

**京都大学教育研究振興財団助成事業
成 果 報 告 書**

平成20年8月25日

財団法人京都大学教育研究振興財団
会 長 辻 井 昭 雄 様

所属部局・研究科 情報学研究科

職 名・学 年 博士後期課程3回

氏 名 武 笠 知 幸

事業区分	平成20年度・国際研究集会派遣助成		
研究集会名	2008年国際コンピュータビジョン夏の学校		
発表題目	Acquisition of Kinematic Structure and Description of Complex Human Motion		
開催場所	イタリア共和国 シチリア州 サンピエリ		
渡航期間	平成20年7月13日 ~ 平成20年7月22日		
成果の概要	タイトルは「成果の概要/報告者名」として、A4版2000字程度・和文で作成し、添付して下さい。「成果の概要」以外に添付する資料 無 有()		
会計報告	交付を受けた助成金額	200,000 円	
	使用した助成金額	200,000 円	
	返納すべき助成金額	0 円	
	助成金の使途内訳 (使用旅費の内容)	研究集会参加費	125,730 円
		宿泊料	70,410 円
		バス賃	670 円
日 当		3,190 円	

成果の概要 / 情報学研究科知能情報学専攻 武笠知幸

1 国際コンピュータビジョン夏の学校 2008 (以下, ICVSS2008) 概要

コンピュータビジョン分野における, トップレベルの研究者による自身の現在最先端の研究成果のレクチャーや, 古典的内容についての非常に示唆に富んだレクチャーを受けることができた。(後述) 通常国際会議と異なり, 学生向けであるため基礎的な内容にまで遡った解説であり, 大変勉強になった。

また, 当該分野における著名なアルゴリズム(下記)に関する学生同士による討議にも参加した。

- ・ SIFT
- ・ Mean-Shift

自身のみでは到達できなかったであろう, 多角的な深い視座が得られたにとどまらず, 現時点での自身の理解度を世界基準において計ることができたことは興味深い。

2 レクチャー概要

Seeing what you want to see: Bayes and priors in vision problems

Andrew Fitzgibbon

Microsoft Research Ltd, Cambridge, UK

ステレオ視に用いる画像の組を選択するといった, コンピュータビジョンにおける多くの問題では, 可能な選択肢が非常に多くなり一意に解を求めることが難しい。

これらは, 予期された観測結果から事前確率を求めるベイズ推定問題として考えることが出来る。

しかしこうして得られる解答は信頼に足るものであろうか。偏りが無いと言えるだろうか。

こうした疑問に対し, 講演者はコンピュータビジョン分野における主な事前確率を分類・紹介し, さらにこれらに基づく推定手法について述べた。

特に最大事後確率推定 (MAP: maximum a-posteriori) に力点を置き, 仮想視点の生成やステレオ形状復元といった問題における有効性を述べた

The Scene in Front of You

Jan Koenderink

Utrecht University, Netherlands

人間の視覚に基づく外界認知がいかに行われているかという問題は, 何世紀にも渡って探求されてきた, 非常に重要な未解決問題である。この問題に対する現在主要なアプローチは, デイビッド・マーによって提唱された「視覚は逆光学である」という立場である。

コンピュータビジョン分野もこの観点に基づいて大きな発展を遂げてきたが, この立場の基礎の脆弱性もまた明らかである。逆光学は多くの曖昧性をもった解をもたらすし, 現実的でない, ときには物理的に不可能でさえある, 光学的に検証できない仮定に基づいているのが通例である。さらに最も重大な問題点は, 原理的に意味というものを計算できないことである。人が外界をある意図を持って観測すると

き、得られる情報は観測者の心に起ち現れる意味である。人の視覚はほぼ確実に逆光学では無く、むしろ制御された幻覚に近い。従って、意図というものは外界認知の結果ではなく起点である。コンピュータビジョンが人の視覚を再現することであるならば、現在主流であるマー的な視点は新たなパラダイムによって置きかえられなければならない。以上の観点から、講演者は陰影からの形状復元といったよく知られた例をひきながら、コンピュータビジョンシステムの構造について議論を展開した。また、コンピュータビジョンシステムの構築に当たり興味深いと思われる人の視覚特性についても論じた。

Object representation and tracking

Andrea Cavallaro

Queen Mary, University of London, UK

外界の意味的記述、イベント記述、ビデオデータからの情報抽出には、対象物体の種類と位置の情報が必要となる。しかしながら、構成要素に基づくビデオへの注釈付加は、用いる対象検出・追跡手法に依存する。対象検出・追跡の精度は逆に注目対象の種類をいかに絞り込むかによる。講演者は対象追跡に適した様々な対象物体の表現法について分析し、対象追跡におけるスケール、回転に対する頑健さと、記述の容易さとの間の発生するトレードオフについて論じた。更に、対象追跡過程への対象検出の統合についても言及した。また、対象記述の高精度化のために如何に時間的用法を用いるか、対象追跡の高精度化に如何に対象検出と時間的フィルタリングが寄与するかについても述べた。

Local Image Descriptors: Techniques and Applications

Matthew Brown

University of British Columbia, Canada

画像中の特徴点のマッチング手法とその応用についての講演であった。

講演者は、まずハリス特徴点や SIFT といった、広く用いられている特徴点抽出・記述手法について述べた。続いて、訓練データにより画像記述子を学習させる手法について議論した。これらは LDA や AdaBoost といった教師付き学習アルゴリズムを用いるものである。古典的手法に比して、識別学習によって如何にして、より少ない次元とパラメータ補正によった最適な記述子を求めることが出来るのか解説し、これらの記述子がマッチングおよび物体認識に有効であることを述べ、最後に応用例としてパノラマ生成、3次元形状復元、画像検索手法を紹介した。

Multi-view stereo for 3D reconstruction

Carlos Hernández

Toshiba Research Europe Ltd,

Cambridge, UK

画像から3次元形状を復元するという問題は、過去20年以上にわたって多くの研究者が取り組んだきたコンピュータビジョン分野における古典的問題である。近年ようやくこうした手法は成熟の時期を迎え、非常に正確な形状復元結果を得られるようになってきた。講演者は、現時点で最高レベルの精度を誇る、ステレオ対応に基づく画像からの3次元形状復元アルゴリズムを紹介した。このアルゴリズムの形状復元過程は4つの段階に分けられる。カメラ校正、画像領域分割、画像からのフォトコンシステンシ推定、フォトコンシステンシに基づく表面形状抽出の4つである。講演者はこの中でも特に、後半の2つの段階に力点を置いていた。すなわち、どのようにして画像集合から幾何学的情報を抽出するか、

そして如何に最適に異なる幾何学的推定結果を統合するか、という点である。

Detection and Tracking and Human Machine Interaction

Claudio Domenico Marchisio

Advanced System Technology

STMicroelectronics

この講演では、ヒューマンマシンインタラクション分野およびコンピュータビジョン分野における ST マイクロエレクトロニクス社の取り組みが紹介された。その中でも特に物体検出とその追跡に焦点が置かれていた。

市場動向の分析から製品の広告戦略にいたるまでの産業界からの視点から、ヒューマンマシンインタラクションアプリケーション開発の概要について述べた。

例としては、デジタルカメラにおけるオートフォーカス等の撮影自動化技術、複合現実感技術、ヒューマンマシンインタラクションシステム、自動車におけるセキュリティ向上、自動監視システム、認証システムが挙っていた。

また、対象物体検出・追跡のデモもあった。

3D Representation for Object and Scene Categorization

Silvio Savarese

University of Illinois at

Urbana-Champaign, USA

対象物体およびシーンの分類は近年コンピュータビジョン分野における主要な研究課題となってきた。これらの課題は非常に難しい問題を秘めている。単体の物体もシーンも、光学的あるいは幾何学的条件によって見た目、構造ともに多彩に変容する。さらに、同様のクラスに属する物体でも、クラス内の変数により様々に異なりを見せる。

近年、対象物体やシーン固有の多視点で観測される 3 次元の特性を効率的に記述する最適な表現法を求め、その表現法を物体認識・分類に活かす、という目的に対し、新たなモデルや機械学習手法が提案されている。講演では、いくつかの異なるアプローチを比較分析がなされた。

Randomized Decision Forests for Segmentation and Recognition

Jamie Shotton

Toshiba, Multimedia Laboratory

Corporate R&D Center

Kawasaki, Japan

この講演では randomized decision forests について考察を行い、そのアプリケーションとしての物体認識および画像領域分割について詳述した。

Radomized forests は一種の決定木の集合である。その決定木はランダムに選ばれたデータの部分集合と特徴群によって学習済みのものである。決定木は識別器を直列したものを一般化したものと言え、高速に動作するという特徴がある。講演で、decision forests が画像中の特徴領域の分類とクラスタリングに如何に用いることが出来るか、そしてそれにより高精度に画像の領域分割と物体認識が同時に実現できる、ということが解説された。

Trading off Invariance and Discriminative

Power in Low-Level representations

Stefano Soatto

UCLA, USA

局所特徴は、多くの認識システムにおいて、認識対象が特定の物体かあるいはある物体のカテゴリに関わらず必要不可欠なものとなった。この講演では、なぜ局所特徴が有用なのか、なぜ画像を微小な領域に分割して用いる必要があるのか、そして、これらの微笑領域を記述するにあたってどのような表現形式が必要となるのか、といった観点に沿って解説が行われた。

ある物体の画像は、その物体自身の形状や反射特性に基づいて決まってくる。しかし、隠蔽や視点、照明といった外的要因の影響も存在する。局所特徴はこれらの外的要因による変動をおさえると同時に、可能な限りの物体自身の情報を獲得するように設計される。

講演では、どのように特徴パラメータを設計するか、そしてそのパラメータを設定するか、局所特徴の選択が識別器の設計においてどのような役割を果たすのか、その頑健さと限界について述べられた。

Computer Vision for Gesture-Based Interaction

Björn Stenger

Toshiba Research Europe Ltd,

Cambridge, UK

ディスプレイと対面しジェスチャーによりインタラクションを行う単眼カメラシステムについての解説。プロトタイプでは複数のコンピュータビジョンシステム（顔検出・認識システムはコンテンツの個人適応に、手追跡システムはディスプレイ上のポイントの操作に用いられた）を統合した。提案システムでは、ユーザがジェスチャーにより画像やビデオ、および3次元コンテンツを操作できる。

Human detection and motion capture

Bill Triggs

Centre National de Recherche Scientifique (CNRS),

Laboratoire Jean Kuntzmann (LJK), Grenoble

画像中や、ビデオ映像中から人物を検知し、さらにその姿勢や動作を推定する手法について、その研究の最前線について概観、解説が加えられた。

Photometric methods for 3D reconstruction

George Vogiatzis

Toshiba Research Europe Ltd,

Cambridge, UK

本講演では陰影に基づく3次元形状復元手法について解説がなされた。これらの手法はある程度の光環境の制御が必要となる。したがって、対応点に基づく形状復元手法に比べ、能動的なものであると言える。その利点は、受動的な手法では困難な、高精度の形状復元が可能となるということである。講演者はまず、陰影からの形状復元手法や、古典的なフォトメトリックステレオを含む、当該分野の概要について述べた後、照明の調整やセルフシャドウへの対応といった詳細について解説を加えた。さらにこれらの手法に基づく2つの有用なアプリケーションについても紹介を行った。一つはテクスチャを持たな

い物体の詳細な3次元形状を復元できる多視点フォトメトリックステレオシステム、もう一つは顔や衣服といった自由変形をする物体の形状復元を行うカラーフォトメトリックシステムである。

Object recognition and localization

Andrew Zisserman

University of Oxford, UK

近年、視覚情報に基づいて対象物体の分類（例えば、車や、牛、バイクや人）を行う研究が盛んに行われ、多くの成果が得られている。物体認識では、画像中に含まれる対象物が何であるかにより画像を分類するか、あるいはさらにその対象物の画像中での位置を特定する。

近年の成果は新しい視覚的表現法と強力な学習アルゴリズムの組み合わせによりもたらされた。

視覚的表現法のひとつは、対象物をシンプルに局所的な画像パッチの集合としてモデル化するものである。各パッチは対象物の各部を記述する「視覚的単語」としてとらえることができる。やや驚くべきことだが、各対象物はパッチの空間的な配置、構成を用いることなく分類することができる。これらのモデルは、文章理解のための統計的モデルである、“bag of words”からの類推によって理解できる。しかしながら、物体の位置を特定するためには、画像中の各構成要素の配置を表現する必要がある。

本講演では、複数の表現手法を紹介した。bag of words, 空間ピラミッド、ピクトリアル構造である。さらに、それらの表現法から構成される記述子や検知器の学習手法やアプリケーションについても論じた。また、特定のクラスに特化した画像領域分割や簡単なあるいは誤りを含む教師信号下での分類学習手法についても解説した。

3D reconstruction from images

Luc Van Gool

ETH Zurich, Switzerland

3次元形状復元に関する基本から応用までを幅広く概説した。

まず、受動的3次元形状復元手法と能動的3次元形状復元の比較紹介を行い、それぞれの手法をさらに使用するカメラ台数により分類し、それぞれの特徴を述べた。

特に、自己校正を行う受動的ステレオ形状復元の詳細と、能動的形状復元手法における構造的照明（structured light）の実際について詳述した。また、レーザレンジファインダ、シルエットベースの形状復元手法、フォトメトリックステレオ、テクスチャからの形状復元手法についても解説を加えた。さらにエピポーラ幾何等を詳述し、単一の画像からの形状復元から複数画像からの形状復元へと順に復元に必要な諸条件を数学的に整理した。

最後に、アプリケーションとして、ARC3D(automatic reconstruction conduit 3D)というユーザが写真をアップロードすることにより形状復元結果が得られるウェブサービスや、全方位カメラを車載することにより都市景観の3次元形状復元をするシステム、さらにこれらの技術により得られる幾何学的・地理的情報や Wikipedia 等の文字情報を統合したサービス(Monumental mining)への取り組み、展望を示した。

他にも、3次元形状復元技術のCGとの統合による、映像制作分野での数多くの応用事例の紹介もあり、非常に興味深い講演であった。

3 成果

事前に登録したポスター発表 (Tomoyuki Mukasa: Acquisition of Kinematic Structure and Description of Complex Human Motion) に加え、思いがけず口頭発表に選ばれるという幸運に恵まれた。

口頭発表では、カリフォルニア州立大学ロサンゼルス校の Stefano Soatto 先生から、対象の大局的構造記述手法に関する質問を受けた。折しもちょうど Soatto 先生の運動と変形の同時記述に関する論文 (Stefano Soatto, Anthony J. Yezzi: DEFORMATION: Deforming Motion, Shape Average and the Joint Registration and Segmentation of Images, International Journal of Computer Vision, 2003) を読み始めたところであり、今後論文内容等について質問させていただきたいとの旨もお伝えすることができた。(現在、再度運動学的構造獲得に関する研究の深化を目指しており、特に対象表面の微細な運動と大局的な運動学的構造に基づく運動の分離、あるいは同時記述が課題となっている。したがって、"Deformotion" というコンセプトは非常に参考になると予想される。)

また、口頭発表自体が英語によるプレゼンテーションの実地練習となり、よい経験であった。やはり英語力が不足しており、特に質疑応答に十分対応できないことを痛感した。今後の課題である。

およそ1週間にわたる半共同生活を通じて、多くの同年代の学生と交流が持てた。

互いの研究に関する討議はもちろんのこと、同様の境遇にいるもの同士として、今後の展望などについてまで話すことができ、今後につながる関係が築けた。

これは今後の研究生活を考えると、非常に大きな資産になると考えられる。

また今回、レクチャー等のスライド資料および録音も得られた。

聴解力の問題により十分に理解できなかった部分の自学自習に活かせるのは勿論のこと、研究室内で共有・活用できる貴重な資産である。来年度以降の下回生の指導等に活用する予定である。