

ACCV2010 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 ACCV2010 の報告 . 修正版 .

キーワード ACCV

A report on ACCV2010

Toru TAMAKI†

† Hiroshima University 1-4-1 Kagamiyama, Higashi hiroshima, Hiroshima, 739-8527 Japan

E-mail: †tamaki@hiroshima-u.ac.jp

Abstract An ACCV2010 report. Corrected edition.

Key words ACCV

1. Introduction

これは , New Zealand で開催された ACCV2010 (the Tenth Asian Conference on Computer Vision) の報告である .

- 開催場所: The Rydges Lakeland Resort Queenstown, Queenstown, New Zealand
- 開催期間 : 8-12 November 2010

最初の 2 日間 (11/8,9) は workshop/tutorial/ICVNZ が開催され , Main conference は 11/10,11,12 の 3 日間にわたって開催された . oral session はすべて single track (会場約 400 席 , スクリーン 3 つ) , poster session は毎日 60 件のポスター発表が 2 部屋を使って行われた . 3 日間とも朝 9:00 から Keynote speech が oral 会場で開催された .

Queenstown は New Zealand の南島の南に位置し , 開催期間中の気候は初夏 . 日中が最高気温 20 ~ 25 ° で町中には半袖の観光客が多いが , 夜になると冷え込んで 10 ° を下回り , 屋外テラスのあるレストランでは毛布を客に貸し出すほどの寒さになる . 北半球から来た人間にとっては , どんな服装をすればいいのかということに最初は戸惑うことも多かった . 幸いにも ACCV 開催期間中には太陽に恵まれ , 多少曇りや風の強い日があったものの , おおむね観光には適した天候であった .



Queenstown 風景



会場となった Rydges

以下では , Main conference の keynote speech と oral session の内容 , poster presentation のいくつかを紹介する . ただし , 紹介する 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。

2. 8, 9th Nov. 2010

ACCV に先立つ 2 日間は、8 つの workshop と 4 つの tutorial が行われた。

- Workshops
 - Application of Computer Vision for Mixed and Augmented Reality
 - Computational Photography and Aesthetics
 - Computer Vision in Vehicle Technology: From Earth to Mars
 - Event Categorization, Tagging and Retrieval (VECTaR 2010)
 - Gaze Sensing and Interactions
 - Subspace 2010
 - Visual Surveillance 2010
 - e-Heritage (Electronic Cultural Heritage)
- Tutorials
 - Advances in Visual Tracking
 - High Dynamic Range Imaging (Cancelled)
 - Human Centred Vision Systems
 - Implementing Image Processing on FPGAs

なお New Zealand では毎年 Image and Vision Computing New Zealand (IVCNZ) が開催されているが、今年は ACCV があるために、ACCV workshop と同じ 11/8,9 に開催された。会場も全く同じで、IVCNZ と ACCV workshop が隣近所の部屋で行われていた。IVCNZ submission deadline は ACCV notification の後に設定されており、開催には苦労したと思われる。

Workshop 2 日目の夕方には Welcome Reception が会場の一室で行われた。椅子はほとんどなく、メインは drink (food はあまりなし)、参加者は皆熱心に会話を続けていた。



Welcome Reception

3. 10th Nov. 2010: 1st Day

3.1 Registration

今回の ACCV proceedings は、CD/DVD ではなく、USB メモリ (ACCV ロゴ入り)。しっかりした conference bag は赤

と青の 2 色。お土産には 4D view のマグカップ。



Conference bag, USB proceedings, Magcup

3.2 Opening

General Chair の Brendan McCane (地元 New Zealand の University of Otago) が登場。

Brenden: Good morning!

audience:

Brenden: Ok, 練習しよう (爆笑) 会場の右側から, Good morning!

右側の **audience:** Good morning!

Brenden: 左側も, Good morning!

左側の **audience:** Good morning!



右側を練習する Brenden

Workshop 開催中から、会場では部屋ごとに Wi-Fi が提供されていたが、機器の不具合か、接続数が多すぎたためか、繋がりにくい現象が多発していた。そのため、今日からは oral session 中は Wi-fi を止めるので、talk に集中できますね、とのアナウンスが Brenden から。



Keynote / Oral 会場は横長なのでスクリーンが 3 つ

3.3 Keynote

Deformable Object Modelling and Matching
Tim Cootes (Univ. of Manchester, UK)



Tim Cootes

ACCV 初日の keynote は AAM で有名な Tim Cootes . 目的は, model を image に fit して, image を統計的に理解すること . そのためには, model をどう構築するか, fit のための matching をどうするか, の 2 点が問題になる .

1. Model building

顔のような対象の場合, 多数の学習画像を与えて統計的に学習を行う . この時に, 顔を表現するモデルは 2 つあり, shape model と appearance model である .

1.1

shape model は, いくつかの点を線で結んだモデルで表現する . この点のばらつき具合を, 以下の手順で統計的にモデル化する .

labeling まず顔上の特徴的な点 (landmark) を手動で与える .
aligning 個人毎の形状のばらつきを補正する . きちんと補正すれば, 点の分散は小さくなる .

PCA 補正された shape model を多数集めて, 主成分分析を行い, 固有ベクトルで表現 .

例: 顔画像 10 人分各人 4 枚, 手の X 線写真, 背骨の X 線写真, 3D 形状 (shape はメッシュで表現) ,

1.2

appearance model は, 形状変形が補正された後のテクスチャ画像である . これは形状情報は持っていないため, 新しい顔を生成するには二つのモデルを組み合わせる shape model + appearance model = new face generation というを行う .

2. Model matching

model と画像のマッチングには, ASM と AAM の 2 つ手法がある .

Active Shape Model (ASM)

顔形状は 3 次元点で表現し, 各点を画像にマッチングさせる . これは Toyota の研究らしい?

Active Appearance Model (AAM)

これが Cootes の研究 . AAM のアイデアは, モデルパラメータの変形を画像から予測するために学習を行うこと . つまり, パラメータ p の変化分 δp と画像特徴 S が与えられたとき,

$$\delta p = F(S)$$

を満たすような関数を学習によって得る . 基本的には回帰ベースの手法で, 線形回帰でも非線形回帰 (Harr 特徴を用いる) で

もよい . 非線形回帰を用いた手法 (BMVC2010) は, 実は線形回帰と性能がほとんど変わらないのに計算時間はかかること .

1 つの AAM だけでは対応できない場合は, いくつかの AAM をつなげればよい . まず低解像度・少数パラメータの AAM でマッチングして, 徐々に高解像度・多数パラメータの AAM を用いる sequence of AAMs は, 現在は ms で実行可能とのこと .

例: 顔検出, 脳の structure, 背骨 . ノキア製の mobile phone で AAM Tracking が実時間で動作している動画も紹介された .

3. Groupwise Registration

AAM で最も手間がかかるのが labeling, つまり点を手動で与える部分である . これを完全になくすためには, 画像の集合を入力したら対応する点を自動的に見つける .

3.1 find corresponding points

対応点を見つけるためには, どの程度よい対応なのか? どれが best か? という評価基準が必要である . ここでは MDL を用いて,

$$\begin{aligned} & total \text{ のコード長 } C_{total} \\ &= \text{モデルの記述長 } C_{model} \\ & \quad \text{パラメータの記述長 } C_{param} \\ & \quad \text{モデルによる画像の復元誤差 } C_{residual} \end{aligned}$$

が最も短くなる対応を採用する .

3.2 変形の表現

画像を変形するために, ここでは piecewise linear warp field を採用 .

3.3 アルゴリズム (PAMI2010)

- 初期化
- 制御点を生成
- 以下 loop
 - モデルを生成
 - 最適なパラメータを推定

3.4 画像特徴

画素値そのもの (raw image) のほかに, 以下のものも同時に使用する .

- linear normalization
- linear z-normalization
- x,y の gradient

3.5 例

- 顔画像
- 3D MRI 画像の registration

3.6 Parts+Geometry model

初期化は affine registration を用いているが, local minimum が多く, ダメな場合が多い . そこで, 局所特徴 (Parts) と, その相対位置関係 (Geometry) を初期マッチングに用いる . どの Parts を利用するのは, Genetic Algorithm を用いて選択する . 手の X 線写真の場合, 3 parts で可能, 5 parts で better , 30 parts で best なマッチングができた .

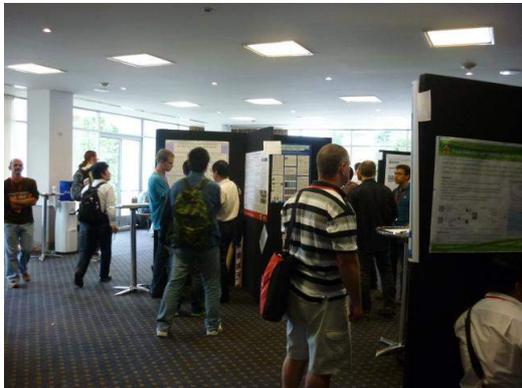
4. Medical Application

Synovitis in Rheumatoid Hand という例を紹介 . リウマチ

の手の3次元復元の可視化(医学専門用語が多く理解困難...)

3.4 Break and Posters

Keynote や oral session の間の break (coffee とデザートあり) は、20 分程度と短いながらも poster セッションの時間がとられていた。以外にもこの時間が、poster 会場はにぎわっていた。おそらく lunch time は外へ食事に出るため、break の時間に見ておこうという参加者が多かったためではないだろうか。



Poster sessionの様子

3.5 Geometry & Correspondence

New efficient solution to the absolute pose problem for camera with unknown focal length and radial distortion.

Martin Bujnak, Zuzana Kukelova, Tomas Pajdla (Microsoft LIVE Engagement Slovakia & Czech Technical University)

P4P + unknown focal length + radial distortion のための新しい解法。射影行列 P をパラメータ化しなおし、シーンが non-planar の場合には 4 equations for 4 unknowns, planar の場合には 3 equations for 3 unknowns が得られる。これをグレンナ基底 solver(ECCV2008) にかける。一般のシーンに対応するために、planar/non-planar の combination も提案。

注: P4P (Perspective 4 points (だったか?)) は、画像と3次元点の対応が4点与えられたときのカメラ姿勢を求める問題。

Efficient large-scale stereo matching

Andreas Geiger, Martin Roser, Raquel Urtasun (KIT, Germany);

画像サイズ 2Mpix までの大きさのステレオ復元を数秒で実現 (graph-cut では 2000 秒)。グリッド上の点でまず粗く簡単なマッチングを行っておき、これを prior として (piecewise linear manifold で近似・補間して) 密なマッチングを行う。likelihood には laplacian。

3D デジカメで撮った画像から顔の 3D を復元したり、街のストリート (100m 程度? の広範囲) を復元したりする例を紹介。

Towards full 3D Helmholtz stereovision algorithms

Amael Delaunoy, Emmanuel Prados, Peter Belhumeur (INRIA Rhone-Alpes, France);

普通のステレオは物体表面がランバート面であることを仮定するが、Helmholtz stereo は任意の反射特性を持つ表面に適用可能 (2002 年以降研究が盛ん)。しかしこれまでは 2.5D の復元しかできなかった (つまりカメラから見た奥行きのみ)。これを、物体をぐるっと撮影した画像から 3D 形状を復元する方法へ拡張。物体形状はメッシュで表現し、各メッシュ状で Helmholtz stereo 方程式を解く変分法アプローチ。合成画像 (ハイライトが激しい CG 球の形状もきれいに復元できている) と実画像で復元を評価。計算時間についての質問には、数時間かかるとの回答。

Image-based 3D modeling via Cheeger sets

Eno Toppo, Martin Oswald, Daniel Cremers, Carsten Rother (TU Munich, Germany);

画像 1 枚から 3 次元形状を復元する。形状は曲面 (平面に限定しない)、ユーザー入力は極力排除する、という方針。まず、物体と背景のそれぞれに短いストロークをユーザー描いて、graph cut で物体のシルエットを抽出。形状を決めるための基準は、表面はなめらかで表面積が小さく、シルエットと同じ輪郭を持ち、体積はユーザーが指定した値を持ち、(オプションで) ユーザーが指定したエッジ部分では表面の不連続を許す、というもの。さらに問題を緩和して convex problem にしてから ICCV2009 の手法 (by D. Cremers) を適用。いくつかの例では、丸っこい物体はうまく復元できているように見える。

Network connectivity via inference over curvature-regularizing line graphs

Maxwell Collins, Vikas Singh, Andrew Alexander (UW-Madison, USA)

各 voxel が方向を持つような Diffusion MRI のデータでは、その方向を接続した path を復元することが重要 (脳内の神経線維がその方向に延びていることがわかるため)。そこで、与えられたデータに対して path を見つけるための graph-based な手法を提案。

いろいろな話が散漫して内容は理解できず。

3.6 Lunch

Lunch time の 3 時間 (12 時から 15 時) は、poster session のメインの時間帯に割り当てられていた。街までは歩いて 10 分程度かかるため、参加者には Rydges の Buffet 形式 lunch。



Oral session の様子



Lunch

3.7 Posters

Generic object class detection using boosted configurations of oriented edges

Oscar Danielsson, Stefan Carlsson

エッジ断片を特徴に使う、AdaBoost + V.J. カスケード方式で、形の似た二つのカップをエッジ情報で判別。

Emphasizing 3D structure visually using coded projection from multiple projectors

Ryo Nakamura, Fumihiko Sakaue, Jun Sato

MIRU と同内容。複数のプロジェクタである範囲のシーンを投影し、そのシーン内の物体の位置によってあてる照明を変化させる。カメラなし、PC なしの AR。

Specular surface recovery from reflections of a planar pattern undergoing an unknown pure translation

Miaomiao Liu, Kwan-Yee Wong, Zhihu Chen, Zhenwen Dai

固定カメラで鏡面反射する表面形状を計測するために、平面パターンを用いる。この平面パターンを適当に平行移動して2枚撮影すると、三角測量が可能。アイデアは、符号化平面パターンを使うこと。

Indoor scene classification using combined 3D and gist features

Agnes Swadzba, Sven Wachsmuth

部屋の3次元形状を平面パッチに分割し、パッチごとに3D

特徴量を算出。これと gist を組み合わせ、かつ時間的に識別を積算して安定した認識結果に収束させる。

Closed-form solutions to the minimal absolute pose problems with known vertical direction

Zuzana Kukelova, Martin Bujnak, Tomas Pajdla

カメラの上方向がわかれば、P2P 問題が 3 linear equations for 3 unknowns になり、マイクロ秒で計算可能。最近のデジカメには内臓センサで方向がわかるため、これを利用すると計算が高速・式が簡単・ロバスト、というアイデア。

Learning rare behaviours

Jian Li, Timothy Hospedales, Shaogang Gong, Tao Xiang

Topic モデルを使って異常検出。Multiclass Delta Latent Dirishlet Allocation を適用。

Real-time robust image feature description and matching

Stephen Thomas, Bruce MacDonald, Karl Stol

SURF より速い特徴量。円状の descriptor だが、画素のサンプリング位置を工夫し、固定サンプリング位置でも orientation 算出が高速。

Feature quarrels: The Dempster-Shafer evidence theory for image segmentation using a variational framework

Bjorn Scheuermann, Bodo Rosenhahn

Bayesian では、同じクラスの分布が競合すると（近くに存在すると）ピークが中間的な位置ずれた、元とは違う分布が生成されてしまう。これを解消するために、Dempster-Shafer 理論を適用。

Gait analysis of gender and age using a large-scale multi-view gait database

Yasushi Makihara, Hideoshi Mannami, Yasushi Yagi

25 視点 168 人の歩行データベースを作成。それをを用いた男女・世代認識結果と、男女間・世代間の特徴の違いを解析。

A system for colorectal tumor classification in magnifying endoscopic NBI images

Toru Tamaki, Junki Yoshimuta, Takahishi Takeda, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Yoshito Take-mura, Shinji Tanaka

大腸内視鏡画像中の癌の判別。dense SIFT を特徴量にした Bag-of-features を用い、SVM で識別、3 種類のカーネルで比

較・最高認識率 95.11% .

Efficient clustering earth mover's distance

Jenny Wagner, Bjorn Ommer

foreground/background の 2 クラス分離問題を、ヒストグラムベースで行う。高速・単純。ヒストグラムのピンに対する特徴選択のようなもの。

Minimal representations for uncertainty and estimation in projective spaces

Wolfgang Foerstner

SE(3), SO(3) の変換 (回転・並進など) の推定に用いられていた tangent space analysis を、projective geometry のすべての要素 (点・線・平面・cone など) に適用するフレームワークの提案。

3.8 Computational Photography & Low Level Vision

Image and video decolorization by fusion

Codruta Ancuti, Cosmin Ancuti, Chris Hermans, Philippe Bekaert (Hasselt University, Belgium);

カラーをグレースケールに変換するとき、カラーで色相が同じ場合にはグレーで差が出なくなってしまう、分かりにくい。それを何とかするのが decolorization。ここでは画像を R,G,B,H-Z チャネルに分解し、saliency map と exposedness から各チャネル各画素の重みを計算し、重みに従って各チャネルを blend(fusion) することでグレースケール画像を生成する手法を提案。

Video temporal super-resolution based on self-similarity

Mihoko Shimano, Takahiro Okabe, Imari Sato, Yoichi Sato (Institute of Industrial Science & University of Tokyo, Japan);

MIRU2010 オーラルと同内容。時間方向の自己相似性を利用して、時間方向の超解像を実現。オリジナルの動画 V0 を、時間方向に 2 倍積分した V1, 4 倍積分した V2, などを生成し、これらの間の関係を学習。MAP 推定で、時間方向に新しいフレームを生成する。

Temporal super resolution from a single quasi-periodic image sequence based on phase registration

Yasushi Makihara, Atsushi Mori, Yasushi Yagi (Osaka University, Japan);

日本語発表既出? (記憶が定かではないです)。低フレームレートの動画が周期的であれば、位相を補正してレジストレーションすることで超解像度ができる、という手法。人物の歩行者輪郭形状を対象にして、周期的な manifold を N-spline で表現し、位相を合わせる。このときに、まず linear approximation, 次に Dynamic programming, 最後に Quadratic Approximation という 3 段階を踏む。段階を経るごとに精度は向上する。

Solving MRFs with higher-order smoothness priors using hierarchical gradient nodes

Dongjin Kwon, Kyong Joon Lee, Il Dong Yun, Sang Uk Lee (Seoul National University, Korea);

一般的な MRF で使われる 1st order smoothness prior は、なめらかな対象には不適切なので、Higher Order Prior (HOP)

が最近使われている。ここで提案している方法は、Factor Graph 上の HOP を MRF エネルギーに利用して、MAP 推定するというもの。そのために、graph conversion method を利用する。通常の Factor Graph は Factor と node からなっているが、それを階層的に gradient nodes(HGN) を持つグラフに変換する。実験では、1st+2nd+3rd order prior をすべて使う mixed order prior がよい結果を出している。

3.9 Detection and Recognition 1

An efficient RANSAC for 3D object recognition in noisy and occluded scenes

Chavdar Papazov, Darius Burschka (TU Munich, Germany);

物体モデルとシーンを両方とも 3D point cloud で表現し、マッチングにより認識・姿勢推定を行う。3D surface の特徴量に (角度, 角度, 内積) という 3 組を提案し、モデルに対してはそれを DB に保存し、シーンに対してはその DB から検索する。このとき、RANSAC に必要な回数を通常 $N = O(n^2)$ から $N = O(n)$ へ削減、高速化を図る。従来法の spin image などよりもロバスト・高性能。

Change detection for temporal texture in the Fourier domain
Alexia Briassouli, Ioannis Kompatsiaris (Informatics and Telematics Institute, Thessaloniki, Greece);

temporal/video texture のノンパラメトリックなモデルを Fourier 領域で行う。各フレームをフーリエ変換し、特性関数を算出。それを基に、cumulative sum of log-likelihood ratio を計算し、これをしきい値で判定することで変化検出を行う。

スライドは文章だらけ、理解困難 ...

Stream-based active unusual event detection

Chen Change Loy, Tao Xiang, Shaogang Gong (Queen Mary Univ. of London, UK);

まれにしか起らないデータを検出する。従来は outlier detection を基にしていたが、ここでは active learning を利用する。まずシーンを、optical flow/block clustering/motion features (words)/region graph nodes などを用いて分割し、各分割領域で bayesian classifier を学習、それぞれでユーザーに対して query を行い、フィードバックを得る。

Asymmetric totally-corrective boosting for real-time object detection

Peng Wang, Chunhua Shen, Nick Barnes, Hong Zheng, Zhang Ren (Beihang University, Beijing, China);

対象となる object (face など) の学習サンプルは少数しか得られないのに、non-object のサンプルは膨大。この偏った (asymmetric) な状況に対処するために、asymmetric な loss function を用いる boosting を提案。名付けて "AsymmBoost_{TC}"。AdaBoost よりも性能が良いという結果をさまざまな例で得た。

4. 11th Nov. 2010: 2nd Day

4.1 Keynote

The Application of Vision Algorithms to Visual Effects Production

Sebastian Sylwan (Weta Digital, Wellington, New Zealand)



Sebastian Sylwan

2日目の keynote は、New Zealand の VFX 企業 Weta Digital の CTO である Sebastian Sylwan。現在上映されている 3D 映画「アバター」のビジュアルエフェクト（視覚効果 VFX）について、CV・CG と産業応用とのギャップについて、Sylwan はイタリア生まれ、USA を経て NZ の Weta に来て 2 年になる。CV や CG は小さい時から興味を持っていたそうである。

Challenge は、CVCG 技術を産業へ応用する (adopt) こと。VFX はここ 10 年で急増しており、映画において 1 分あたりの VFX ショットの数をグラフにするとその増加ぶりがわかる。VFX を使った映画は興行成績でもトップに位置する（1 位はアバター、2 位はタイタニック）。2007 年から 2008 年の 1 年間だけでも、VFX の割合は 45 % から 48 % に増えている（数字の記憶はあいまいです）。

技術が生まれてから産業で発展するにはどの分野も 30 年かかる。自動車は、1885 年から 30 年かかったし、飛行機は 1902 年から 25 年かかった。VFX も、1976 年からアバターまで約 30 年かかっている。

Weta ではこれまでアドホックな VFX も多かったが、CV では最先端の技術が生まれている。このギャップを何とかして埋めたい。そこで、VFX で問題となっていることを紹介する (share the challenges in VFX)。

Scaling の問題：

アバターの 1 シーンを見ると、森の木の枝や葉が多く、ポリゴン数は膨大になる。あまりに膨大なので、普通に研究発表されている手法はほとんど役に立たないから、自分たちで手法を作り直すことになる (Re-invent techniques by ourselves)。

Workflow：

VFX で使われる画像処理技術は多岐にわたる (optical flow,

mocap, etc)。これらを一連の処理 (workflow) の中にツールとして取り込まなければならない。たとえば、

- capture → 3D
- assets → 3D → 2D
- scan ←→ 3D

キャプチャされたデータやこれまでの蓄積を 3D データ処理に流し込み、scan を繰り返して、2D 画像にレンダリングする。

Capture：

よく見かける green back screen を使う場合にも、さまざまな処理が必要になる。スクリーン上にあるマーカーでカメラ校正を行うが、CV で使われているカメラモデルではあまりにシンプルであるため、映画用カメラには使えない。レンズモデルも複雑になる。ズームやフォーカスなどは CG 合成のためには必要で、これらをすべてモデル化しなければならない。校正に使うマーカーも、一般的な追跡手法や運動推定手法では実用には耐えない。顔にマーカーをつけて表情をモーションキャプチャする場合にも、CV 手法は非常に制御された状況でしか動作しないが、映画撮影では照明が複雑に変化する。

研究を実用化する場合の問題：

VFX で必要となる・不要である要素は、他の産業とは異なることが多い。

- ユーザーはすべて熟練者である
- アルゴリズムの処理結果は、その後さらに加工処理される
- カメラモデルは非常に精密なものが要求される（ピンホールでは役立たず）
- 環境をかなりコントロールしてよい
- 高解像度が必要
- ほかのツールと統合して一連の workflow の中で使用する
- 大規模データの高速処理のために並列化は必須

例：アバターの 3D シーンを撮影するためにステレオカメラで撮影し、シーンの奥行きを求める。一般的なアルゴリズムでは、とても top conference に accept されるような結果は得られないが、それを熟練者が後処理すれば非常にきれいな奥行きデータが得られる。VFX ではそれでよい。

解像度：

映画には高解像度が必要。論文で使われるような CIF は論外で、以下のようなフォーマットを処理する必要がある。

- Full HD 1920x1080
- 35mm 2048x1556
- 4K Red ...
- 5K Red 5120x2560 (さらに 48fps, 16bit/channel=7BG/s!)

論文では「マシンパワーがあれば」「データが十分にあれば」と逃げるができるが、産業ではそうはいかない。

1 フレーム以下のずれ：

ステレオの左右画像で 1/2 フレームずれている、という問題が生じた。あまり見た目に重要ではないように思えても、fix しなければならない。これはずれている物体を optical flow によって補正して解決した。

Control(後処理):

アルゴリズムは結果の良さを追求するが、その結果が容易に扱えるようなものでなくては意味がない。たとえば dense points は当然だめ。メッシュでもだめ。映画の最終版ではかなり細かいポリゴンのレンダリングをしなければならないが、アルゴリズムで粗いメッシュが生成されたとしても、写実に耐える細かいポリゴンをそれ以上追加することはできない。”make results more manipulable.”

Computational power :

3次元復元のアルゴリズムは、計算資源を使いすぎ、大規模データに対しては実用に耐えない。あるエンジニアが、2・3日かかっていた復元アルゴリズムを並列化して1時間に短縮したが、使用しているリソースを調べたら100,000コアも使っていた、という笑話で紹介された(ジョークの解釈が間違っているかも知れません)。Wetaの計算資源はかなりのもので、NZ全体の計算資源のうち80%はWetaであるというデータも示された。ただしいくら並列化しても、I/Oとネットワークにボトルネックが生じてしまう。

User guidance の重要性 :

VFXのためには perfect な結果が求められるため、user interaction は必須である。ある調査では、アルゴリズムが「95%正しく処理される場合」が90%だが、残りの10%は「完全におかしな結果が得られる」場合であるという。これはVFXには致命的な問題である。そのため、できるだけいろいろなパラメータを調整して perfect にしたいが、一般的なアプリはパラメータが全くなく(簡単に使えるように)、商用アプリでもパラメータ数は少なく設定してある(sonyのスマイルシャッターは笑顔度というパラメータが一つだけ)。パラメータは調整できるほうがいいし、その数も多いほうがいい。アルゴリズムの初期化もユーザーが与えればよいし、制御された環境を用意することもいとわない。

Temporal stability :

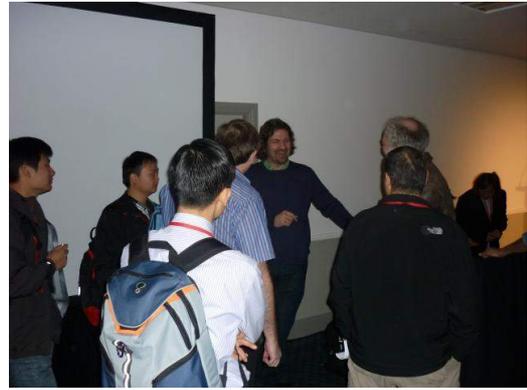
映画は動画なので、1フレームごとに処理するアルゴリズムは適さない。たとえばフレーム毎にマッチングを行うと、当然ちらつきは発生する。

Summary :

パラメータ調整は、今よりももっとやりたい。どんなアルゴリズムも必ず処理の workflow の中で使用するのだから、できるだけある特定の問題を(それだけを)処理する小さなアプリケーションのほうが、かなり使いやすい。ただし、このような問題を解決したとしても CV paper にはなりにくく、VFX publication にしかならないかもしれない。

”Try to make a bridge between academic researches and industry”

発表後、産業界で必要とされる問題を解くアルゴリズムを募集する workshop や contest を企画してはどうか、と聞いてみた。そんなものでは解決しないから、人を雇っているのだ、と一笑に付された。



講演後に質問者に囲まれる Sebastian Sylwan

4.2 Applied Computer Vision

Automatic workflow monitoring in industrial environments
Galina Veres, Helmut Grabner, Luc Van Gool, Lee Middleton
(ETH, Switzerland & University of Southampton, UK);

倉庫や工場の作業をモニタリングしたいが、各作業の時間的長さも異なるし、順番も入れ替わるし、開始・終了もあいまいである。既存の人物検出・追跡は、多数の荷物・機械などで、ほとんど失敗する。そこでここでは、シーン記述子として local motion monitor を提案(計算が簡単でロバスト)。これを Echo State Network (ESN) を用いて識別する。7クラスの task を認識し、NN よりも良い性能を得た(f 値で ESN 78%, NN 60%)。

Context based support vector machines for interconnected image annotation
Hichem Sahbi, Xi Li (CNRS TELECOM ParisTech, France);

ACCC2010 best paper。同じ画像でもコンテキストが異なるとカーネル値は異なるべき、というアイデア。通常の RBF カーネル $\exp(-D/\beta)$ に、コンテキストに依存する factor $\exp(-\alpha/\beta \sum PK^{t-1}P')$ をかけ、カーネル K^t を得る。コンテキスト P は、Flickr のタグを用いる。コンテキストを使う割合は α で制御できる。 α を大きくすると性能は向上するが、大きすぎると反復が終了しない。

Finding human poses in videos using concurrent matching and segmentation
Hao Jiang (Boston College, USA);

物の姿勢推定と、人物領域の segmentation を同時に行う。従来手法は PoseCut(IJCV2008), ObjectCut(CVPR2005) など、他にもあるが、segmentation が姿勢推定の前処理とされている。これを最適化で同時に行う。最適化する目的関数は、体の各部分をマッチングするコスト(部分同士の位置関係、体の対称性などもコストに含める)、foreground ラベルのコスト(各画素が foreground かどうか、foreground 領域の境界の長さも短くなるようにコストを作る)、またこれらが実際に一致しているかどうかのコスト。これを全部合計すると、スライド1枚に10行にわたってコスト関数がびっしり埋まるぐらいになる。これを branch and bound で最適化。

Modeling visual pose disambiguation: Recognizing action at a distance by key poses

Snehasis Mukherjee, Sujoy Kumar Biswas, Dipti Prasad Mukherjee (Indian Statistical Institute, Kolkata, India);

低解像度での人物の action 認識 . Bag-of-words モデルの word には新しく提案する pose descriptor を使用 . keypoint を検出した後 , optical flow を重みづけし , pyramid 的に optical flow histogram を作る (第 1 レベル : 矩形内で 8 方向 bin , 第 2 レベル : 1/4 領域ごとに 8 方向 bin , 以下同様) . ただしすべての pose を codebook に入れてしまうと性能が落ちる . そのため , codebook となる kd-tree には , あいまいな姿勢は捨て , keypose となるものだけを登録する . 周期的な姿勢も扱えるようにするため , 各 action で action graph を作り , そのノードがグループを持つようにする . 各 action のヒストグラムを作るのは従来手法 (PAMI2010) を利用 .

Social interactive human video synthesis

Dumebi Okwechime, Eng-Jon Ong, Andrew Gilbert, Richard Bowden (University of Surrey, UK);

会話での人 (アバター) のやり取り (social interaction) を , 自動的に動画として作成する . texture 合成や video texture , motion synthesis などと同様に , 学習ベースで生成 .

流れ:(入力)何人がが会話している動画 → Multimodal Interactive Motion Controller (MIMC, IEEE T-Multimedia2010) + Human video texture synthesis → social interaction model → Animation

提案する Human video texture synthesis では , 人物の動きの軌跡を交点 (cut point) でいくつかのセグメントに分割し , セグメントをノードとする遷移グラフにして , さらにいくつかのノードをグループ化 (cut point group) して texture motion graph を作成 . social interaction の動画は , このグラフのノードをたどることで生成する . 実験では , 3 人 (スタジオ内 , ブルースクリーン , マイクスタンド前) が 30 分話す動画を撮影し , 解析 . それらしい会話を「話している人」と「聞いている人」の動画を自動的に生成 . 「興味なさそうに」聞いているのか「熱心に」聞いているのか , も生成仕分ける . 質問ではどのような応用があるのかが聞かれ , 群衆 CG 生成 (学校など) などが挙げられた .

4.3 Posters

Monocular template-based reconstruction of smooth and inextensible surfaces

Florent Brunet, Richard Hartley, Adrien Bartoli, Nassir Navab, Remy Malgouyres

reference image を使い , 3D 曲面を復元 . 曲面が S 字に曲がっても (folding) , reference があるので復元可能 . 学生ではなく R. Hartley 自身が poster の前に立っていた .

A hybrid supervised-unsupervised visual vocabulary algorithm for concept recognition

Alexander Binder, Wojciech Wojcikiewicz, Christina Mueller, Motoaki Kawanabe

BoW の codebook 作成のために , 階層的 K-means と同程度の性能で , やや高速に計算する手法 . クラスタリング階層の第 1 段階は K-means と同じ , 第 2 段階は supervised な線形手法 .

Image inpainting based on probabilistic structure estimation
Takashi Shibata, Akihiko Iketani, Shuji Senda

エッジを保存するように , 勾配の確率 (尤度) を定義 , さらに類似パッチ選択も行い , 既存手法と同程度の性能で高速に inpainting を行う .

Class-specific low-dimensional representation of local features for viewpoint invariant object recognition

Bisser Raytchev, Yuta Kikutsugi, Toru Tamaki, Kazufumi Kaneda

同じクラスは近く , 違うクラスは別れるような次元削減手法を MDS で実現 .

Optimal regions for linear model-based 3D face reconstruction
Michael De Smet, Luc Van Gool

3D morphable model で顔の各部分ごとに線形和で表現するとき , 部分を hard に {0,1} で分けずに soft に [0,1] で表現 , より自然な表情の生成が可能 .

Color kernel regression for robust direct upsampling from raw data of general color filter array

Masayuki Tanaka, Masatoshi Okutomi

ベイヤパターンから upsampling を実現 . ベイヤを仮定しないため , スパースに原色の点が与えられても超解像度が可能 .

TILT: Transform invariant low-rank textures

Zhengdong Zhang, Xiao Liang, Arvind Balasubramanian, Yi Ma

画像 (= 行列) の rank は , テクスチャが複雑になるほど , また変換がかかるほど大きくなる . そこで rank を小さくする変換を推定すると , 自動的に画像の補正が可能 .

A convex image segmentation: Extending graph cuts and closed-form matting

Youngjin Park

領域を {0,1} で分ける graph cut と , [0,1] で分ける matting を組み合わせ・統合 . graph cut と matting のエネルギーの和を最小化 . GPU 実装あり .

Identifying surprising events in video using Bayesian topic models

Avishai Hendel, Daphna Weinshall, Shmuel Peleg

Bayesian surprising model を Latent Dirichlet Allocation (LDA) に組み込み , まれなイベントを検出 .

Continuous surface-point distributions for 3D object pose estimation and recognition

Renaud Detry

3D point cloud の物体を表現するために , 3D 座標 + 方向の 5 次元の確率分布を定義 (3D Gauss * von Fisher) . 3 次元シーンとのマッチングには MCMC を利用 .

Time-domain compressive sensing and its reconstruction for high-sensitivity imaging by using combined long/short exposure type single chip image sensor

Sanzo Ugawa, Takeo Azuma, Taro Imagawa, Yusuke Okada

高解像度低フレームレート (G チャンネル) と, 低解像度高フレームレート (RB チャンネル) を組み合わせると, 高解像度高フレームレートを構成 .

Phase registration of a single quasi-periodic signal using self dynamic time warping

Yasushi Makihara, Ngo Trung, Hajime Nagahara, Ryusuke Sagawa, Yasuhiro Mukaigawa, Yasushi Yagi

周期的な歩行者の位相を合わせるために, 複数の多様体を用意, 歩行速度が変わったら多様体を切り替える .

Robust photometric stereo via low-rank matrix completion and recovery

Lun Wu, Arvind Balasubramanian, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, Yi Ma

ランバートを仮定する $N = DL^\dagger$ という方程式において, 鏡面反射・cast/attached shadow を sparse 成分として表現, 分離 .

4.4 Tracking & Categorization

Efficient visual object tracking with online nearest neighbor classifier

Steve Gu, Ying Zheng, Carlo Tomasi (Duke University, USA);

LKT トラッカーの T である Carlo Tomasi が登壇 . 追跡における Challenges (困難な問題) には, occlusion, clutter, scale change, appearance change などさまざまある . そこで, 最近提案されている online classification に基づく手法を拡張し, これらを解決する . 基本的なやり方は, 現在のフレームから追跡対象の appearance model (Bag-of-SIFT) を計算し, 次のフレームの model で NN のものを追跡する . そして対象の model を update する必要があるが, この研究では新しい update scheme を提案 . マッチングには window の 4 次元パラメータ (w, h, x, y) の全探索が必要になるため, branch and bound に基づく Efficient subwindow search (ESS, CVPR2009 by Lampert) を利用 . Limitation は, Motion blur はダメ (SIFT が原因), 矩形 window しかダメ (ESS を使うから) . ただし, いろいろな問題はクリアし, かつ高速 (SIFT 計算を除くと MATLAB で 10fps) .

Robust tracking with discriminative ranking lists

Ming Tang, Xi Peng, Duowen Chen (NLPR, Institute of Automation, China);

Tracking においては, 追跡対象のモデルが drift することが問題, また localisation の精度も問題になる . そこで Discriminative Ranking List (DRL) tracker を提案 . 画像中の 2 つの大きさの異なるパッチを用いて, pair of object models を構築 . 追跡に利用する特徴は RGB+gradient .

スライドが文章ばかりで, プレゼンはそれを朗読するばかりだったので, 理解できず .

Analytical dynamic programming tracker

Seiichi Uchida, Ikko Fujimura, Hiroki Kawano, Yaokai Feng (Kyushu University, Japan);

MIRU で発表済みの解析的 DP トラッカー . 九州大学の内

田先生登壇 . online tracking は効率的だが, local min の問題が常にある . offline tracking は非実時間, 非効率だが, global optimum が得られる . そこでここでは, 解析的 DP を使って効率的にトラッキングする . 通常の DP トラッカーは組み合わせ最適解を求めるため $O(N^4T)$ であるのに対して, 解析的 DP は解析解を計算するために $O(T)$ である . アイデアは, 各フレームでの DP の local cost を 2 次近似する . smoothness constraint も 2 次形式で与えると, DP 更新式全体がすべて 2 次関数になり, 実質的に探索が不要になる (global min が $O(1)$ で求まるため) . さらに複数の 2 次 local cost を利用するとロバストになる . local cost をどうやって 2 次近似するかの説明はスキップされた .

A multi-scale learning framework for visual categorization

Shao-Chuan Wang, Yu-Chiang Wang (Academia Sinica, Taiwan);

Bag-of-words の spacial pyramid model (SPM) は有名だが, 各 pyramid レベルを同等に扱っている . そこで, クラス情報を基に, レベル毎の重みを学習する . 方法は, 各 pyramid レベルで特徴量を計算し, MKL にかける . codebook 作成には sparse coding も利用 . Caltech-101 で 61.43% の性能 .

発表後, レベルだけでなく同レベル内の位置によっても重みを変えたら, と質問したら, shift invariant でなくなる, との回答 .

4.5 Image Sensing

Fast spectral reflectance recovery using DLP projector

Shuai Han, Imari Sato, Takahiro Okabe, Yoichi Sato (University of Tokyo, Japan);

MIRU2010 poster と同内容 . 物体表面の分光反射特性を計測するには, 固定照明でカメラにフィルタを取り付ける方法があり, 高精度だが時間解像度が低い (高速物体には不向き) . 高時間解像度を得るために照明を変化させる方法もあるが, 特殊な照明デバイスが必要 . そこでここでは, DLP プロジェクタを使う (RGB フィルタがプロジェクタ内部のランプ前で高速回転している) . これを物体に照射して, 高速度カメラで撮影すればよいが, RGB のフィルタが切り替わるところ (color switch) で色のブレンドが生じてしまう . そこで, RGB の 3 つの照明であらかじめ記録しておき, color switch の部分では, その 3 つの線形和が生じていると仮定する (これの計測には diffuse white plate を使う) . これを解くには, 9 個の観測でよい (9 measurement = 3 color filters * 3 illuminations) . 応用例では, スペクトルの relighting を動画像に対して行っている .

ACCV2010 Best Student Paper に選ばれた .

Hemispherical confocal imaging using turtleback reflector

Yasuhiro Mukaigawa, Seiichi Tagawa, Jaewon Kim, Ramesh Raskar, Yasuyuki Matsushita, Yasushi Yagi (Osaka University, Japan);

MIRU2010 oral と同内容 . カメラの aperture を半球にすると, カメラの視野角 (FOV) が 180 度になり, Depth of Focus (DOF) が 0, つまりある depth の物体だけをクリアに撮影す

ることができ、occlusionがあってもその向こうの物体を見ることが出来る。問題は、仮想的な半球状の aperture を作るためにはカメラが多数必要になることであるが、それをここでは多数のミラー+カメラ1台で実現する。半球を近似する geodesic dome の面上に仮想的なカメラを配置するために、turtleback (亀甲) reflector という多数の鏡を組み合わせた小球を作成。3D プリンタで外枠を作り、鏡を加工して貼り付ける作業は Ph.D の学生が2日ばかりで作成。製作費は US\$50 程度。ただしこれを利用してはまだクリアな映像が得られないため、特殊な照明を利用。プロジェクタから投影した高周波パターンが、ある Depth だけに高周波成分が照射されるようにすればよいが、そのためにはやはり turtleback reflector を利用。現在の欠点は、仮想カメラも仮想プロジェクタも解像度が小さいこと。

4.6 Banquet

Banquet (conference dinner) は、Queenstown の観光名所である Skyline Gondola に乗って Bob's Hill 頂上にあるレストランで行われた。ACCV の会場からゴンドラの駅まで徒歩で10分かかるので、「Walking Bus に参加する人は 18:00 に集合」というアナウンスが General Chair の Brendan McCane からあった。どういう意味かを尋ねたら、「みんなで歩いて行くんだよ」とのこと。



Skyline Gondola の麓側の駅

Banquet は着席式、Entree, Main, dessert で、その間に以下のアナウンスが行われた。司会はやっぱり General Chair の Brendan McCane 。



4.6.1 Statistics

投稿数・参加者数などが紹介された。投稿数 739, oral 35, poster 180, acceptance rate 29.1%。投稿数の top は中国, 参加者数のトップは日本。参加者の 55%がアジアから。

表 1 submission statistics (countries)

Country	# of papers
China	183
Japan	116
US	56
Korea	48
Australia	38
Germany	33
France	33
UK	26
Singapore	25
Canada	21
Taiwan	17
Hong Kong	14
India	13
New Zealand	12
Iran?	10
Israel	8
Austria	8
Sweden	7
Italy	7
Greece	6
Spain	6
Belgium	5
Czech	5
Hungary	5
Others	37
Total	739

表 2 submission statistics (subject area)

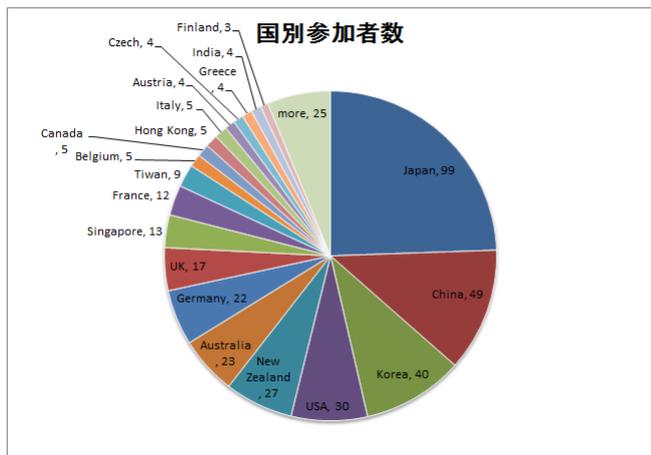
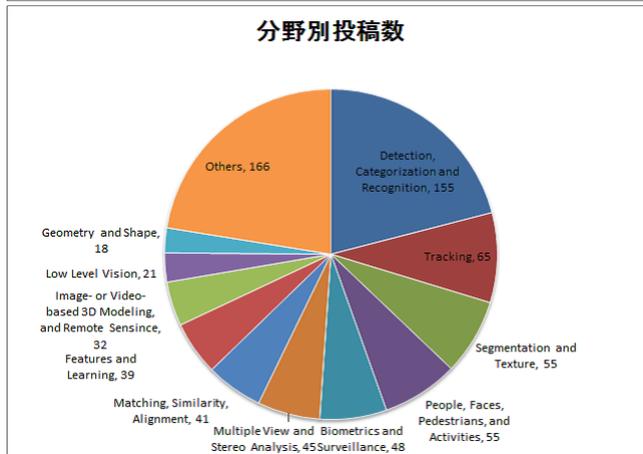
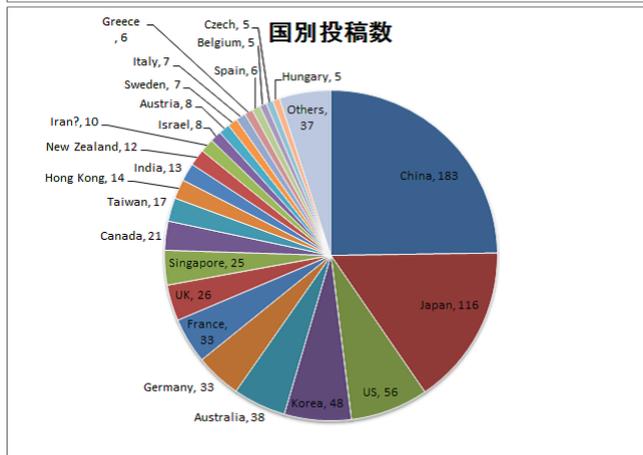
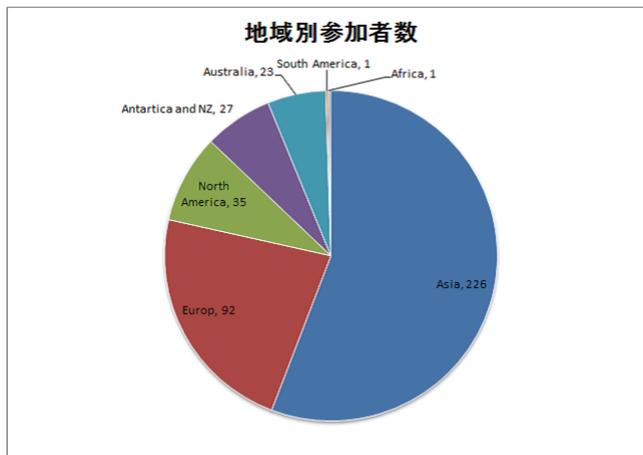
Area	# of papers
Detection, Categorization and Recognition	155
Tracking	65
Segmentation and Texture	55
People, Faces, Pedestrians, and Activities	55
Biometrics and Surveillance	48 ^(注1)
Multiple View and Stereo Analysis	45
Matching, Similarity, Alignment	41
Features and Learning	39
Image- or Video-based	
3D Modeling, and Remote Sensing	32
Low Level Vision	21
Geometry and Shape	18
Others	166

表 3 Registration statistics (countries)

Country	# of people
Japan	99
China	49
Korea	40
USA	30
New Zealand	27
Australia	23
Germany	22
UK	17
Singapore	13
France	12
Tiwan	9
Belgium	5
Canada	5
Hong Kong	5
Italy	5
Austria	4
Czech	4
Greece	4
India	4
Finland	3
more?	25
Total	405

表 4 Registration statistics (areas)

Area	# of people
Asia	226
Europ	92
North America	35
Antartica and NZ	27
Australia	23
South America	1
Africa	1
Total	405



4.6.2 Awards

ACCV2010 award が Brendan から紹介され, その場で Honorary Chair の Richard Hartly から賞状が手渡された. が, 会場に来ていた受賞者が少なかったのが残念である.



Award を説明する Richard Hartly

- Honorable mention

Image-based 3D modeling via Cheeger sets

Eno Toeppe, Martin Oswald, Daniel Cremers, Carsten Rother (会場不在)

- best application paper

Network connectivity via inference over curvature-regularizing line graphs

Maxwell Collins, Vikas Singh, Andrew Alexander (会場不在)

- best student paper

Fast spectral reflectance recovery using DLP projector

Shuai Han, Imari Sato, Takahiro Okabe, Yoichi Sato

- outstanding reviewer award

several reviewers (しかし会場にいたのは Peter Roth だけ) .

- best paper

Context based support vector machines for interconnected image annotation

Hichem Sahbi, Xi Li

4.6.3 AFCV

池内先生から, Asian Federation of Computer Vision (AFCV) の紹介が行われた. AFCV は ACCV 開催の母体 (full support and sponsor) で, 総会は ACCV の Banquet で開催, member は ACCV 参加者, 次回の ACCV 開催場所を承認する権利がある. ただし次回の ACCV に参加しなければ membership はなくなるとのこと (会場爆笑) .



AFCV を紹介する池内先生

4.6.4 Next ACCV

次回の ACCV2012 の proposal が In So Kweon から紹介さ

れた. 韓国についてのトリビアクイズがいくつも出され, 正解すると韓国土産がもらえる. Proposal の内容は以下の通り.

- 場所: Daejeon, Korea
- 時期: 5-9/Nov/2012
- General chairs: In So Kweon, Chilwoo Lee, Akihiro Sugimoto

• Program chairs: Kyoung Mu Lee, Yasu Matsushita, Zhanyi Hu, Jim Rehg

- Venue: Daejeon Conference Center (DCC)

– Conference halls: 1000 seats, 236 seats

– Exhibition hall: 2520m²

– Seminar rooms: 11 (70 seats), 6 (20 seats)

• Conference Theme: Computer Vision for the Smart Services and Systems

– paper quality は保つ

– new emerging application を重視する: Smart-IT, Cultural Tech, Bio-medical App, Digital Vehicle

– young generation のために”Young Leaders Night” を設ける. もし young なら (自薦?), 夕食がもらえる.

- スケジュール

– 11/5: Workshops & Tutorials, Icebreaking Reception

– 11/6: Workshops & Tutorials, Welcome Reception

– 11/7: Main conference, Young Leaders Night

– 11/8: Main conference, Banquet

– 11/9: Main conference, Closing



In So Kweon (右) と司会の Brendan (左)

会場の拍手によって, Proposal は承認された. しかし, 前日のポスターセッションですでに, ACCV2012 のパンフレットとお土産が配布されていた. 数が少なく目立たないところがあったため, あまり知られていなかったようである. なおこのお土産は Banquet 会場を出る参加者に配布された.



ACCV2012 パンフレットとお土産

日程的には、ACCV2012 の次の週に筑波で行われる ICPR2012 (11-15 Nov 2012) と、そのサテライトワークショップ S+SSPR2012 (7-9 Nov 2012 ?) との重複が心配である (個人的に)。日程的にも重なるし、投稿分野も重なる。逆に、欧米からは 2 週間のパカンスツアーを組んで来ることできるというメリットもある。

5. 12th Nov. 2010: 3rd Day

5.1 Keynote

Image-Based and Sketch-Based Modeling of Plants and Trees
Sing Bing Kang (Microsoft Corporation, Redmond, USA)



Sing Bing Kang

ACCV2010 の最終日午前の Keynote は、Sing Bing Kang . マレーシア生まれで、CMU で Ph.D を取得し、現在は Microsoft Research に所属。ACCV2007 でも関連イベントで Talk のあった、image based の tree modeling について。

Why Trees?

なぜ tree をモデリングするのか? それは、住宅地の設計、建物のデザイン、web ゲーム、映画など、さまざまな応用が必要とされるからである。しかし、とても challenging な課題である。形状は複雑で、オクルージョンも多い。そこで、ユーザーインタラクションを利用する。ただし easy to use なもの、出力がリアルでかつ使いやすいもの、を目指す (前日の keynote を意識?)。

従来手法:

- ルールベース手法: Botanical ルール (統計的) L-system, Geometric ルールなどがある。しかし、使用される言語 (ルール) はシンプルでコンパクトなため、何が rendering されるのかはやってみないと予測ができない、という大きな欠点がある。
- インタラクティブな手法: L-system base やスケッチベースなど
- Image based な手法: Volumetric (使いにくい、結果を修正できない), stereo, など。

以下ではまず Image-based 手法について解説し、その後 sketch-based 手法について述べる。

1. Image based

Plants spectrum :

我々がとっているアプローチは、植物の大きさによって手法を変えること。葉の大きさに連続的に tree をとらえた時に、両極端には次の 2 つがある。

- Bushes : leaf が大きく映る (観葉植物など)
- Trees : leaf が小さく映る (大樹)

1.1. Bushes : leaf が大きい場合の手法

流れ:

- 画像の capture
- dense Structure-from-Motion(SfM)
- leaf segmentation graph モデルを用い, Ncut で 3D dense point を segmentation . 2D の leaf 領域はユーザーが指定する半自動方式 .

- leaf reconstruction leaf 部分の 3 次元平面近似
- Branch editing 枝をユーザーが描画

出力結果例: 観葉植物のような葉っぱが大きい植物 (ポインセチアなど) のレンダリング結果が示された .

1.2. Trees : leaf が小さい場合の手法

流れ: bushes の場合とほとんど同じ

- 画像の capture
- dense Structure-from-Motion(SfM)
- occluded branch の合成
- leaf analysis
- Branch editing

ただし, 隠れて見えない枝 (occluded branch) の合成と, 細かい葉の合成 (leaf analysis) が必要となる点が, bushes の場合と異なる .

そのために, まず 3D branch model を構築する . まず point cloud の最も下部に位置する部分をユーザーが指定する . その後は自動で枝の先の部分まで point cloud を追跡し, 枝ごとに分割する . しかしまだ細かい先端までは枝が伸びないので, leaf 上の point に向かって cone 状に枝を派生して leaf と枝をつなぐ . その後, 各 leaf point において小楕円を当てはめ, そこに画像のテクスチャを投影して, それらしい leaf に見せかける .

「vision の人から見れば, ユーザー操作が多すぎるように思うかもしれない」とつぶやいたように, 枝の editing や, leaf point の指定など, ユーザーが指定する操作が多い . しかし, 結果の tree model は, それらしいものが出来上がる .

1.3. Recap (まとめ)

bushes と trees のモデリング手法を別々に紹介した . 結果はそれらしくリアルに見えて OK である . だが, 問題もある . まず, 出来上がる tree は現実に存在する tree のみ (その画像から構築するため) . 次に, 簡単なスケッチだけで, tree を作ってほしい, という要求がある . だから, sketch based 手法も望まれる . これを次に紹介 .

2. Sketch based

ここでは, ユーザーが入力する簡単な sketch を基に, example ベースで tree modeling を行う手法を説明する .

流れ:

- ユーザーの sketch 入力 (枝ぶり, 葉の茂り具合 (crown))
- データベース (DB) から example を選択
- branch reconstruction & propagation
- leaf population
- 3D model の完成

features :

DB に登録する example の特徴は以下のものを使用する .

- 3D geometry
- tree parameters (節目から枝が伸びる方向, 本数などを

gaussian で表現)

- 2D シルエット
- leaf template

Template selection :

sketch で入力された枝と crown を基に, DB から template を選択する . ここでは bottom-up inference と, ICM(?) による refinement を反復する . 当然ながら, 入力する crown を変えると別の tree が出来上がるし, DB の example を変えると同じ sketch でも別の tree が出来上がる .

tree parameter の調整を間違えると, 現実にはありえない tree が出来上がることが多い . この調整は重要である .

3. 比較

- image-based
 - 出力はリアル
 - 限界は, vision アルゴリズムに依存する
- sketch-based
 - 出力は (微妙なパラメータ調整をしないと) おかしなものになりやすい
 - 限界は, prior に依存する
 - しかし easy to use

もっと重要なことは, どちらも scalable ではない! (森林の多数の樹木を合成することは現実的に計算不可能) ということ .

4. summary

unsolved problems

- 画像から, tree parameter を学習すること
- ありえない tree を出さないような parameter の範囲を学習すること

user assistance について

Let's face it! CV アルゴリズムは perfect ではない, とわきまえるべき . ユーザー入力を必要としたとしても, CG ではそれ以上にできえが重要 .

Crowd source を利用しよう

Mechanical Turk のようなものを利用すると, net 上の人々の feedback が数秒以内に得られる . UIST2010 の 2 つの best paper はどちらもそれを利用している .

vision アルゴリズムは, 単体では perfect ではないので, 人間の判断と merge するような方向性を模索する必要がある .

講演後, あなたのアルゴリズムを搭載した商用アプリケーションはあるか? と聞いてみたところ, 知らないとのこと . まだまだ実用化には程遠いらしい .

5.2 Segmentation & Texture

MOMI-cosegmentation: Simultaneous segmentation of multiple objects among multiple images
Wen-Sheng Chu, Chia-Ping Chen, Chu-Song Chen (Academia Sinica, Taiwan);

複数の画像中に共通して存在する物体を抽出する cosegmentation の拡張 . 従来とは異なり, 共通する物体の数は一つではなく, 複数かも知れないし, 0 かも知れない . まず初期化のために, common pattern discovery (CVPR2010) を利

用・局所特徴量 (OpponentSIFT, PAMI2010) の相対的関係のスコア (relational score) を計算する relational score には、2つの descriptor の類似度と、それらが空間的に近いかどうか、が含まれる。これが似ているものをクラスタリングし、confidence Map を作成する。これを基に segmentation を行うため、smoothness/color/confidence/locality の各項目を持つエネルギーを min-cut で最小化する。実験結果例では、多数の starbacks の写真から、共通する starbacks ロゴが抽出されていた。

Spatiotemporal closure

Alex Levinshstein, Cristian Sminchisescu, Sven Dickinson (University of Toronto, Canada & University of Bonn, Germany);

video から空間的に連続した領域を抽出する (ECCV2010) だけでなく、時間的にも連続した領域を抽出する手法。まず video から時空間 superpixel を抽出 (PAMI2009), superpixel flow を計算, superpixel の時空間の類似度 (面積, appearance, 動き) のグラフを構築, それを基に foreground と background の分離を parametric maxflow (ICCV2007) で行う。評価は Stein (ICCV2007) を利用。watershed とどう関係するのかという質問に対しては、とてもよく似ているとの回答。

Compressed sensing for robust texture classification

Li Liu, Paul Fieguth, Gangyao Kuang (University of Waterloo, Canada & National University of Defense Technology, China);

Bag-of-words におけるヒストグラムを compressed sensing で圧縮し, codebook にする。識別器は NN。テクスチャDBでの評価では, Brondatz で 97%程度 (50 次元)。特徴を回転不変にする工夫はいろいろあるが, ここでは local patch を構成するためのサンプリング点 (通常は画素位置のグリッド状) を円状にとり, 勾配計算は放射状が接線方向にとる。

Interactive multi-label segmentation

Jakob Santner, Thomas Pock, Horst Bischof (Graz University of Technology, Austria);

ストロークベースの segmentation は一般的になっていて, office2010 にも grabCut が入っている。しかし, カラー特徴しか扱えない, 2つより多いラベルは扱えない。そこで, マルチラベルで texture 特徴も扱えるようにする。処理速度は GPU を使うと高速 (ここで demo 実演あり)。手法: 使用する特徴は GPU 実装のある既存のもの多数 (color, texture)。モデルは, pixel 毎に各ラベルに所属する尤度。これを Random Forest で学習。Segmentation は, smoothness 制約を入れた MAP 推定を行うために, GraphCut の weighted Total Variation 拡張を利用。さらに, multi-labeling を行うために, convex relaxation を行った近似 Potts regularity を利用。評価に利用したオリジナルデータセット (と binary も) は gpu4vision.org からダウンロードできる。評価は Dice score で行い, 色は Lab, texture は LBP を使った組み合わせが best performance。この手法は Random Forest を使っているのだから, 高速で認識率が高い。一方, 比較した NN は遅いが認識率は高く, linear SVM は早い認識率は引く。

Four color theorem for fast early vision

Radu Timofte, Luc Van Gool (KU Leuven, Belgium);

4色問題 (定理) (FCT) を vision に導入したのは IJCV2002 が初めて。ここでは, FCT を MRF の loopy-BP の計算に利用できることを示す。アイデアは, どのような segmentation 結果も 4色で塗ることができるのだから, loopy-BP のメッセージ伝播にも 4色で塗ることができる制約条件を入れてしまう。そこで, 4色からなる 4次元ベクトルをノードに利用することを提案。名付けて BP-FCT。early vision への応用例として, stereo, segmentation, restoration, optical flow が紹介された。

5.3 Posters

Active learning for solving the incomplete data problem in facial age classification by the furthest nearest neighbor criterion

Jian-Gang Wang

training set に最小限の sample をどう追加するか? という問いに, 最も遠い NN 距離を持つものを入れる, というアイデア。似たような sample ばかりで training set を作っても意味がないため。

From local features to global shape constraints: heterogeneous matching scheme for recognizing objects under serious background clutter

Martin Klinkigt, Koichi Kise

局所的な shape context 同士の, global で loopy な位置関係を利用して, rigid object のマッチングを行う。

3d structure refinement of nonrigid surfaces through efficient image alignment

Yinqiang Zheng, Masatoshi Okutomi, Shigeki Sugimoto

3D 曲面を 3角パッチに区切り, 各パッチでの warp の Jacobian を計算, ICA で高速に registration。

Reconstructing mass-conserved water surfaces using shape from shading and optical flow

David Pickup, Chuan Li, Phil Willis, Peter Hall, Darren Cosker

水面の 3D 復元。shape-from-shading で粗く復元してから, 体積保存制約を入れた optical flow 方程式を解く。

Earth mover's morphing: Topology-free shape morphing using cluster-based EMD flows

Yasushi Makihara, Yasushi Yagi

二つの形状をヒストグラムで表し, EMD 距離を計算する途中段階を形状として見せると, 形状モーフィングになっている。穴なし円盤と穴あきドーナツのモーフィングも可能。

Totally-corrective multi-class boosting

Zhihui Hao, Chunhua Shen, Nick Barnes, Bo Wang

boosting における weak learner の重み更新ベクトルを, 今の時刻 t だけではなく, 過去 $1, \dots, t-1$ の更新とすべて直交するようにとる (これが totally-corrective)。

Image de-fencing revisited

Minwoo Park, Kyle Brocklehurst, Robert Collins, Yanxi Liu

画像からフェンスのような細くて規則性のある格子のようなものを除去。繰り返しという条件だけを与えて, online learning

でなにがフェンスかを自動で学習 .

Feature-assisted dense spatio-temporal reconstruction from binocular sequences

Yihao Zhou, Yan Qiu Chen

時刻 t の左右画像だけでなく, $t+1$ の左右画像 4 枚を使ってステレオ復元 . まず feature-based で粗く復元しておき, それを使って dense に復元 .

Improved spatial pyramid matching for image classification

Mohammad Shahiduzzaman, Dengsheng Zhang, Guojun Lu

画像を 1 回ぼかして SPM を使う . 新規性は ?

Fast recovery of weakly textured surfaces from monocular image sequences

Oliver Ruepp, Darius Burschka

複数の keypoint を追跡して, その軌跡の subspace を求める . 基底となる軌跡を求めるためには変分法を用いる . これができたら, 近似した軌跡の動きから形状復元 .

Ghost-free high dynamic range imaging

Yong Seok Heo, Kyoung Mu Lee, Sang Uk Lee

複数露光 LDR 画像から HDR 画像作成 . ただしカメラや物体が動いても, SIFT やホモグラフィ計算で補正するので, 大丈夫 .

Pedestrian recognition with a learned metric

Mert Dikmen, Emre Akbas, Thomas Huang, Narendra Ahuja

metric learning において, metric が学習サンプルに影響される pull/push の項を新しく提案 .

Affordance mining: Forming perception through action

Liam Ellis, Richard Bowden, Michael Felsberg

perception と action を線形回帰で結びつける .

3D object recognition based on canonical angles between shape subspaces

Yosuke Igarashi, Kazuhiro Fukui

形状部分空間同士の正準角を利用して物体認識 .

Face recognition with decision tree-based local binary patterns

Daniel Maturana, Domingo Mery, Alvaro Soto

LBP の計算を decision tree (DT) で表現 (DT-LBP), Tree を root から leaf まで下る path を一つのヒストグラム bin とみなして学習 .

5.4 Detection & Recognition II

A unified approach to segmentation and categorization of dynamic textures

Avinash Ravichandran, Rene Vidal, Paolo Favaro (Johns Hopkins University, USA);

non-rigid で動く texture object を抽出 (segmentation) し認識するためには, appearance と dynamics をモデル化する必要がある . 難しいのは, scale も color も view point も変化すること . ここでは Dynamic Texture Model (ICCV2001, IJCV2003) を利用 . ただしこのままでは multi-texture を扱えないなどの問題があるため, Bag-of-words と組み合わせる . まず video から Dynamic Appearance Image を複数作成し, それぞれに局所特徴 (dense SIFT) を計算, word を求める . Appearance 以外の

情報は MKL で組み合わせる . しかしまだ single-texture しか扱えないため, local にどの texture に属するかというフラグに smoothness 制約を入れた spatial regularization をグラフでモデル化 . 評価用データセットは存在しないため, 自分たちで作成 . 119 video に, 手動で正解マスク画像も追加 .

Learning video manifold for segmenting crowd events and abnormality detection

Myo Thida, Eng How-lung, Dorothy Monekosso, Paolo Remagnino (Kingston University, London, UK);

人混みの映像を低次元で可視化・モデリングしたい . まず入力映像から local な motion representation を抽出 (パッチ中の optical flow の 8 方向ヒストグラム) . これから, 時空間 laplacian Eigenmap のための距離行列を構築するが, そのために時間 d_t と空間 d_s の距離を両方用いる ($w_{ij} = \exp(-d_t d_s)$) . これで低次元に投影した video が時間方向に描く多様体をクラスタリング . その後, abnormal behavior を検出するには, 多様体までの距離でしきい値判定 . 評価では CVPR2009 よりよい結果が得られた . 質問では当然ながら, 最近流行の Topic model と比較したのかどうか聞かれたが, やっていないとのこと .

A weak structure model for regular pattern recognition applied to facade images

Radim Tylecek, Radim Sara (Czech Technical University, Praha);

厳密ではないがある程度の規則性を持って並んでいるパターン (建物壁面の窓の並びなど) を抽出する . 従来はデザインが制限されたり, 厳密に規則性がなければだめだったが, ここでは weak structure + simple classifier で, 従来と同程度の性能を得る . 確率モデル $p(k, A, X, N)$ を採用, ここで k : complexity, A : element Attribute, X : position, N : structure . シーン中の規則性は, ある $p(A|N, X)$ または $p(N, X)$ の周辺にあると仮定して, その周辺のみを探索する . structure はグラフで表現し, appearance, spatial, structure の規則性を求める (それぞれ $p(A|N, X), p(X|N)p(N), p(N|X)p(X)$ で表現) . 画像モデルはコーナーやエッジの尤度を用いる . 認識のための探索空間は膨大すぎるので, MCMC を用いる . 評価では, プラハでは有名な (?) Dancing House という建物の写真 (窓の配列が列ごとに微妙にずれて踊っているように波打っているため, 規則性はあるものの厳密な格子状にはなっていない) を認識していた .

Multiple viewpoint object categorization and localization

Scott Helmer, David Meger, Marius Muja, Jim Little, David Lowe (University of British Columbia, Canada);

SIFT の作者 David Lowe グループ . 特定物体認識を画像 1 枚でやるのは限界がある, だから multiview で認識し, 3D 位置を localize する . まず各画像において deformable parts model を用いて物体を検出 (false positive が多いが) . 次に multiview の情報を用いて, 物体の category, space, scale, pose を求める . マッチングは, branch and bound を用いて, $\max_h P(F, h|C)$ を最

大にする h を求める . ここで h : matching, F : detector, C : 3D object proposal . 物体かどうかは $P(f|h, C)/p(f|background)$ で評価 . 評価用 multiview データセットは存在しないため, UBC VRS というデータセットを新たに作成 (まだ公開しない?) . single view では false positive が多いが, multiview にするとそれが減る (ただし miss detection も少し増える) .

余談: Lowe の発音は ロー (ローエではない) . ACCV2007@Tokyo のときに会場にいた本人にどう発音するのか聞いてみたら, high and low の low と同じ, と教えてもらった .

5.5 Matching and Similarity

Localized earth mover's distance for robust histogram comparison

Kwang Hee Won, Soon Ki Jung (Kyoungpook National University, Korea);

ヒストグラムを用いた手法では, ヒストグラム同士の距離計算に Earth Mover's Distance (EMD) がよく使われる . しかし計算コストが高いこと, ノイズに弱いことなどが問題 . EMD の変種は多数ある (ECCV2008, PAMI2007, ICCV2009, etc) が, どれもダメ . そこでここでは, EMD をノイズにロバストにするために localized EDM を提案 . EDM での距離計算において, すべての bin を動かすのではなく, しきい値以下のものはそのまま残す . 問題は, しきい値の設定がデータに依存することと, まだ計算が遅いこと .

発表は原稿の朗読でスライドにはテキストが多く, 理解困難 . 結局何が local なのか不明のまま .

Geometry aware local kernels for object recognition

Dimitri Semenovich (UNSW, Australia);

グラフマッチングで使われるアイデアをカーネル設計に取り入れる . local features の集合 A, B 同士のカーネル計算には, 従来は summation kernel や Bag-of-words kernel などがある . BoW kernel に max 計算を用いると性能が上がると報告されている (2009) が, ここではさらに空間情報を導入する . つまり, 集合 A 内の i, i' と, 集合 B 内の j, j' の距離 (?) を利用する difference-of-distance kernel $k((i, i'), (j, j'))$ を提案 . 特殊な場合においては, 従来のもとの等価であることも示された .

5.6 Closing and Award ceremony

最後もやっぱり Brendan McCane が登場, 昨日 award をやってしまったので, もう一回やるわけにもいかない, とのジョーク . "Now, you can go home!"



Closing

6. Conclusions

6.1 雑感

ACCV の内容のレベルが上がっていると言われているのを聞いていたが, 確かにその通りだと思う . 前回参加したのは ACCV2007 であるが, それに比べても見るべきものは多かったように感じる . ただしその理由は主に, top conference に漏れて ACCV に回ってきた欧米の質の高い研究によるところが多い, と個人的には感じる . 特に, 欧米の著名な研究グループの研究発表は CVPR で発表されていてもおかしくない . 相対的にアジアの研究が見劣りするのとは (とくに presentation skill に関しては) 致し方ないところだろう .

一方, 参加者の質は top conference とは言えないだろう . 1 日目の keynote では満席になっていた 400 の椅子も, 最終日の最終セッションでは 1/8 程度が埋まっているだけであった . もっともこれは参加者の母集団の大きさが違うだけなのかもしれない (CVPR2010 の参加者は 1800 人, ACCV2010 の参加者は 400 人) .

また, poster では no show が多かった . poster session のたびに, board に poster が張ってあるかどうかをチェックして回っていた General Chair の Brendan McCane に聞いてみたところ, no show の poster の割合はだいたい 10% とのこと . それらの paper は, website の警告通り Springer online には載らないとのこと . それ以外にも poster だけは張ってあるが誰も一回もその場に来ないというものも多数あった (と思っているが真実は不明) .

6.2 Trend

自分の興味に従って眺めた, ACCV における trend をいくつか紹介する .

- 二値 $\{0, 1\}$ を実数 $[0, 1]$ に relaxation して convex optimization する .
- Latent Dirichlet Allocation (LDA) を使う Topic model で検出 .
 - Bag-of-words 全盛 .
 - branch and bound で最適化を解く .
 - sparseness, L1 norm などはずでに定着した模様 .
 - 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 .

- 最初のスライドは”Challenges or Problem statement” , 最後のスライドは”Recap and Contributions” .
- code や dataset は web で公開 , URL をスライドに掲載 (CVPR/ICCV ではすでに一般的) .

かじめご了承ください .

文 献

[1] Proc. of ACCV2010, 2010.

6.3 Japanese

参加者は日本人が最も多いのにもかかわらず , oral での日本人の質問は少ない (覚えているのは数件) . ICCV2009 でも CVPR2009/2010 でも , 日本人の oral での質問はほとんどなかった .

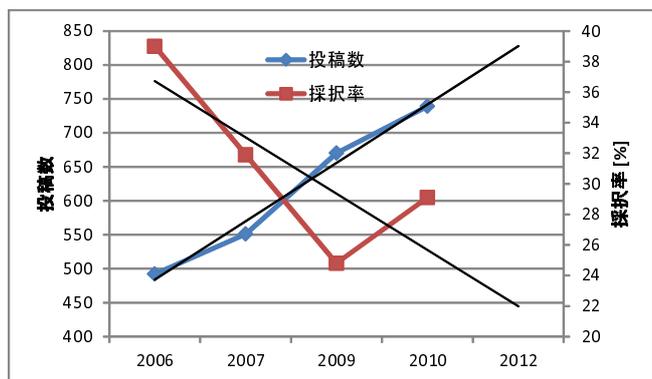
講演が終わった直後の break に , 壇上まで行って講演者に質問する人の列の中にも , 日本人はいなかった . 特に 2 日目 Keynote の Sebastian Sylwan (Weta) は , Keynote が終わった後の 20 分間の break はずっと質問攻めにあい , 次の oral session が始まるまで壇上から離れられないほどであった . その質問者の中には中国か韓国の学生が何人もいたが , 日本人はいなかった .

日本人の発表のほとんどが , oral/poster とともに , すでに MIRU で発表されたものであった .

6.4 History

これまでの ACCV の投稿数と採択率は以下の通り . 2006 年以降のデータを基に次回 2012 年の予測を線形で行うと , 投稿数は順調に 800 を超え , 採択率は 22% 程度になる模様である .

year	location	submissions	orals	posters	rate
2012	Korea				
2010	New Zealand	739	35	180	29.1
2009	China	670	35	131	24.8
2007	Japan	551	46	130	31.9
2006	India	492	64	128	39.0
2004	Korea		61	137	
2002	Australia	189	85	67	80.0
2000	Taiwan				
1998	Hong Kong	over 300	58	112	56
1995	Singapore			450	
1993	Osaka				



投稿数と採択率の予測

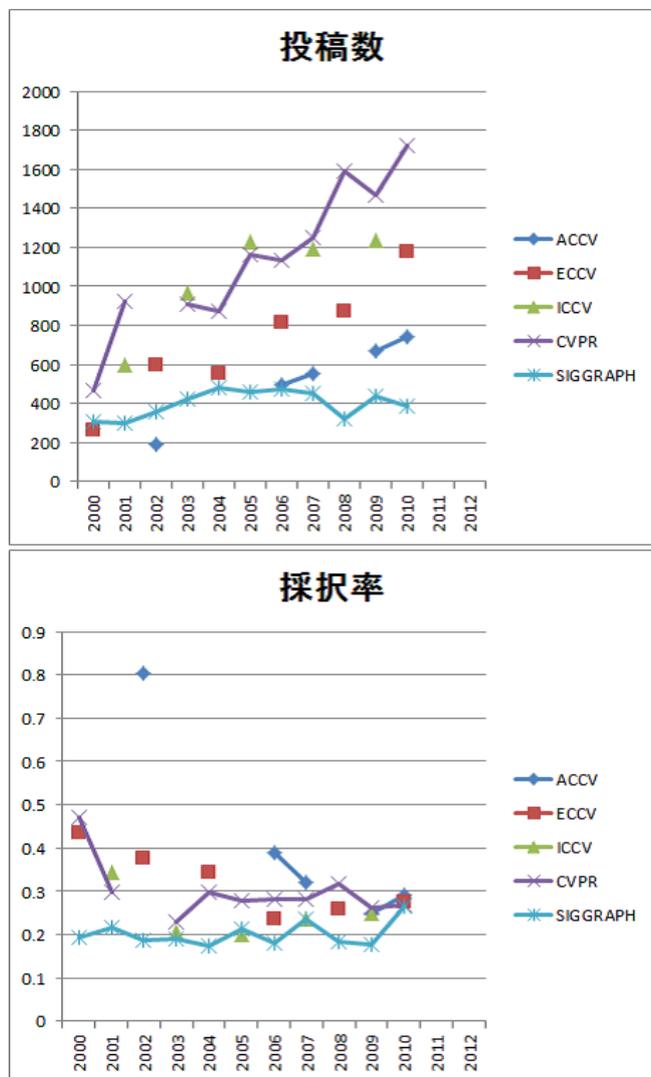
6.5 Acknowledgement

ACCV でお世話になりましたすべての皆様に感謝いたします . この report は personal で unofficial なものであり , 正確を期す努力はしたものの , 正しい情報とは限らないことをあら

付 録

1. 他の会議との比較

Top conference の 2000 年以降の投稿数と採択率をグラフ化してみた。これを見ると、ACCV の採択率が急速に top conference レベルに近づいていることがわかる。意外にも ICCV の投稿数が伸び悩んでいるようである。



ACCV, ECCV, ICCV, CVPR, SIGGRAPH の投稿数と採択率