

ECCV2012 参加報告

向川 康博¹ 奥村 麻由² 中澤 篤志¹ 佐藤 いまり³

概要：2012年10月8日～11日にイタリアのフィレンツェで開催された ECCV2012 の概要を，参加者の視点で報告する．

1. ECCV2012 の概要

European Conference on Computer Vision (ECCV) はコンピュータビジョン分野におけるトップカンファレンスの一つであり，欧州で開催される会議であるが，米国やアジア圏からの発表・参加も多く，国際性の高い会議である．今回の会議は第12回目であり，2012年10月8日～11日の4日間に渡って，イタリアのフィレンツェで開催された．

General Chairs は，Roberto Cipolla (University of Cambridge, UK), Carlo Colombo (University of Florence, Italy), Alberto Del Bimbo (University of Florence, Italy) の3名が務めた．また，Program Coordinator は，Pietro Perona (California Institute of Technology, USA) が，Program Chairs は，Andrew Fitzgibbon (Microsoft Research, Cambridge, UK), Svetlana Lazebnik (University of Illinois at Urbana-Champaign, USA), Yoichi Sato (University of Tokyo, Japan), Cordelia Schmid (INRIA, Grenoble, France) の4名が務めた．

2. 統計情報

ECCV2012 に関する様々な情報が，ホームページに掲載されている．ここでは主な統計情報を紹介する．

2.1 全体の件数

表1に，メイン会議，チュートリアル，ワークショップごとの件数を示す．メイン会議への投稿件数は1437件と相当に増えており，また，参加者も事前登録者だけで1135名であり，規模が相当に大きくなっていることがわかる．これだけの投稿論文の査読のために，50名を超えるエリアチェアと，約1000名の査読委員が動員されている．



図1 会場近くの街並み

表1 全体の件数

Main Conference	Paper submissions	1437
	Oral Program	40
	Poster Program	368
	Demo submissions	25
	Demo Program	21
	Exhibits	11
Tutorial	Tutorial proposals	12
	Tutorial Program	8
Workshop	Workshop proposals	37
	Workshop Program	21
	Submitted papers	450
	Accepted papers	257
Attendees	Registered Attendees (事前登録のみ)	1135

2.2 採択率

表2に，ここ最近の ECCV の論文採択率の推移を示す．投稿数が飛躍的に増加しており，本分野が活発であることが伺える．この投稿数の増加に比例して，ポスターの採択

¹ 大阪大学
² 東芝
³ 国立情報学研究所

表 2 ECCV2008,2010,2012 の採択率の変化

ECCV2008	件数	比率	ECCV2010	件数	比率	ECCV2012	件数	比率
投稿総数	871		投稿総数	1174		投稿総数	1437	
採択数	243	27.9%	採択数	327	27.9%	採択数	408	28.4%
オーラル	40	4.6%	オーラル	39	3.3%	オーラル	40	2.8%
ポスター	203	23.3%	ポスター	288	24.5%	ポスター	368	25.6%

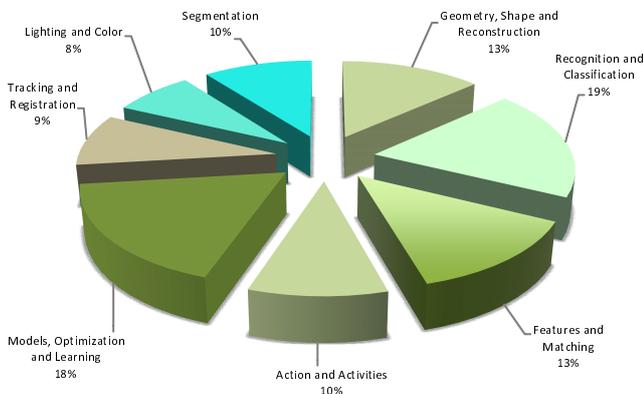


図 2 研究分野の比率 (ECCV2012 ホームページより引用)

件数も増えており、結果として全体の採択率は 28%前後で変化していない。ただし、オーラルは基本的にシングルトラックのため件数を増やすことが難しく、ついに採択率は 3%を下回った。

なお、ワークショップでも、全体としての採択率は約 57%となっており、メイン会議よりも通りやすいものの油断はできない。

2.3 研究分野

図 2 に、研究分野の比率を示す。Recognition and Classification が 19%と最も多く、次いで Models, Optimization and Learning が 18%となっている。比較的、理論的な研究が多いとされる ECCV ではあるが、ポスター会場でも学習や認識に関する理論が目立っていた。

2.4 国別

図 3 に、国別の比率を示す。上から順に、アメリカ (36%)、ドイツ (8%)、イギリス (6%)、中国 (6%) と続く。ただし、アメリカとはいうものの、実際には中国からの留学生の場合も多く、改めて中国の強さを実感した。なお、日本はわずか 1%程度であり、このグラフでは後ろのほうにあり、見つけるのも難しい。

ちなみに、3 位には Microsoft (7%) が目立っており、後ろには Google (1%) も見える。一企業が国家に匹敵するというジョークも聞かれたが、クラウド技術が普及した現在、メールアドレスから国を推定することが難しいことを示しているとも言える。

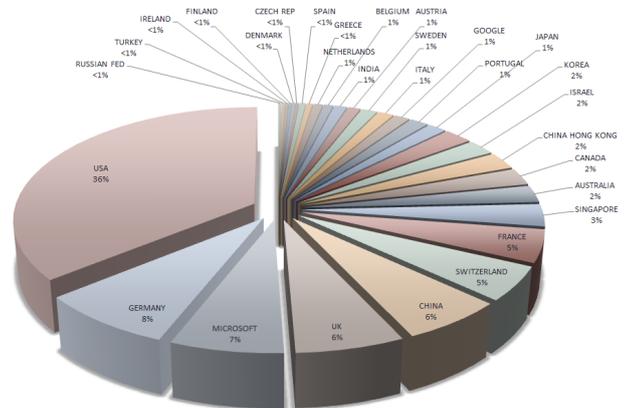


図 3 国別の比率 (ECCV2012 ホームページより引用)

3. 受賞論文

3.1 Best Paper Award

Daniel Kuettel, Matthieu Guillaumin and Vittorio Ferrari, "Segmentation Propagation in ImageNet".

Kuettel らは大規模公開 DB の ImageNet の画像に対して、全自動ですべての画像に対象と背景を分けるセグメンテーションを行う手法を提案した。特徴は全自動であることと、ImageNet の階層型種別ラベルを利用していることで、同じ種別、近接種別の対象特徴は似ていることを利用している。予め用意された画像とそのセグメンテーションの正解を用いて、近接種別の画像に対してセグメンテーションを行う。その結果を近接種別の画像に適用する。さらに、その近接種別の画像に対して... というのを繰り返して、ImageNet 全画像のセグメンテーションを自動で行っている。

セグメンテーションの手法には GrabCut を使い、初期領域の設定に著者が CVPR2012 で発表した画像の一部の類似画像のセグメンテーション結果を用いる手法を用いている。GrabCut の energy function のデータ項に、画像単体での尤度、種別毎の尤度、近接種別での尤度を組み込み、同種別、近接種別の特徴は似ていることを利用して、自動セグメンテーションを実現している。対象の位置情報のない ImageNet の画像に位置と形状の情報を自動で付与することができること、また、階層型の構造をうまく利用できることが評価されたと思われる。

3.2 Best Paper Award - Honorable Mention

Kris Kitani, Brian D. Ziebart, James Bagnell and Martial Hebert, “Activity Forecasting”.

カーネギーメロン大学の Kris Kitani らが発表した Activity Forecasting は、未来の行動軌跡を予測するというもので、これまで画像処理の分野で研究されてきたような、過去または現在の行動がどのような行動かを識別する行動解析とは異なった問題設定となっている。

この論文では、車や花壇といった周囲の物理的な状況が人の行動軌跡に影響を与えるという単純なアイデアを、既存手法の Markov decision process(MDP) モデルに上手く当てはめて解いている。予測対象の位置情報と、クラス分けされたその周囲の物理的な状況を用いて、最適なコスト関数のパラメータを勾配降下法によって学習し、軌跡全体でのコストを加味して次の行動を予測する。

軌跡の予測だけでなく、周囲状況のクラス分けなど、全体的に既存手法を上手く組み合わせられており、実験においては軌跡の予測精度だけでなく、軌跡の滑らかさや目的地の予測精度など、様々な視点で評価を行っている点も評価されたのではと感じた。

3.3 Koenderink Prize for Fundamental Contributions in Computer Vision

Vladimir Kolmogorov and Ramin Zabih, “What Energy Functions Can Be Minimized via Graph Cuts?”

現在の研究に大きく貢献する過去の ECCV 投稿論文に与えられる Koenderink Prize は、コーネル大学の Vladimir Kolmogorov と Ramin Zabih が 2002 年の ECCV で発表した “What Energy Functions Can Be Minimized via Graph Cuts?” に与えられた。一般的なエネルギー関数を、グラフカットを用いて最小化する枠組みを提案しており、ステレオやセグメンテーションなど現在のビジョンや画像処理に盛んに応用されていることから受賞となった。

Vladimir らの発表以前にも、グラフカットはビジョンに導入されており、2 値のノイズ除去などにおいて用いられていたが、いずれも独自のグラフ構成やエネルギー関数を定義していた。Vladimir らの方法では、一般的なエネルギー関数に拡張するため、隣接する頂点間を切断するためのコストに対して、劣モジュラ条件を定義したことにより、一般的なエネルギー関数を大域的に最小化することを可能にしたことが評価されたと思われる。

3.4 Best Student Paper Award

Jianxiong Xiao and Yasutaka Furukawa, “Reconstructing the World’s Museums”.

Xiao と Furukawa らは、“Inverse CSG” という新たなアルゴリズムによる大規模な屋内環境の 3 次元復元手法、および可視化手法を提案した。Inverse CSG は、建物のレー

ザレンジセンサのデータを入力とし、これにフィットする 3DCSG モデルを、エネルギー関数に関する貪欲法により生成する。この手法は、スケラビリティに関しては従来手法を圧倒しており、精度面においても、部屋の角など復元の難しい箇所でも従来手法を上回る性能を示している。いくつかの有名な美術館の実際の復元結果が論文中やプロジェクトページで公開されており、手法の提案だけでなく、実システムとして稼働していることも評価されたと思われる。また、発表はとても上手で感心したが、Furukawa 氏によれば、かなり練習したとのことである。

3.5 Demo Paper Award

David Derisory, Josh Susskind, Lauren Kreiger and Marian Bartlett, “Emotion Mirror: A novel intervention for autism using real-time expression recognition”.

最近、コンピュータビジョン技術を、自閉症の治療に役立てようというという試みを見かけることが多い。自閉症では一般に表情の認識や表出が困難となるが、本研究では、表情のリアルタイム認識によって、画面に表示されるキャラクターが自分の表情を真似したり、あるいは逆に自分がキャラクターの表情を真似したりする Emotion Mirror と名付けられたシステムを開発し、デモを行っていた。表情認識そのものは既存のツールボックスを用いているが、うまく認識や表出ができると、画面中のアイスクリームが豪華になるといったゲーム的な要素や、キャラクターについてもロボットのような無機質なものから動物、人間と、レベルを変えるなどの工夫によって、子供が楽しめるように作りこまれていることが高く評価できる。

4. 注目発表

本節では、我々が聞いて面白いと感じた論文や、多くの聴衆で人だかりができていたポスター発表などをいくつか紹介する。

4.1 重なりのあるシーンでの両手の姿勢推定問題

Luca Ballan, Aparna Taneja, Jurgen Gall, Luc Van Gool, and Marc Pollefeys, “Motion Capture of Hands in Action using Discriminative Salient Points”.

Articulated Object Tracking において、CG で生成された画像と実画像を比較するアプローチはよく取られるが、多くの局所解が存在する事や、自由度が非常に高いため依然困難な問題である。ETH Zurich のグループでは同様のアプローチを用いて両手のポーズ推定を行っており、デモにおいても非常に良く動いていたため、注目論文として紹介したい。本手法は画像からの特徴として、エッジ、オブティカルフローに加え、salient point (指先)を用い、これを最もよく説明する手モデルをレンダリングし、比較する事でポーズ推定を実現する。これは、指先は変形しづらく、

かつ手のポーズに対して特徴的であるためである。この指先を Hough forest classifier を用いて検出し、CG モデルにおいて生成された指先形状と比較する。これに加え、CG において自己干渉および両手指の干渉を評価して避けることで、良好な推定を実現する。本手法が成功しているのは、指先を salient point として見つけ、指の configuration の候補を早めに限定し、その後エッジおよびオプティカルフローを用いて収束させるという組み合わせの妙にあると思われる。

4.2 視線移動と人の意図の関係解析

Alireza Fathi, Yin Li, and James M. Rehg, “Learning to Recognize Daily Actions using Gaze”.

人の行動において視線は重要な役割を果たしている。Georgia Tech. のグループでは、頭部装着カメラと視線検出装置により人の操作タスクを計測し、人の視線移動とタスク間の関係性を示した。彼らは、視線の停留点の変化には行動の種類との相関があると考え、ガウス分布でその確率のモデル化を行っている。これは例えば「ピーナツバターをパンに塗る」動作の場合には、ピーナツバター、ナイフ、パンの間を視線が移動する確率を推定する。頭部装着の画像からは、シーン中のオブジェクトに関する特徴、アピランス特徴、操作予測情報を特徴量とし、実際に計測された視線情報を Ground Truth として SVM の学習を行う。これにより、あるアクションタスクにおいて、視線がどのように移動するかを SVM 分類機で予測することが可能になる。またこの関係を用いて、視線の情報とシーンの情報から、行動の認識を行うことも可能であり、実験では 47% の確率で推定が可能であることが示されている。本研究は、従来では視線情報は、スタティックな画像やビデオ等を対象に行われることが多かったが、実際の操作タスクにおいても有効に用いることが可能であることが示されたという意味で意義深いと考えられる。

4.3 視線移動データベースと Visual Saliency

Stefan Mathe and Cristian Sminchisescu, “Dynamic Eye Movement Datasets and Learnt Saliency Models for Visual Action Recognition”.

視線を取り扱った興味深い研究として、Mathe らの発表を取り上げたい。彼らは映像（静止画・動画）に対する視線移動と Visual Saliency の問題を取り上げ、既存研究の問題点（例えば shooter’s bias）を考慮して実験を行った。その結果、映像の種類により、被験者間の注視点の一致度が異なることが確認された。具体的には、映画等で構成される Hollywood Debase の方が UFS Sports database よりも一致度が高いことが確認されている。これは、映画のようなコンテンツのほうが撮影者の意図が反映されやすいことを示している。また新たな取り組みとして、静止

画のみでなく動画での視線移動問題にも取り組み、複数人物間での視線移動の一致度を評価している。これにより、Action Recognition での新たな Visual Saliency の導出法を提案し、高い効果を得ている。このように近年盛んになっている Visual Saliency と注視点の関係解析問題に対し、動画像に対する解析を導入した点、また動画像に対する視線移動のデータベースが公開されている点は、今後のこの分野のキーとなると思われる。

4.4 視覚的顕著性計算法のベンチマーク

Ali Borji, Dicky N. Sihite, and Laurent Itti, “Salient Object Detection: A Benchmark”.

画像中において注意を引く対象を見つけるための仕組みとして考案されてきた視覚的顕著性 (Saliency) を計算する様々な手法をその長所と短所を紹介しながら、さらにその性能を比較している論文を紹介する。簡単に入手でき、精度も高く、良く知られているモデルについて、様々な視点から評価が行われていて興味深い論文であった。プログラムが公開されているかどうかなど実用的な情報も掲載されており、今後、顕著性を考慮した研究を進めたい方にはとても参考になる論文である。発表にはたくさんの人が聞きに来ていた。

4.5 特殊照明による Depth edges の推定

Yuichi Taguchi, “Rainbow Flash Camera: Depth Edge Extraction Using Complementary Colors”.

この研究では、画像中に観察される Depth edges (depth discontinuities) を Rainbow Flash と呼ばれる 8 個の LED ライトで構築される光源下で撮影された画像から効率良く推定する手法を提案している。従来研究は、空間的に配置された Flash を用いて、それぞれの光源下で観察されるキャストシャドウを用いて Depth edges を推定しノンフォトリアリスティックレンダリング技術への入力として活用した。従来法が異なるフラッシュの方向で撮影された複数枚の画像を必要とするのに対し、この手法では、円状に配置された異なる色の LED ライトを用いることで、観察される影の色合いに基づき Depth edges とその orientation を推定できることを示した。色を使うアイデアがとても興味深かった。この研究の発表者は同じ時間帯にもう一つポスター発表があり、異なるトピックのポスターをふたつ並べて臨機応変に説明しており、とても大変そうだった。

5. 会議の様子

5.1 オーラル発表時間

オーラル発表の持ち時間は 15 分（発表 12 分 + 質疑 3 分）と、一般的な発表時間と比べて短かった。そのため、予稿集の内容に比べて発表内容は絞りこまれており、短時間でも伝わるように工夫されていた。発表者にとっては短かつ

たかもしれないが、聴衆者にとってはテンポよく聞くことができ、集中力を持続できたのではないと思われる。

5.2 ポスター発表でのタブレット端末の利用

iPad 等のタブレット端末の普及に伴い、ポスターに端末を貼りつけて動画を見せるなどの工夫をする人が多いかと思いきや、そのような発表をしている人は意外に少なく、数名程度だけであった。コンテンツを見せる必要のあるマルチメディア系とは異なり、どちらかといえば理論的な研究が多い ECCV ではポスターで十分なかもしれない。

5.3 不運なデモ会場

デモセッションはポスターセッションと時間が完全に重なっており、かつ両セッションは離れた建物で実施されたため、デモ会場の訪問者はほぼ皆無という時間帯があった。実際、まずポスターを見た後、オーラル・デモの建物に移動して、少しデモを見た後にオーラルを聴くという移動パターンが多かったように見受けられる。二日目以降はデモに関して頻繁にアナウンスがあったため、少し状況は改善されたものの、もう少し参加者の移動パターンを考慮した配置のほうがよかったと思われる。

5.4 twitter

オーラル会場には発表用のメインスクリーン以外にサブスクリーンがあり、発表中に twitter で質問を呟くと、サブスクリーンに表示され、質問時間が余った時には、座長が表示された質問の中から選んで質問してくれるという試みがあった。不運にも会場のネットワークが繋がりにくかったためか、あまり積極的に活用はされなかったが、小さな疑問も気軽に質問できたり、発表者が後から聴講者の疑問点を確認することができるという点では、このような取り組みは面白く、積極的に行われるべきだと感じた。ただし、そのための準備は万全にしておく必要があるだろう。



図 4 twitter 用のサブスクリーン

5.5 タイムテーブル

午前と午後の最初はポスターとデモセッションからの開始だったため、あまり時間にとられることなく、昼食も比較的自由に時間を使って過ごすことができた。そのため、各ポスターセッションが始まってすぐは発表者しかおらず、やや寂しい感じを受けた。

5.6 バンケット

カンファレンス会場から少し歩いたところにある旧駅舎がイベント会場として利用されており、元ターミナル駅の巨大な空間がバンケット会場として使われた。料理はケータリングサービスと思われ、廃墟じみた空間がおしゃれなレストランになっていた。

かなりの数のスタッフが働いていたが、即席の厨房で 1000 人を超える客にコース料理を提供するのは大変だったようで、料理と料理の間の時間が非常に長く、ディナーは 8 時前から始まったが、最後のデザートが出てきた頃には深夜 0 時を過ぎていた。やはり、1000 人規模になると、バンケットの準備が急激に難しくなり、開催者の悩みの種であることは間違いなさそうである。



図 5 旧駅舎で行われたバンケット

6. まとめ

本稿では、予稿集を読んだだけではわからない会議の様子を、実際に参加した者の主観的な視点で報告することに努めた。もちろん、ここで紹介した論文以外にも多数の興味深い発表があり、すべてを紹介できたわけではない。幸いにも、招待講演やオーラル発表についてはビデオが公開されているため、ぜひご覧になることをお勧めする。

次回の第 13 回 ECCV は、2014 年にスイスのチューリッヒで開催予定である。日本からも多くの投稿論文と参加者が集まることを願って、本報告のまとめとする。