のインパルス応答（ガウスっぽい形）と表面下散乱のインパルス応答（指数減衰）を分離するのは難しいので，アドホックなルールでインパルス応答を直接反射と表面下散乱の成分に分離．

単にデバイスの性能を売りにしているだけで，手法やアルゴリズムがすごいという訳ではない．

3. Accidental Pinhole and Pinspeck Cameras: Revealing the Scene Outside the Picture, Antonio Torralba, William T. Freeman

画像に写っていないシーンの情報を取得する，という話．窓から光が差し込んでいる部屋の写真には，そこには写っていない「屋外」の情報が shadow として写っている．これを一般化して，窓辺に人が立った時と何もない時の，差し込む光の差（単純な差分！）を利用すると，人が遮蔽物となり，その結果 pinspeck カメラとなって（図がややこしいので論文参照）画像には写っていない屋外の風景を，画像に写っている屋内の壁に写し戻すことができるということをデモンストレーション．

（電通大の高橋先生より）目の付け所が良い．こったことはやっていないが．

4. Jigsaw Puzzles with Pieces of Unknown Orientation, Andrew C. Gallagher

矩形ピースでジグソーパズル．各ピースの回転は不明，という点が従来と異なる問題設定．照合には，ピース境界での RGB 色差を利用．

なぜ CVPR に来てまでオーラルでジグソーパズル（何年も前に PRMU アルコンでやったネタ）を聞かなければならないのだ？と思って途中で退出．

3.6 Orals 1B: Shape Representation & Matching

2. Progressive Graph Matching: Making a Move of Graphs via Probabilistic Voting, Minsu Cho, Kyoung Mu Lee

（柳川・オムロン）点同士のマッチングではなく，周辺の点で構成したグラフ同士でマッチング．

3. The Shape Boltzmann Machine: A Strong Model of Object Shape, Seyed M. Ali Eslami, Nicolas Heess, John Winn

（名城大の堀田先生より）「 Deep Boltzman Machine の変形版を用いた方法。生成型モデルなので、対象の一部しか見えていなくてもそこから元の形を推定できる。発表時のデモは非常にうまく行っていたが、実際は位置ずれに弱いと考えられる。」

3.7 Posters1B:Color &Texture,Early & Biological Vision, Image Based Modeling, Segmentation & Grouping

ポスターセッションは 13:00 から．ランチタイムに行われていた Doctoral Consortium は 13:40 まで．どちらも同じ部屋の同じポスターボードを使用... だから一部では CVPR ポスター発表の人がポスターを張れないという事態が発生していた模様．13:00 になったら，客は Doctoral Consortium だろうが CVPR poster だろうが関係なく聞きにくるので，「どこに何が張ってあるんだ？」と混乱．

9. Example-based Cross-Modal Denoising, Dana Segev, Yoav Y. Schechner, Michael Elad

Video を使って，ノイズのひどい audio のノイズ除去．

11. The Image Torque Operator: A New Tool for Mid-level Vision, Morimichi Nishigaki, Cornelia Fermuller, Daniel DeMenthon

各画素の「トルク」を，画像フィルタの一種として定義．いろいろな応用ができる．

12. FREAK: Fast Retina Keypoint, Alexandre Alahi, Raphael Ortiz, Pierre Vandergheynst

ORB, BRISK よりもコンパクトでマッチング性能の高い検出器．

（中部大の藤吉より）ORB とあまり変わらないか．延長線上の研究だろう．

18. Figure-Ground Segmentation by Transferring Window Masks, Daniel Kuettel, Vittorio Ferrari

学習画像のマスクをテスト画像に transfer（平均）して，テスト画像のマスクを作成．

21. Efficient Inference for Fully-Connected CRFs with Stationarity, Yimeng Zhang, Tsuhan Chen

全結合の CRF は計算コストが高い．$O(N^2)$ の計算ボトルネックの部分を，画像のフィルタリングで解決するのがミソ．

35. Higher Level Segmentation: Detecting and Grouping of Invariant Repetitive Patterns, Yunliang Cai, George Baciu

ユーザーが指定したパッチに似た部分を，レジストレーションベースで探索・検出・領域分割．

3.8 Orals 1D: Segmentation and Grouping

2. On Multiple Foreground Cosegmentation, Gunhee KIM, Eric P. Xing

(Yuan, Hiroshima Univ.) ベイズモデルを用いて，先に特徴点検出を行い．領域分割に用いる．スライドの作り方がすばらしかった．

3.9 Orals1C: Illumination & Reflectance

1. Discriminative Illumination: Per-Pixel Classification of Raw Materials based on Optimal Projections of Spectral BRDF, Jinwei Gu, Chao Liu

分光 BRDF で画像の領域分割（画素毎の認識）．従来は BRDF5 次元データを線形判別 $w^T x + b$ で識別．x が BRDF なので，この w^T を照明と思って w を実現する LED 照明を実現．つまり照明を当てるだけで $w^T x$ が実現できる．6 色の 25LED を使って，線形「SVM 光」で照らすと，material が識別できる．

2. Camera Spectral Sensitivity Estimation from a Single Image under Unknown Illumination by using Fluorescence, Shuai Han, Yasuyuki Matsushita, Imari Sato, Takahiro Okabe, Yoichi Sato

カメラ分光感度を計測するのは大変．従来は，多数の狭帯域光で計測するか，校正カラーチェッカーを 1 枚撮影．そこで蛍光色で校正カラーチェッカーを作ると，照明に依存しなくなり，非常に簡単に推定できるようになる．

3. Micro Phase Shifting, Mohit Gupta, Shree K. Nayar

Structured light の phase shift 法は，通常は低周波から高周波まで多数投影が必要．しかし相互反射やボケに弱い．そこで周波数解析．ボケや相互反射の周波数よりも高い周波数で，かつ周波数の近い（最低 2 つの）光を使うと，相互反射やボケに強くなる．周波数が近くてもうなり (beat) 周波数を利用，7 枚の入力で OK.

4. A Closed-Form Solution to Uncalibrated Photometric Stereo via Diffuse Maxima, Paolo Favaro, Thoma Papadhimitr

照明が未知の場合の照度差ステレオは，GBR ambiguity が発生，困難になる．そこで，ランバード拡散反射 (LDR) の極値を使うといアイデア．この極値は法線方向が一致するはずなので，解が制約できる．non-convex な問題だが closed-form の解を得た．

3.10 Posters1C:VisionforGraphics, Sensors, Medical, Vision for Robotics, Applications

35. Icon Scanning: Towards Next Generation QR Codes, Itamar Friedman, Lihi Zelnik-Manor

（東北大の伊藤先生より）QR コードの次として，アイコンを用いる．照明変化等は学習で対応．

4. 19th June, 2012: 2nd Day

4.1 Demo

Auto Face Re-Ranking By Mining the Web and Video Archives, NII

佐藤真一先生チーム．Google 顔画像検索結果の再ランキング．同一画像内にある顔は，クラスは分からなくても違うクラスのはず，という制約を入れる．

A Text Detection System for Urban Scenes and Related Applications, Nokia

LiDAR データに検出テキストを張りつけ，立体地図を作製．応用は分かりやすいナビゲーション（3 つ目の角を曲がる，ではなく「この看板」という目印を出せる）．

FREAK: Fast Retina Keypoint, EPFL

昨日のポスターのデモ．コードも公開中．ORB と BRISK の発展形．ランダムに 2 点を決めるのではなく，中心は小さい円で密に，周辺は大きい円で粗に取る．

（オムロンの柳川さんより）あまりマッチング精度が良くない．高速コンパクトでも性能が悪ければ意味がないのでは．

4.2 Posters 2A: Video Analysis, Stereo & Structure from Motion

6. Action Bank: A High-Level Representation of Activity in Video, Sreemananth Sadanand, Jason J. Corso

（名城大の堀田先生より）「著者は違うが Object bank の action 版である。205 個の action detectors の出力を特徴量として識別を行う。全ての実験で共通の action bank を用い、KTH データセットで 98.2%、HMBD51 で 38%等の高い精度が出た。」

17. Social Behavior Recognition in Continuous Video, Xavier P. Burgos-Artizzu, Piotr Dollar, Dayu Lin, David J. Anderson, Pietro Perona

（名城大の堀田先生より）「 spatio-temporal bag-of-words とパーツ追跡の軌跡を用いて Adaboost により識別器を学習。これに加えて時間的な context を用いることにより精度が 14 %向上した。」

24. A Combined Pose, Object, and Feature Model for Action Understanding, Ben Packer, Kate Saenko, Daphne Koller

3 レシピ（オムレツ，サラダ，スープ）の 13 アクションを認識．Saenko がプレゼンしていた．

33. Dense Reconstruction On-the-Fly, Andreas Wendel, Michael Maurer, Gottfried Graber, Thomas Pock, Horst Bischof

（東北大の伊藤先生より）タブレットの画像をサーバに送って，すぐに 3 次元復元．DTAM か PTAM かを使っている．東工大の鳥居さんの研究に近い．

43. Real-Time 6D Stereo Visual Odometry with Non- Overlapping Fields of View, Tim Kazik, Laurent Kneip, Janosch Nikolic, Marc Pollefeys, Roland Siegwart

反対方向を向いた 2 台のカメラ（校正済み）で形状復元．不定のスケールは，カメラが 2 台あることを利用してうまく推定．

4.3 Invited Talks

CVPR で Invited Talk は今までに見たことがない（少なくとも 2009 年から）ので，非常に珍しい．

David Mumford: Where are we in Vision? Some thoughts about the "Big Picture"

1 つ目の talk は，領域分割問題の名前に「Mumford–Shah」として登場する Mumford 氏.

David Marr が問題を High/Middle/Low レベルに分割し，それぞれの問題はだいたい解けた．だが，一般の vision machine には近づいたのだろうか？ここでは Marr に足りなかったものを 4 つ挙げている：トップダウンであること（Marr はボトムアップ），部分的・あいまいなものも扱うこと，階層的 tree になっていること，学習に基づくこと．これらのアイデアを元々言い始めたのは Ulf Grenander であり，「pattern analysis = pattern synthesis」つまりトップダウンとボトムアップが協調しなければならない．そして，画像解析のための文法（階層的な re-usable part の集合）が必要であるとの信念を協調.

内容が散逸していて，あまりよく理解できず．ある人によれば，10 年前と talk の内容が変わっていないとのこと....

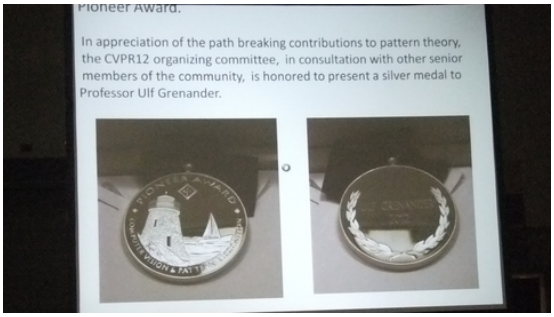
David Mumford の講演.

Grenander に代わって Geman: Two Lessons from Ulf Grenander's Second Carrier

2 つ目の講演は，Ulf Grenander の予定だったが，体調が悪いため（89 才！）欠席とのこと（1 日目の opening で説明があった）．その代わりに，MRF で有名な「Geman&Geman 論文」の Stuart Geman が登場[注4]．Grenander の研究から得られた 2 つのことを解説：abstraction の必要性，complex な combinatorial を扱うこと.

Stuart Geman の講演.

その後，Co-general Chair の Song Chung Zhu から Grenander の研究や業績等の紹介があり（ある著書は 900 ページもあるとのこと[注5]），committee から Grenander に silver medal が送られた．本人に代わって，娘さんとお孫さんが壇上でメダルを受け取る.



メダルと授与式.

Jitendra Malik が呼ばれて，壇上で熱心に説明していた．ある人によれば，Grenander の研究は誰にも理解されていなかったが，David Mumford やその学生だった Song Chung Zhu が広めたとのこと．今ではその名を知る人は少ないのでは.

4.4 Orals 2B: Optimization Methods

1. Incremental Gradient on the Grassmannian for Online Foreground and Background Separation in Subsampled Video, Jun He, Laura Balzano, Arthur Szlam

短いタイトルは online subspace learning from subsampled data．これを low-rank matrix completion として解く（nuclear norm 最小化問題にして）．subspace を逐次的に求めるので grassman.「GRASTA」で検索するとコードがある．応用は背景モデリングなので，背景差分を会場でデモ（macbook で）.

CVPP のオーラルでその場でデモをする人を始めてみた.

2. Curvature-Based Regularization for Surface Approximation, Carl Olsson, Yuri Boykov

SfM 点群を surface にしたい．その場合に，3 次元点をグラフのノードとして，2 次 smoothness（つまり曲率）を推定する．通常は 3 つのノードが必要になるが，2 ノードと各ノードでの tangent を用いる（この方がクリークサイズが小さい）．曲率計算は三角形の辺の長さの比で代用．計算時間は 50,000 点で 3 時間.

3. General and Nested Wiberg Minimization, Dennis Strelow

$\min_{U,V} f(U,V)$ という問題を解く場合，通常は U と V を交互に推定する（EM, alternating LS, alternating LP）か，同

(注4)：Grenander も Geman も Mumford も，今の所属は Brown University. http://www.dam.brown.edu/ptg/participants.shtml この大学は今回の学会会場のすぐ近く.

(注5)：おそらく Ulf Grenander, General Pattern Theory: A Mathematical Study of Regular Structures, Oxford University Press, 1994 のこと.

時に推定する（LM, Newton-Raphson）．もう一つの方法がWiberg で，片方を消去する方法．従来は行列のみを対象にしていた（線形だから）が，ここでは非線形の一般の問題を扱う．

4. A-Optimal Non-negative Projection for Image Representation, Haifeng Liu, Zheng Yang, Zhaohui Wu, Xuelong Li

NMF で得られた特徴を Regression でどう使うか？という問題を立て，回帰予測誤差を細小にするような NMF を求めるANP を提案．

4.5 Posters2B: OptimizationMethods, Motion & Tracking

5. A Tiered Move-making Algorithm for General Pairwise MRFs, Vibhav Vineet, Jonathan Warrell, Philip H. S. Torr

任意の pairwise term を扱うための t-拡張を提案．

10. Robust Maximum Likelihood Estimation by Sparse Bundle Adjustment using the L1 Norm, Zhijun Dai, Fengjun Zhang, Hongan Wang

L1-バンドル調整のために，L1 ノルム内の関数を Taylor 近似，内点法で解く．

12. A Bundle Approach To Efficient MAP-Inference by Lagrangian Relaxation, Jorg Hendrik Kappes, Bogdan Savchynskyy, Christoph Schnorr

単一ではなく，複数の劣微分方向をまとめて（バンドルして）用いる．劣微分方向を求めるのは組み合わせ最適化を解く．

17. Fast Dynamic Programming for Labeling Problems with Ordering Constraints, Junjie Bai, Qi Song, Olga Veksler, Xiaodong Wu

領域分割の境界線を DP パスとして求める．

4.6 Orals2D: StatisticalMethods & Learning

1. Learning Rotation-Aware Features: From Invariant Priors to Equivariant Descriptors, Uwe Schmidt, Stefan Roth

回転不変であるべき応用は多い（ノイズ除去や検出等）．そこで，特徴を学習する際に，特徴間の変換（この場合は回転）を陽にモデル化する．回転を取り入れた R-FoE は，画像が回転していても同じノイズ除去性能を実現．同じく回転を取り入れた・回転に不変な特徴量 EHOF/IHOF を提案し，航空画像中の車両（方向がバラバラ）を検出．

2. QsRank: Query-sensitive Hash Code Ranking for Efficient ε-neighbor Search, Xiao Zhang, Lei Zhang, Heung-Yeung Shum

多数のデータを検索する場合にはハッシュがよく用いられる．しかしハッシュバケツをどうランキングするのかはあまり研究されていない．そこで，PCA ＋バイナリ化でハッシュを生成し，Hamming 距離でランキングする．これを ε-NN に応用．

3. Geodesic Flow Kernel for Unsupervised Domain Adaptation, Boqing Gong, Yuan Shi, Fei Sha, Kristen Grauman

Adaptation の時に，ソースドメインとターゲットドメインで分布が違う場合にどうするか（しかも unsupervised）．ここでは，データを線形部分空間でモデル化し，ソースとターゲットの部分空間の対応をグラスマン多様体上で移動させる（これがGeodesic flow）ことで実現．この geodesic flow での移動をサンプルしてカーネルにすることで．Domain invariant なカーネルを実現．

4. Supervised Hashing with Kernels, Wei Liu, Jun Wang, Rongrong Ji, Yu-Gang Jiang, Shih-Fu Chang

LSH などは unsupervised なハッシング．ここでは supervised なハッシングとして，similar/dissimilar ペアを学習．± 1 のデータベクトルの内積は Hamming 距離で表現できることを利用して，ハッシュ計算をカーネル計算で置き換える．

4.7 Posters 2C: VideoSurveillance, Statistical Methods & Learning

ポスター会場入り口の前で，Marc Pollefeys がポスターセッションの間ずっと誰かと立ち話をしていた．30 分後に行ってみると，まだ同じ場所で話をしていた...

9. Power SVM: Generalization with Exemplar Classification Uncertainty, Weiyu Zhang, Stella X. Yu, Shang-Hua Teng

Uncertainty を考慮した SVM．定式化で dual 問題と primal 問題を対比しており，今までに見たことがない SVM の見方だった．

10. Active Image Clustering: Seeking Constraints from Humans to Complement Algorithms, Arijit Biswas, David Jacobs

active learning のクラスタリング版．ユーザーに「同じクラスになるべきか否か」を問い合わせ，クラスタリング．

14. Image Categorization Using Fisher Kernels of Non-iid Image Models, Ramazan Gokberk Cinbis, Jakob Verbeek, Cordelia Schmid

画像中のパッチは iid ではない．そこでこれまでの bag-of-words モデルを拡張して，non-iid の BoW を fisher kernel でモデル化．

18. Semi-Coupled Dictionary Learning with Applications to Image Super-Resolution and Photo-Sketch Image Synthesis, Shenlong Wang, Lei Zhang, Yan Liang, Quan Pan

dictionary learning と，マッピング（画像・スケッチ間，低解像度・高解像度間）を同時に学習．

20. Efficient Discriminative Learning of Parametric Nearest Neighbor Classifiers, Ziming Zhang, Paul Sturgess, Sunando Sengupta, Nigel Crook, Philip H. S. Torr

（名城大の堀田先生より）「nearest neighbor の距離に重みを付けたものを parametric nearest neighbor と呼んでいる。重みだけでなく、プロトタイプも学習するが、局所解しか得られない。複数の初期値から得られたものを ensemble することにより、非線形 SVM に近い精度を実現した。」

4.8 Lobster banquet

バンケットの会場では，アンケートに答えて Microsoft Kinect が当選した人や Google の旅費補助に当選した人，Outstanding reviewers が紹介された．

4.8.1 Best Open Source Code Award

続いて Willow Garage 出資による Best Open Source Code Award の紹介．

- 1 位（$800）FREAK: Fast Retina Keypoint, Alexandre Alahi, Raphael Ortiz, Pierre Vandergheynst
- 2 位（$600）A New Mirror-based Extrinsic Camera Calibration Using an Orthogonality Constraint, Kosuke Takahashi, Shohei Nobuhara, Takashi Matsuyama
- 3 位タイ（$300）Evaluation of Super-Voxel Methods for Early Video Processing, Chenliang Xu, Jason J. Corso

Action Bank: A High-Level Representation of Activity in Video, Sreemanananth Sadanand, Jason J. Corso

4.8.2 Best papers

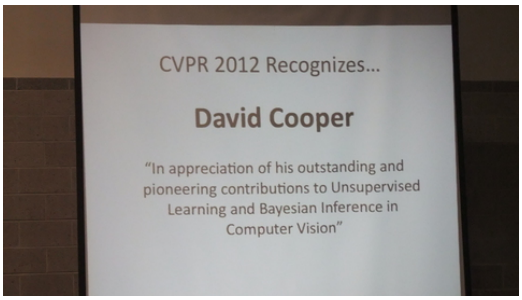
続いてはベストペーパーが紹介された.

- best student award: Max-Margin Early Event Detectors, Minh Hoai, Fernando De la Torre
- best paper award: A Simple Prior-free Method for Non-Rigid Structure-from-Motion Factorization, Yuchao Dai, Hongdong Li, Mingyi He

Honorable mention は無し.

4.8.3 ?

続いて，何の賞なのかよく分からなかったが，表彰されたのは David Cooper.



David Cooper の紹介.

Cooper による受賞スピーチがあったのだが，会場が大きすぎて音が反響しあまりよく聞き取れないことと，バンケットの最中だったので誰もが話をしていてうるさくてよく聞こえず（つまり誰も聞いていない），内容は分からなかったまま．悲しいことに，この状況でスピーチは 10 分程度続き，結局 Co-general chair の Zhu に早く終わるように促されていた.

5. 20th June, 2012: 3rd Day

5.1 Posters 3A: Face & Gesture, Human ID, Document Analysis, Scene Understanding

ポスター発表していたので他の研究は聞けず.

8. Linear Discriminative Image Processing Operator Analysis, Toru Tamaki, Bingzhi Yuan, Kengo Harada, Bisser Raytchev, Kazufumi Kaneda

MIRU2011 の発表内容＋解析.

35. Automatic Discovery of Groups of Objects for Scene Understanding, Congcong Li, Devi Parikh, Tsuhan Chen

（名城大の堀田先生より）「Visual phrase に刺激され（本人談)、個別の対象ではなく対象の集合をモデル化する方法。アノテーションデータを使って対象間の位置、大きさ、視点の関係を投票することにより自動的にグループを発見する。」

5.2 Invited Talk: Sebastian Thrun

2005 年の DARPA グランドチャレンジ（自動走行の車で長距離を走る）で優勝したチームを率いていた人．今回は Google car についてのトーク.

直前のポスターセッションが終わっても，ポスター会場にはまだまだ人が残っていたので，このトークの最初からは参加できなかった.

「TED で見せた動画」と言って紹介していたのはおそらくこれ.

- Sebastian Thrun: Google's driverless car http://www.ted.com/talks/sebastian_thrun_google_s_driverless_car.html
- A spin in the Google Self-Driving Car at TED2011 http://blog.ted.com/2011/03/05/a-spin-in-the-google-self-driving-car-at-ted2011/

会場には多数の椅子があったにもかかわらず，多数の立ち見が出現．Antonio Trallba は床に座りながら壁に寄りかかって見ていたし，Fernando Te la Torre も床に座ってノート PC で仕事していた.

超満員の Sebastian Thrun の講演会場.

5.3 Orals 3A: Video Analysis & Event Recognition

1. Detecting Activities of Daily Living in First-person Camera Views, Hamed Pirsiavash, Deva Ramanan

First Person Camera での認識のための object centric feature を提案．First person camera に写る物体の見えは，遮蔽が多く，扉が開いたり閉じたりする機能な変形が多い．またデータは短いショットではなく長時間の映像．そのために，bag-of-object を passive/active state detector を組み合わせたものと，更に temporal pyramid 特徴（spacial pyramid にヒントを得たもの）を提案.

2. Discriminative Virtual Views for Cross-View Action Recognition, Ruonan Li, Todd Zickler

active recognition は view が変わってしまうと特徴も変わってしまう．そこで adaptation．特徴ベクトル x に変換 W^T（縦長の行列）をかけて，次元を拡張．この W の学習には，陽に discrimination を考慮した項と unsupervised ペアの項を組み合わせる.

3. Max-Margin Early Event Detectors, Minh Hoai, Fernando De la Torre

Best student award の論文．イベントの早期認識．全ての部分イベントを学習するが，アイデアは時刻 t のイベントのスコア $f(x_t)$ はそれ以前の全ての時刻のスコアよりも高くなるべき，というもの．数式では $f(x_t) - f(x) > 0$ で，この右辺を少し変えて $f(x_t) - f(x) > \delta(p, p_t)$ とする．この δ が adaptive margin で，さらに SVM のようにスラック変数を入れて $f(x_t) - f(x) > \delta(p, p_t) - \xi$ という定式化にする．これが convex なので効率的に解ける．イベント認識をマージン最適化問題に置き換えたところが評価されたか.

4. Understanding Collective Crowd Behaviors: Learning a Mixture Model of Dynamic Pedestrian-Agents, Bolei Zhou, Xiaogang Wang, Xiaoou Tang

駅構内で大量の人が行き来する映像の解析．Dynamic pedestrian agent モデルで歩行者を表現するので，ルールはシンプルだし歩行者シミュレーションもできる．しかし全ての人をずっと追跡することは困難なので，軌跡の fragment を扱う．これ

をつなげるために１．明確なスタート位置とゴール位置がある，２．その位置間を歩行する，３．スタートするタイミングは確率モデルに従う，という仮定を利用．応用として，歩行軌跡のクラスタリング，歩行者シミュレーション、異常検出など．詳細は www.crowdbehavior.org を参照.

5.4 Posters3B: Image & Video Retrieval, Object Detection

4. Multi-Attribute Spaces: Calibration for Attribute Fusion and Similarity Search, Walter J. Scheirer, Neeraj Kumar, Peter N. Belhumeur, Terrance E. Boult

SVM decision score を補正して，正規化されたアトリビュート空間を作る（校正）.

5. D-Nets: Beyond Patch-Based Image Descriptors, Felix von Hundelshausen, Rahul Sukthankar

点同士のマッチングではなく，網目状に descriptor をつなげて Net にしてしまう.

7. Spherical Hashing, Jae-Pil Heo, Youngwoon Lee, Junfeng He, Shih-Fu Chang, Sung-Eui Yoon

超球を用いたハッシング．LSH よりも高精度.

11. Nonparametric Kernel Estimators for Image Classification, Barnabas Poczos, Liang Xiong, Dougal J. Sutherland, Jeff Schneider

BoW を生成する高次元分布を仮定して，それを分類に用いる support distribution machine を提案.

26. Fast Search in Hamming Space with Multi-Index Hashing, Mohammad Norouzi, Ali Punjani, David J. Fleet

全探索ではないハミング距離での n-近傍探索．2bit までの違いを許容する場合には，ビット列を 3 つのチャンクに分けるとどれか一つのチャンクは同一になるはず．これを利用して高速化.

38. Large-scale Knowledge Transfer for Object Localization in ImageNet, Matthieu Guillaumin, Vittorio Ferrari

ImageNet の画像に bounding box がついているものが少ないので，bb がついているデータから pos/neg サンプル，位置，objectness 等を学習，500M 毎分の bb を作成.

41. Steerable Part Models, Hamed Pirsiavash, Deva Ramanan

少数のフィルタで表現する steerable フィルタの考えを用いて part model を構成.

5.5 Orals3C: Vision Systems

1. Street-to-Shop: Cross-Scenario Clothing Retrieval via Parts Alignment and Auxiliary Set, Si Liu, Zheng Song, Guangcan Liu, Changsheng Xu, Hanqing Lu, Shuicheng Yan

女性の全身像のスナップ写真から，その人が着ている服の候補一覧を作成，オンラインショップまで 1 クリック．これを実現するために，スナップ写真（背景雑多，姿勢多様）とオンラインショップの写真（背景は一様の白，姿勢は類似したモデル）との差を埋めなければならない（この 2 つがシナリオが違うために，クロスシナリオと呼ぶ）．直接この 2 つをつなぐのは難しいために，多数のスナップ写真から構成される aux セットを用意して，まずテストのスナップ写真を aux セットで検索．次に aux セットとオンラインショップ写真の類似度をスパース性を用いて定義．応用例として，デートのときに着るトップスに合うボトムズを選んでくれる.

2. Autonomous Cleaning of Corrupted Scanned Documents - A Generative Modeling Approach, Zhenwen Dai, Jorg Lucke

文書に落書きされたら，どう修復するか？一般的には OCR もインペインティングもデノイジングもだめ．そこで，落書きは同じものはないが，文書中の character は同じものが繰り返し出現するという性質を利用して，文書を修復．アルファベット数が少なければ言語によらない．文書でなくても良い（例として丸い細胞の画像から細胞を検出）.

3. A Theory of Multi-Layer Flat Refractive Geometry, Amit Agrawal, Srikumar Ramalingam, Yuichi Taguchi, Visesh Chari

カメラと物体の間に屈折がある場合の幾何．屈折物体が多層平面であることを仮定して，屈折率や各層の厚さや層の数が未知でも，3D 点がどのように 2D 点として写るのかをモデル化.

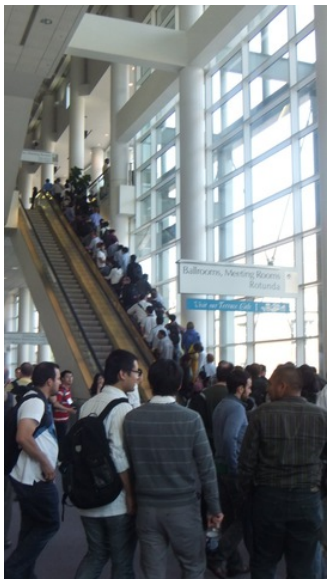
4. Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite, Andreas Geiger, Philip Lenz, Raquel Urtasun

KITTI データセットの紹介．車の自動運転の研究はあるが，レーザースキャナと詳細地図が必要．vision として解くべき（カメラを用いる）問題はどこまで解けているのかを計るためのベンチマークを作成.

・キャリブレーション済：カメラ間，カメラ―レーザースキャナ間，GPS ―レーザースキャナ間の校正.

・アノテーション：22 人で 3 次元物体をアノテート．遮蔽ラベルは Amazon mechaical turk を利用．ほとんどのラベルは車，人，自転車.

なぜ今更ステレオデータセットを作るのか?: Middlebury データセットは現実的ではない．それでうまく行っているアルゴリズムでも KITTI データセットでは全く性能が悪くなってしまう.



最後のオーラルが終わって，最後のポスターセッションへと向かう人々.

5.6 Posters 3C: ObjectRecognition, Performance Evaluation

10. A Codebook-Free and Annotation-Free Approach for Fine-Grained Image Categorization, Bangpeng Yao, Gary Bradski, Li Fei-Fei

（名城大の堀田先生より）「コードブックもアノテーションも使わない画像認識法。ランダムに生成したテンプレートによるマッチングスコアと Bagging を用いて識別する。一周回って戻ってきた感じ。実験によりアノテーションに基づく方法や codebook に基づく方法よりも精度が高いことを示した。」

11. Discovering Localized Attributes for Fine-grained Recognition, Kun Duan, Devi Parikh, David Crandall, Kristen Grauman

ユーザーインタラクションを用いて画像にアトリビュートを付ける．

35. Pose Pooling Kernels for Sub-category Recognition, Ning Zhang, Ryan Farrell, Trevor Darrell

poselet 同士のカーネルを定義．

8. Hedging Your Bets: Optimizing Accuracy-Specificity Trade-offs in Large Scale Visual Recognition, Jia Deng, Jonathan Krause, Alexander C. Berg, Li Fei-Fei

（名城大の堀田先生より）「画像のラベル付けではなるべく間違えないことが重要であるという考えに基づく方法。例えば、キツネの画像が与えられて、ハイエナと間違うよりもイヌ科と答えた方が確信度が高いというもの。ただし、キツネやハイエナの上の階層がイヌ科というの木構造は人間が事前に与えている。」

15. Discriminative Spatial Saliency for Image Classification, Gaurav Sharma, Frederic Jurie, Cordelia Schmid

（名城大の堀田先生より）「 spatial pyramid matching の各領域の重要度を latent SVM を用いて学習した。得られた重要度をその領域のヒストグラムにかけたものを特徴量とした。精度は向上しているが、位置ずれに弱いと考えられる。」

6. 21th June, 2012: Workshop / Tutorial day

main conference が終わった翌日もワークショップが盛りだくさん．参加者はどのくらいいるの気になっていたのだが，実際には CVPR 前日のワークショップ程度の参加者が残っていた．

post conference workshop のブレイクの様子．

7. 雑　　感

- 時代は今やビッグデータ．論文タイトルにも large scale というキーワードが多い．大量データを高速に扱うための binary 特徴量，hashing や indexing，nearest neighbor 等も盛ん．
- カーネルの発表もあったが，large scale にどうやってカーネルを用いるのかという視点が多い．カーネルについての 2 件のオーラル (Orals 2D) は，domain adaptation と hashing についての研究である，という点が象徴的．
- SfM が花盛り．Bundler の登場で一旦は SfM の発表が少なくなったように見えたが，昨年から Kinect や PCL が登場し，point cloud 処理の重要性が増してきたことで，3 次元形状取得の研究が盛り返してきたように思える．同様に photometric stereo も多数発表されていた．
- シーン理解は outdoor 一辺倒から indoor 台頭へ．First person camera の研究が増えてきたことと，Kinect（屋内限定デバイス）を使って depth を使う研究が登場してきたたため．action 認識の研究として料理シーン（屋内）を対象とした研究が 2 件あったので，今後増えるかも．
- グラフィカルモデルはあちこちで利用されている．単純なグラフィカルモデルを作って contex と言っていた時代から，多数のノードを多層につなげた Deep network/learning の時代へ移行か．

なお USB 予稿集に不備があったとのこと．CVPR2012 の web サイトで修正ファイル (zip) が公開されてるので，予稿集を入手された方は注意してほしい．

8. CVPR に accept されるまで

今回初めて発表者として CVPR に参加した．それまでの道のりを簡単に記しておく．

- 最初の投稿は MIRU2010．査読結果はオール 1.（ちなみに 5 段階評定で 5 がもっとも良い.）発表を取り下げようとしたら「今更ダメです」という返答をもらった....
- 次の投稿は MIRU2011．これまで MIRU に投稿した論文の中でも最もよい査読点数をもらったが（オーラルに通ったもう一本の論文よりも点数は良かった），結果はポスター．ポスター賞もとれず（同じセッションで発表されていた CARD が受賞）．
- ICCV2011．結果は definitely reject，weakly reject，weakly accept．エリアチェアからのコメントは “I read the paper and rate it also “definitely reject” for minor novelty and too weak experimental validation.” つまりダメダメ．出直してこい.（おそらく読んでないけど読んだことにして reject 用のコメントを使い回していると想像.）
- ICASSP2012．京都であるしレベルもそれほど高くないし 4 ページだし，手堅くここに発表してさっさと始末してしまおうと思った．でも共著者の Raytchev 先生に「それはもったいない，面白い論文だし，通るかどうか分からないけれど出さなければ通らない．もう一回 CVPR にチャレンジしてみましょう」と言ってもらった．これがなければ投稿していない．
- CVPR2012．結果は borderline, borderline, weakly accept．エリアチェアからのコメントは以下の通り．

> The theory behind the paper is well presented, well motivated and sound.
> The main concerns are at the exps side, including both database and features.
> The rebuttals of the authors are effective, and re-

solved some concerns from the reviewers.
Given that most reviewers are rigorous to face recognition papers in CV area, this paper is recommended to be accepted as Poster. But the authors are encouraged to improve the exps part in final version.

判定は Definitely Accept．現在追加実験をして投稿準備中.

謝　　辞

このレポートの執筆には多くの方から頂いた意見を参考にさせて頂いた．特に堀田一弘先生（名城大）からはメールで一言コメントを送っていただいたものをそのまま掲載させて頂いた．ここに感謝致します.

文　　献

[1] 玉木徹, ACCV2010 報告, 第 2 回広島画像情報学セミナー, 2010/11/17. http://ir.lib.hiroshima-u.ac.jp/metadb/up/ZZT00001/accv2010report.pdf

[2] 玉木徹, CVPR2011 報告, 第 7 回広島画像情報学セミナー, 2011/7/1. http://ir.lib.hiroshima-u.ac.jp/metadb/up/ZZT00001/cvpr2011report.pdf

[3] 安倍満, 石川博, 岩村雅一, 坂上文彦, 佐藤いまり, 佐藤真一, 杉本茂樹, 玉木徹, 西山正志, 阮翔, CVPR2011 報告, Vol.2011-CVIM-179/2011-CG-145, No.18, pp.1–8, 2011. http://www.f.waseda.jp/hfs/CVIM-CVPR2011rep.pdf

[4] CVPR2012 Pocket Guide. http://www.cvpr2012.org/announcements/pocketguideavailablepdf

[5] Laurent Charlin, Richard S. Zemel, Craig Boutilier, A Framework for Optimizing Paper Matching, UAI2011, pp.86-95, 2011. http://uai.sis.pitt.edu/papers/11/p86-charlin.pdf