

ISM 2006 参加報告

北原 鉄朗

平成 18 年 12 月 19 日

1 General information

正式名称 The Eighth IEEE International Symposium on Multimedia. IEEE Computer Society 主催のマルチメディア全般を対象とした国際シンポジウムである。開催日は 2006 年 12 月 11~13 日、開催地は San Diego, California で、Paradise Point Resort & Spa という季節外れの大規模リゾート施設での開催であった。チュートリアル等はなく、3 日間の各々の朝に keynote speech と panel discussion があつた後、通常発表セッションとなる。通常発表セッションは 3 部屋でのパラレルセッションで、1 つが general session, もう 1 つが special track で、最後の 1 つが workshop である。Workshop が本会議の前後の日程であるのではなく、本会議の一部として開催されるのが特徴的である。また、オーラルセッションが中心で、ポスター/デモセッションは 2 日の夕方 の 1 セッションのみであった。採択率は 35%以下とのことで、少数精鋭のレベルの高い会議を期待したが、参加した範囲では、それほど活発な議論が繰り広げられているようには感じなかった。

対象分野はマルチメディア全般と非常に広く、マルチメディア信号処理からネットワーク/セキュリティ、e-learning や virtual reality までカバーしている。しかし、印象としては画像処理の占める割合が大きく、大半が画像処理系、数割がネットワーク系 (peer-to-peer, プロトコル, セキュリティなど) で、音響処理系は数件 (我々含め) であった。しかも、その中にエンコーディングなどの発表も含まれているため、我々のモチベーションに近い研究 (音響処理+パターン認識) は、我々の他にはほとんどなかった。また、“マルチ”メディアと言うわりには複数メディアの統合を扱った研究は少なかった (少なくとも、私が聴講したセッションにはなかった)。

私にとって初体験の形式として、reception や banquet の中に keynote speech があつた。内容は (陽には出さないが) 企業宣伝を意図していると思われるものであつた。keynote speech 中は私語は慎まなければならない、せつかくの交流の機会を奪われ、非常に不快であつた。

2 自分の発表

あまり自分の満足の行く出来ではなかった。自分自身の英語力不足により、質問内容が聞き取れず、実りのある議論には至らなかった。

最初の質問者の質問は数割聞き取れたのだが、内容が全く理解できなかった。発表後に再度確認したところ、HMM のモデル化の部分で、倍音の各々を別々の HMM でモデル化して何らかの方法で統合するのだと勘違いしていたようだ。次回、同内容を発表するときはこのような誤解が生じないよう工夫したい。

2 人目の質問は、どのくらい複雑な音楽には対応できるのか、といった主旨のものであつた。

Audio processing のセッションは聴講者が少なく (10 人強ほど?), あまり盛り上がっていません。このような会議で発表するのであれば, visualization をメインに据えるなどの工夫が必要と思われる。

3 他の発表

いくつかの発表のみを紹介する。

- **A New Multimedia Content Skimming Method based on Speech Emphasis Extraction and Its Application to Content Variations**

音声の目立ち具合を信号処理により計算し, 目立っているところに瞬時に頭出しができるシステム。用いる特徴量 (たとえばパワーの計算に用いる周波数帯域) やデータの種類 (サッカー中継, ホームビデオなど) による精度の違いを検討していた。

- **A Semi-automatic Video Annotation Tool with MPEG-7 Content Collections**

MPEG-7 の特徴量セットを用いて映像の各カットを階層的クラスタリングを行って, カット間の階層的構造をアノテーションするものである。手法自体はいたってオーソドックスなものだが, MPEG-7 の特徴量セットを用いているとのことなので, 実装について質問したが, 発表者自身が実装したわけではないのでわからないとのこと。MPEG-7 の特徴量セットを使って, という研究は画像処理の研究では他にもいくつかあったが, MPEG-7 完全準拠のアノテーションツールあるいは検索ツールというのはいまだあまり扱われていないようだ。ポスターにも “Due to the inflexibility of the MPEG-7 standard, a lot of uncompliant formats have been proposed.” と書かれていた。

- **Blind Detection for Additive Embedding Using Undetermined ICA**

情報ハイディングの一手法。元となる画像とそれに埋め込む情報を独立と仮定して, 観測された画像からそれらを分離する。この場合, 観測データ (1ch) より出力データ (2ch 相当) の方がチャンネル数が多いので, undetermined ICA という方法を用いている。発表では文献名のみ書いてあり, 詳細はわからなかった。

- **OreDesk: A Tool for Retrieving History Data based on User's Operations**

データの内容には一切踏み込まず, データへのアクセスのされ方だけからデータの関係性を見いだすことを目的としている。たとえば, ある web ページとある文書ファイルが同時に開かれていたらそれらは関連があるとか, ある文書からある文書へコピー&ペーストがされたら関連があるとか, そのような方法である。

- **The Effect of Key and Tempo on Audio Onset Detection Using Machine Learning Techniques: A Sensitivity Analysis**

我々の発表を除いて唯一の音楽信号処理の研究発表である。この研究では onset detection を machine learning で行うが, 学習データとテストデータにおける key や tempo の違いの影響や認識手法 (neural net, SVM, naive Bayes) による精度の違いの検討をしていた。