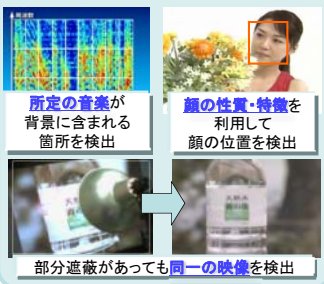
音や映像から「部品」を取り出す
ーメディアシーン学習が切り拓く次世代メディア解析ー

背景・課題：メディア情報処理の新しいパラダイムとして メディアの「中身」を自動解析する枠組を提案
メディアシーン学習 *(Media Scene Learning, MSL)*：複雑かつ多様な音・映像（メディアシーン）から，事前知識や参照情報を用いずとも「聞こえる音・映っているもの」を自動的に取り出し学習する枠組

アプローチ：メディアデータに内在する性質に基づき メディアを構成する主要な要素を自動的に抽出
音響信号：個々の音の生成 / 混合過程を考慮して，所定の信号を最もよく表現する構成音の組合せを特定
映像信号：重要性の判断基準として「目立つかどうか」に着目し，映像中の主要な物体 / 建造物 等を特定

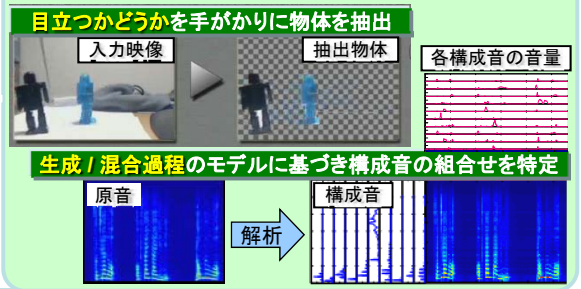
到達点：メディアの生成・観測機構を統計的にモデル化した新しいメディアシーン学習技術を考案
音響信号：音声の生成過程の統計モデルに基づく混合音声解析技術 CARS *(Composite Auto-Regressive System)*を考案
映像信号：初期視覚特性の統計モデルに基づく物体領域自動抽出技術 SBIL *(Saliency-Based Image Learning)*を考案

メディア探索・認識
・事前知識/参照情報を活用
・高速/頑健な一致性の検出



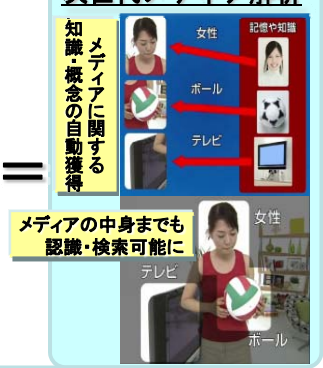
＋

メディアシーン学習のアプローチ
・メディア内在の性質に基づき主要要素を自動抽出
⇒ 広範な適用範囲・斬新な応用可能性の開拓

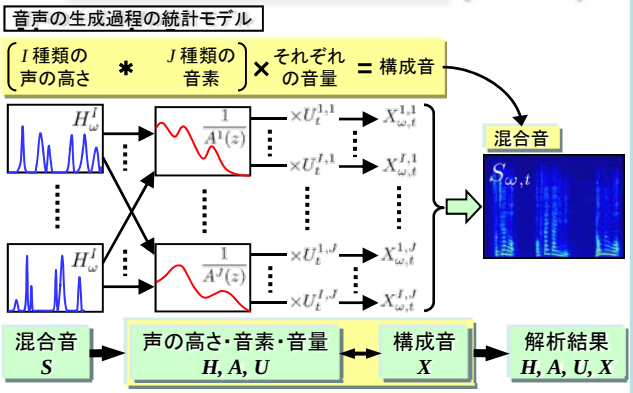


＝

次世代メディア解析



音響メディアシーン学習技術 [CARS]



構成音分解 / 雑音抑圧 / 構成音に基づく検索 などが可能に

映像メディアシーン学習技術 [SBIL]

構成物体検索 / 構成物体に基づくマイニング / メタ情報付与 が的確かつ容易に

関連文献

1. Kameoka & Kashino, "Composite Autoregressive System for Sparse Source-Filter Representation of Speech," In Proc. ISCAS2009, pp. 2477-2480, 2009.
2. Kimura, Pang, Takeuchi, Miyazato, Yamato & Kashino "A Stochastic Model of Human Visual Attention with a Dynamic Bayesian Network," *http://arxiv.org/abs/1004.0085*
3. 福地, 宮里, 木村, 高木, 大和, 柏野 "グラフコストの逐次更新を用いた映像顕著領域の自動抽出", 画像の認識・理解シンポジウム (MIRU2009) 予稿集, OS5-4, 2009年7月

連絡先

木村 昭悟 *(Akisato Kimura)*，亀岡 弘和 *(Hirokazu Kameoka)*，
大石 康智，ルルー ジョナトン，竹内 龍人※，柏野 邦夫，大和 淳司

メディア情報研究部 メディア認識研究グループ （※ 人間情報研究部）