ACM Multimedia（ACMMM）2014参加報告

（学生会員）加賀谷北斗†，（正会員）山崎俊彦††

2014年11月3日から7日まで，米国・フロリダ州・オーランドで開催されたThe 22nd ACM International Conference on Multimedia（ACM Multimedia 2014）の様子を報告する．ACM Multimediaはマルチメディア分野では最高峰の国際会議の一つであり，最先端技術の研究発表はもちろん，さまざまなマルチメディアに関するイベントがあることも特徴である．

キーワード：ACM Multimedia，マルチメディア，コンピュータビジョン，画像処理，アート

1．まえがき

ACM Multimediaは，Association of Computing Machinery（ACM）の分科会Special Interest Group in Multimedia（SIGMM）が主催するフラッグシップ会議であり，1993年に開催以来，今回は22回目を迎えた．

ACM Multimediaは採択率が低く，非常に競争が激しい国際会議の一つであることはすでに多くの読者がご存じだと思うが，その具体例を一つご紹介しようと思う．中国やオーストラリアなど一部の国では国際会議にランク付けを行い，博士号取得や若手教員の登用時などの点数化に利用している．ACM MultimediaはそれぞれSection A, Rank A+と最上位を獲得している．

ACM Multimediaは基本的に北米，アジア，ヨーロッパを順番に廻って開催されている．本年は米国・フロリダ州・オーランドのHilton Orlando Lake Buena Vistaで開催された（図1）．当該ホテルはWalt Disney Worldのオフィシャルホテルの一つで，敷地内にある．Disney Downtownと呼ばれるショッピング・レストラン街と目の鼻の先であり，各パークまでの無料送迎バスも出ている（ただし，Walt Disney Worldの敷地面積は山手線内側の1.5倍程度であるため，ホテルから各パークへの移動にはバスでも片道30分ほどかかる）．

本稿では，2章で会議の概要，3章で2014年の研究動向などを紹介する．4章・5章では会議に参加しての雑感や来年の開催地について紹介する．（山崎）

†東京大学 大学院学際情報学府 学際情報学専攻
††東京大学 大学院情報理工学系研究科 電子情報学専攻
"ACM Multimedia (ACMMM) 2014 Report" by Hokuto Kagaya (Graduate School of Interdisciplinary Information Studies, the University of Tokyo, Tokyo) and Toshihiko Yamasaki (Department of Information and Communication Engineering, Graduate School of Information Science and Technology, the University of Tokyo, Tokyo)



図1　会場となったHilton Orlando Lake Buena Vista

2．会議概要

(1) 基調講演・受賞記念講演

基調講演は本会議初日と3日目の午前に行われた．初日はMicrosoft CorporationのHarry Shum氏によるBing（Microsoft社が展開している検索エンジン）の技術的発展の変遷やその中で使われている画像認識技術の紹介であった．Harry Shum氏はImage Based Renderingを中心とするグラフィックスやコンピュータビジョンの分野の第一人者でMicrosoft Research Asiaの初代所長も歴任しているが，その後本社に戻りBing開発の責任者，Microsoftの研究部門の統括として活躍している人物である．3日目はMIT Media LabのRosalind W. Picard教授によるウェアラブルデバイスや顔表情解析による人間の内部状態推定に関する一連の研究成果の紹介が行われた．特に興味深かったのは，てんかん患者は発作前に比較的長時間皮膚の伝導率に変化がみられ，ウェアラブルデバイスで発作の予測が可能であるという研究である．しかもその現象発見に至った経緯は，たまたま兄弟に転換発作をもつ学生が興味本位でデバイスを自宅に持ち帰ったことによるというのだから面白い．ほかにも，クラウドソーシングによる顔の表情と感

情のデータ収集システムなどの研究についても紹介があった．それぞれの研究成果はIndiegogoを通したデバイス発売[1]や無料WEB API[2]の提供などにより一般に還元されている．

2日目にはSIGMM Awardの受賞記念としてUniversity of Illinois at Urbana-ChampaignのKlara Nahrstedt教授の講演が行われた．彼女は長年マルチメディアのネットワーク配信に関する研究を行っており，ほかにも数々の受賞を誇る当該分野の第一人者である．（山崎）

(2) チュートリアル

チュートリアルは，11月3日に半日・全日のもの合わせて下記の7つが提供された．どの部屋も30～50名程度収容の小さめの部屋であったため，賑わい感があった．ちなみに，著者らは＃4にてライフログ・食事ログに関する講演と，ソーシャルメディアにおけるジオタグ処理に関する講演を行った．

1. Immersive 3D Communication
2. Social Multimedia Computing
3. Learning Knowledge Bases for Text and Multimedia
4. Emerging Topics on Personalized and Localized Multimedia Information Systems
5. An Introduction to Arts and Digital Culture inside Multimedia
6. Video Hyperlinking
7. Over-the-Top Content Delivery: State of the Art and Challenges Ahead （山崎）

(3) オーラル・ポスタセッション

2日目すなわち11月4日から6日までの3日間が本会議にあてられた．例年，Full Paperがオーラルセッション，Short Paperがポスタセッションという形で発表される．オーラルは最大同時2並行にて行われ，ポスタはその裏で常時展示の上，コーヒーブレイク中の30分にプレゼンテーションタイムが割り振られるという形での発表であった．ポスタセッションの発表者に設けられた時間は少なく，聴講する側としても所望の発表の著者と必ずしもコンタクトをとれないため，時間を上手に使う必要があった．ポスタ会場の様子を図2に示す．

今年のACM Multimediaでは以下の14のテーマが設定された．

1. Media Transport and Delivery
2. Multimedia Systems and Middleware
3. Big and Broad Multimedia
4. Deep Learning for Multimedia
5. Emotional and Social Signals in Multimedia
6. Multimedia and Society
7. Multimedia and the Crowd
8. Multimedia Art, Entertainment and Culture
9. Multimedia HCI and QoE
10. Multimedia Search and Recommendation
11. Multimedia Security, Privacy and Forensics
12. Multimodal Analysis and Description
13. Music, Speech and Audio Processing in Multimedia
14. Social Media and Collective Online Presence



図2　ポスタ会場の様子

表1　Full Paper, short paperの各テーマへの投稿数
（プログラム・ブックレットより抜粋）

Area	Full papers	Short papers
Big and Broad Multimedia	10	8
Deep Learning for Multimedia	23	21
Emotional and Social Signals in Multimedia	16	29
Media Transport and Delivery	18	15
Multimedia and Society	11	13
Multimedia and the Crowd	7	8
Multimedia Art, Entertainment and Culture	19	19
Multimedia HCI and QoE	20	38
Multimedia Search and Recommendation	56	80
Multimedia Security, Privacy and Forensics	10	23
Multimedia Systems and Middleware	17	25
Multimodal Analysis and Description	43	75
Music, Speech and Audio Processing in Multimedia	8	18
Social Media and Collective Online Presence	14	21

参考のために，各テーマへの投稿数を表1に示す．特段目を引くのは“Deep Learning for Multimedia”であろう．Deep Learningとは，多段に接続したニューラルネットワークを用いた機械学習手法群の総称であり，近年機械学習コミュニティだけでなくコンピュータビジョン，自然言語処理等の界隈でも注目されているトピックである．そのDeep LearningがついにMultimediaというユーザに近いアプリケーションの分野にも登場したという事実は興味深い．実際，このテーマでの投稿数も相当数あったようである．その他では，Deep Learning同様アカデミックにとどまらない流行であるBig Dataの流れを汲んでか，Big and Broad Multimediaというテーマも登場した．（加賀谷）

(4) Multimedia Grand Challenge

例年開催されているMultimedia Grand Challengeは，企

業等があるテーマを設定し，それに対して科学的なアプローチによるソリューションを提案するというものである．多くの場合，テーマ設定を行った企業から貴重な実験用データなどが提供される．一般的なコンテストなどでは参加者全員で同じ問題を解き，その性能を競うことが多いが，このグランドチャレンジでは問題設定そのものを参加者が行うため，同一のテーマに対しても多種多様な問題と解法が提案されるのが特徴である．Multimedia Grand Challengeはシングルセッションで開催される．会議内でどれだけ重要視されているセッションであるかおわかりいただけるであろう．

今年は以下に挙げる5つのテーマが設定された．

1. New York City 360° Challenge（IBM）
2. MSR-Bing Image Retrieval Challenge（Microsoft）
3. VideoLecture.NET Challenge（VideoLectures.NET）
4. Emotional Response to Multimedia Content Grand Challenge（Technicolor）
5. Machine Learning for Computational Photography（Huawei）

それぞれの課題に対し，合計10件の発表が行われた．1st Prizeに輝いたのは，IBMの課題に取組んだ“New Yorker Mélange: Interactive Brew of Personalized Venue Recommendations”である．ニューヨークを旅行する際，どんな観光地を訪れるべきなのか，初めての場合は混乱することがあるだろう．それを解決するために，自分と似たような志向の人のデータを利用して，上手にナビゲートしようという取組みである．データはFlickr, Foursquare, Picasaといったsnsから画像とタグを収集し，視覚的な特徴と文書特徴を用いてクラスタリングを行っている．特筆すべきは，インタラクティブなインタフェースで，実際にWebサイトで試すことができる点である[1]．蛇足だが，この発表では画像特徴抽出の際にDeep CNNが用いられており，こういったところにも流行の波が垣間見える．2nd Prizeにはニューヨークのファッショントレンドを詳細に解析した“What are the Fashion Trends in New York?”，3rd Prizeにはtechnicolorの課題に取組み，GIFアニメを見た際に視聴者が抱く印象を推定した“Predicting Viewer Perceived Emotions in Animated GIFs”が選ばれた．（加賀谷）

(5) Open Source Software Competition

Open Source Software CompetitionはACM Multimediaの面白い試みの一つであり，オープンソースとして公開しているソフトウェアのコンペを行うイベントである．本年は，Convolutional Neural Networkのソフトウェアとして有名なCaffe（Caffe: Convolutional Architecture for Fast Feature Embedding）が受賞した．（山崎）

(6) Technical Demo

ACM Multimediaのデモセッションは，Technical Demoと称して，最先端の技術をバックグラウンドにもつシステム・アプリケーションなどを，会議参加者にインタラクティブな形で体験してもらうという目的で行われている．当日はポスタセッションと並行して同じ部屋の一角を利用して発表が行われた（図3）．筆者が食事に関する研究を行っているため，食事を題材としたデモを二つ紹介しておこう．電気通信大の河野氏らによるFoodCam-256[4]は，スマートフォンベースのリアルタイム食事品目認識システムであり，独自のデータベースに含まれる256品目を認識することができる．リアルタイムに処理を行う工夫として，SVMの重み行列に圧縮を行っている．ニューヨーク大・シンガポール大のグループによるTaste+は，独自のデバイスを用いて味覚を拡張するデモを行っていた．（加賀谷）



図3　デモの様子

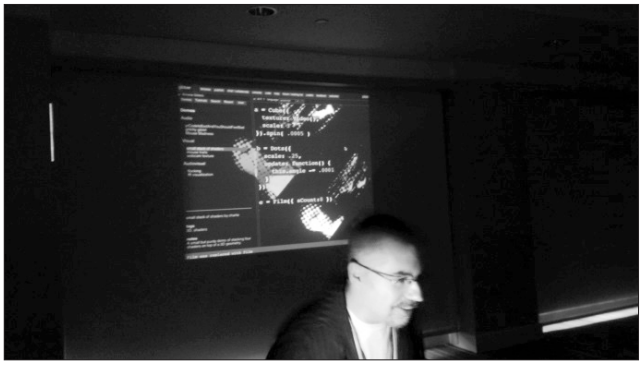


図4　アート展示の様子（写真はSound-Light Giblet）

(7) アート展示

ACM MultimediaではMultimedia Art Exhibitionというメディアアート展示があるのも大きな特徴である．Art exhibitionは本会議期間中別フロアで開催され，7件の展示があった．特に目を引いたのがUCサンタバーバラのC.Roberts氏によるSound-Light Gibletという作品で，WEBブラウザ上でJavaScriptを用いてリズムやビートなどを記述していくだけで，即興で音楽と視覚効果を生成できるというものである（図4）．言葉で表現しても伝わらないと思うので，ぜひ氏のプロジェクトページにあるでもビデオをご覧いただきたい[5]．バンケットの際にもライブデモ

が披露された．国際会議でのバンケットでは現地の伝統芸能などが披露されることは多いが，このような形で発表作品を体験できるのは筆者自身も初めてのことである．（山崎）

(8) ワークショップ

学会最終日11月7日には9つの全日，3つの半日ワークショップが開催された．本会議も終わり，会議日程としても最終日のため参加者は少ないかと思いきや，どのワークショップも賑わいを見せていた．今年はワークショップのみに参加する人のためにワンデー・チケットを販売したところ好評だったという報告がなされたが，それを裏付ける形となった．（山崎）

3．研究動向

まずはBest Paper Awardをはじめとする受賞論文について簡単に述べたい．受賞作品の一覧はACM Multimedia 2014のWebサイトにて閲覧可能である．さて，受賞論文を含む3件の論文のプレゼンテーションがBest Paper Sessionとして会議2日目の午前中に行われた．Best Paper Awardを受賞したのはAmirhossein Habibian氏らによる“VideoStory: a New Multimedia Embedding for Few-Example Recognition and Translation of Events”[6]である．こちらはまさにMultimediaの正統派の研究であり，映像とそれについた少ない説明文を基に対応関係を学習し，新たな映像に対してフレームごとにもっともらしいタグを付与するといったものである．

他のFull Paperから筆者の目に留まったものをいくつか紹介したい．まず，先に述べたDeep Learningに関するセッションから，Ji Wan氏らによる“Deep Learning for Content-based Image Retrieval: a Comprehensive Study”[7]．こちらはDeep Learningを画像検索に応用したものである．画像認識の分野ではすでに高い精度が示されているConvolutional Neural Network（CNN）を特徴抽出器として用いている．何も工夫をせずとも従来のBag of Wordsベースの手法より精度が高い上，Metric Learningや，新たなタスクのデータセットを用いたPre-trainingを行うことにより，より良い結果となることを報告している．まだこの分野での研究は始まったばかりであり，今後の同グループの動向も興味深い．

さらに同一セッションから，北京大のZheng Zhang氏らとMicrosoft Researchのグループによる“Error-Driven Incremental Learning in Deep Convolutional Neural Network for Large-scale Image Classification”[8]も大変興味深い発表だった．より実生活に近い一般物体認識を行う上で残る課題として，snapshot的に学習するのではなく，識別すべきクラスが次々と登場する場合をどうするかというものがある．この論文では，画像認識のstate-of-the-artであるDeep CNNをベースとして，そのような状況に対応するような仕組みを提案している．具体的には，最初段階の状態からモデルを木構造的に分割し，階層的なフローを作る．彼らによれば，これによりBengioのいうところのカリキュラム学習[13]になっているのではないかということである．古くからオンライン学習の研究がなされているように，最初から識別したいクラスが定められていたり，充分な学習データが用意されていたりでない問題設定は，今後重要なタスクとして取組まれるだろう．

さらに，コンピュータビジョン系の学会でもブームの1つになっているFine-Grained Visual Categorization（FGVC）についての発表も散見された．例えば，上海交通大のXiaopeng Zhang氏らによる“Fused One-vs-All Mid-Level Features for Fine-Grained Visual Categorization”[9]では，認識対象物体のパーツごとの識別結果をMid-Level特徴として用いる手法を提案した．さらに，画像特徴的に類似したクラスの特徴を併合して利用している．

さて，ポスタセッションの形式で発表が行われたShort Paperでは非常にバラエティに富んだ研究を見ることができた．Short Paperにおいても，Deep Learning的な手法を取り入れたものは多く，今後は単なる構成要素として使われるようになっていくものと推測される．その意味で，Caffeなどのライブラリーの開発が進むことはコミュニティにおいて大きな意味をもつため，Open Source Software Awardにおける受賞は必然のものと言えよう．研究内容に話を戻すと，感情推定やアトリビュート推定などの画像認識応用の話題や，Twitterの解析，効率的なハッシュに関する話題，新たな推薦システムなどなどテーマは多岐にわたる．題材となる対象も衣服，食事，音楽，映像，風景，人間の姿勢など飽きないほどであり，ポスタセッションの会場が非常に色味豊かであった．

筆者が惹かれたものをいくつかご紹介する．マルチメディアの学会らしくファッションに関する研究も多かったが，東北大の山口光太氏らによる“Chic or Social: Visual Popularity Analysis in Online Fashion Networks”[10]は興味深かった．近年日本でもその日のコーディネートをSNSにアップロードして楽しむ文化が普及しつつあるが，そういったコーディネートの人気を左右するのはどういった要素なのかという部分に着目している．その写真の画像特徴や投稿者によるタグといったコンテンツ要素と，投稿者自身に関する情報や，投稿者のフォロー・フォロワー数などといったソーシャル要素について解析を行っている．なお，画像解析には筆者らが提案しているパーシング技術を用いている．結果としてはSNS上の人気度にはソーシャルな要素が大きく影響しているのではないかと述べられている．

トレント大学のBo Zhang氏らによる“You Talkin' to Me? Recognizing Complex Human Interactions in Unconstrained Videos”[11]では，非実験室環境下で撮影された映像に対して，握手，ハグ，キス，ハイタッチといった複数名の人間によるインタラクションを検出している．

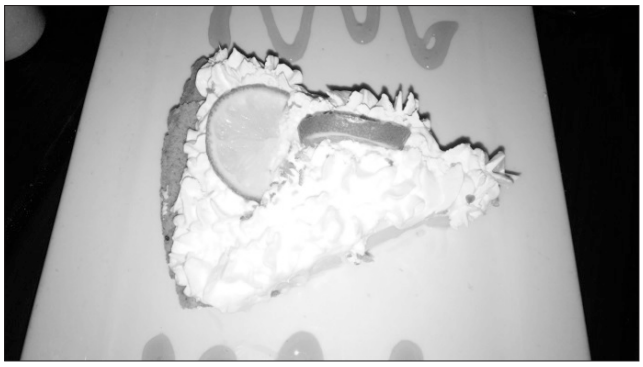


図5　フロリダ名物「キーライムパイ」

人間の行動を検出する研究の中でも，これらはテレビやYouTube等のエンタテインメント寄りに生かされる研究であり，ACM Multimediaらしさが見えると感じた．

画像へのアノテーション問題はよく題材にされる課題だが，台湾大のHong-Wun Jheng氏らの“Automatic Facial Image Annotation and Retrieval by Integrating Voice Label and Visual Appearance”[12]では，画像，特に人物の顔の写った写真のアノテーションに人間の声を用いることでそれを補助しようと試みられている．顔検出というCV的手法と，音声認識という音声情報処理的手法の組合せというマルチモーダルな視点が印象に残る．どの研究にも言えることだが，実用化された場合のビジョンが門外漢にも見えやすいところがこの分野のいいところだろう．（加賀谷）

4．雑　感

今年は，ビザ発行の遅れのために中国本国からの参加が極端に少なく，代理発表やビデオプレゼン，スカイプによる質疑応答などが多く目についた．それ自体は仕方のないことであるが，どうしても臨場感にかけるプレゼンとなってしまった．高臨場感遠隔会議システムは，マルチメディアの分野で長く研究されてきたにもかかわらず，まだ充分満足な臨場感が得られる会議システムが実現できていないということは，われわれの分野はまだまだ発展途上だね，と一部の研究者らの間でジョークのように語られていた．

蛇足であるが，Walt Disney Worldのおひざ元ということで，レストランは充実していたように思う．フロリダはストーンクラブというカニやキーライムパイ（図5）などが有名である．キーライムは，アメリカ合衆国最南端であり文豪ヘミングウェイが滞在したキー・ウェストへ続くフロリダ・キーズと呼ばれる列島に由来している．ヘミングウェイも「老人と海」を執筆しながらこのキーライムパイを食べたのだろうか，などと思いをはせながら食べるとまた違った気分を楽しめる．（山崎）

5．ACM Multimedia 2015

来年のACM Multimediaはアジア地区担当で，オーストラリア・メルボルンで開催予定である．フルペーパは2015年3月末，ショートペーパは2015年4月7日の締切予定である．そのほかにも上記に紹介したとおりグランドチャレンジやデモ，ワークショップなどさまざまな参加形態が可能である．ぜひ読者の皆様の参加を促したい．（山崎）

（2014年12月12日受付）

〔文　献〕

1）http://www.empatica.com
2）http://www.affectiva.com
3）http://newyorkermelange.com
4）Y. Kawano, K. Yanai: "FoodCam-256: A Large-scale Real-time Mobile Food Recognition System employing High-Dimensional Features and Compression of Classifier Weights", ACM Multimedia（2014）
5）http://www.charlie-roberts.com/index.htm
6）A. Habibian, T. Mensink, C.G.M. Snoek: "VideoStory: A New Multimedia Embedding for Few-Example Recognition and Translation of Events", ACM Multimedia（2014）
7）J. Wan, D. Wang, S.C.H. Hoi, P. Wu, J. Zhu, Y. Zhang, J. Li: "Deep Learning for Content-based Image Retrieval: A Comprehensive Study", ACM Multimedia（2014）
8）T. Xiao, J. Zhang, K. Yang, Y.P.Z. Zhang: "Error-Driven Incremental Learning in Deep Convolutional Neural Network for Large-scale Image Classification", ACM Multimedia（2014）
9）X. Zhang, H. Xiong, W. Zhou, Q. Tian: "Fused one-vs-all mid-level features for fine-grained visual categorization", ACM Multimedia（2014）
10）K. Yamaguchi, T.L. Berg, L.E. Ortiz, C.O. Social: "Visual Popularity Analysis in Online Fashion Networks", ACM Multimedia（2014）
11）B. Zhang, Y. Yan, N. Conci, N. Sebe: "You Talkin' to Me? Recognizing Complex Human Interactions in Unconstrained Videos", ACM Multimedia（2014）
12）H.-W. Jheng, B.-C. Chen, Y.-Y. Chen, W. Hsu: "Automatic Facial Image Annotation and Retrieval by Integrating Voice Label and Visual Appearance", ACM Multimedia（2014）
13）Y. Bengio, J. Louradour, R. Collobert, J. Weston: "Curriculum learning", ICML（2009）



加賀谷北斗（かがやほくと）　東京大学工学部電子情報工学科修了．現在，東京大学大学院学際情報学府学際情報学専攻修士課程在学中．食事画像認識等の研究に従事．学生会員．



山崎　俊彦（やまさきとしひこ）　東京大学工学系研究科電子工学専攻博士課程修了．現在，同大学情報理工学系研究科電子情報学専攻准教授．2011年～2013年，コーネル大学訪問研究員．ソーシャルメディアデータを用いた処理，物体認識，3次元映像処理などに従事．博士（工学）．正会員．