



Web動画ショットの動作分類のための 時空間特徴抽出手法の提案

CVIM研究会

電気通信大学 情報工学専攻

野口 顕嗣

2010年3月19日



アウトライン

- はじめに
 - 背景, 研究の目的, 関連研究
- 提案手法
 - 時空間特徴抽出手法の提案
 - 特徴統合による分類手法の提案
- 評価実験
 - データセット
 - 動作認識に関する実験
 - Web動画分類に関する実験
- おわりに
 - 考察, 今後の課題



背景

- **Web上には大量の動画が存在**
 - Youtube
 - ニコニコ動画
- **見たい動画を探すためにはどうすれば良い？**
 - 現状ではテキストベースの検索手法
 - 動画を完全に特定することは困難
- **コンテンツベースのアプリケーションの必要性**
 - 動画の内容によって分類することは重要
 - 大量のWeb動画の分類を行った研究は少ない
- **Web動画分類の困難さ**
 - 背景ノイズ, 低解像度, カメラモーション
 - 新しい時空間特徴抽出手法を提案
 - 特徴を統合することによる精度向上



目的

- 本研究の目的は以下の二点にある
 - Web動画分類のための新しい時空間特徴の提案
 - カメラモーションに対する対応
 - 処理の高速さ
 - 大規模Web動画ショット分類
 - 教師信号ありのランキング付け
 - 教師信号なしクラスタリング
 - 特徴統合によるショット分類手法の提案

教師信号ありの大量Web動画ショット分類

大量のWebショット

Runningの学習セット

分類器

1 位



2 位

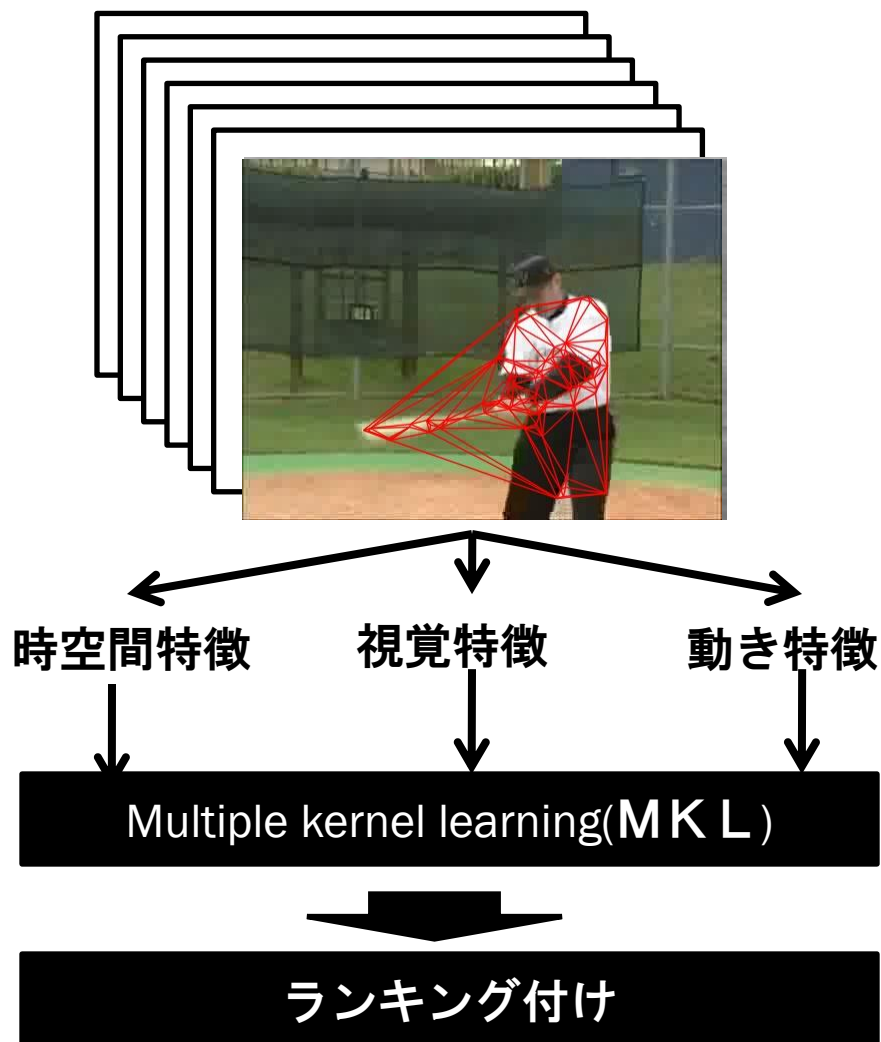


3 位





特徴統合による動作認識



教師信号なしのWeb動画ショットクラスタリング

大量なSoccerタグのWebショット

ドリブルシーン



シュートシーン



インタビューシーン





関連研究

■ Web動画における分類の研究

- CinbisらはWeb上から動作を自動学習する手法を提案[Cinbis et al 2009]
 - Web動画の動作分類
 - 学習には静的な画像からの特徴量を使用



本研究では、視覚特徴のみではなく、動き特徴も考慮

- LiuらはPage Rankに基づいて重要な点を選択する手法[Liu et al 2009]
 - 時空間特徴と視覚特徴を統合することでWeb動画の動作認識
 - 動作認識を行うために特徴をどのように利用するかに重点



分類のための時空間特徴抽出手法の提案



アウトライン

- はじめに
 - 背景, 研究の目的, 関連研究
- 提案手法
 - 時空間特徴抽出手法の提案
 - 特徴統合による分類手法の提案
- 評価実験
 - データセット
 - 動作認識に関する実験
 - Web動画分類に関する実験
- おわりに
 - 考察, 今後の課題



時空間特徴抽出手法

■ Web動画の特徴

- データ量が非常に大きい
- カメラムーションを含む
- 手振れなどによる動きのノイズ
- 低い解像度
- 雑多な背景ノイズ
- 撮影の視点変更

■ Web動画からの特徴抽出に重要なこと

- 高速に抽出可能
- カメラムーションに対する対応
- 撮影の位置に対して頑健
- 高い分類精度



時空間特徴抽出手法(概要)

カメラモーション検出

カメラモーションを検出したフレームは破棄



時空間特徴抽出



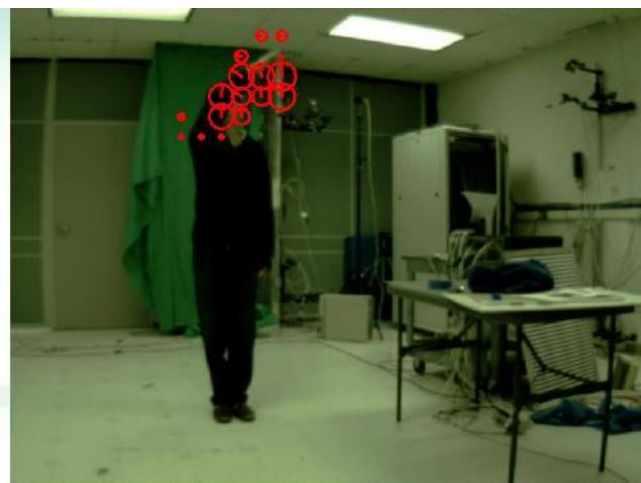
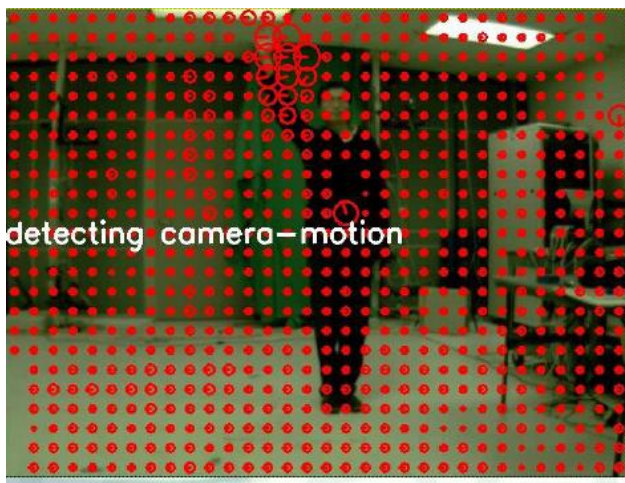
特徴をBag-of-Spatio-Temporal-Features(BoSTF)で表現



カメラモーション検出

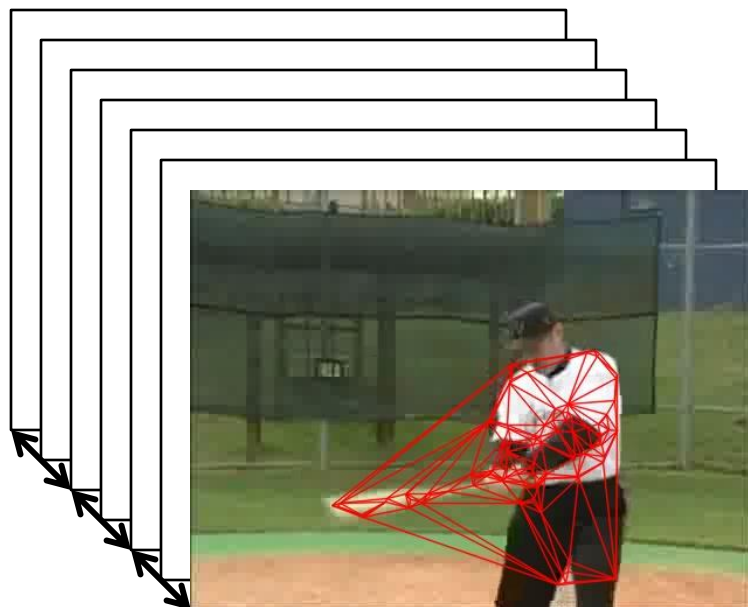
■ グリッドで動きを計算

- 動いていた領域が一定割合以上ならカメラモーション
- カメラモーションが検出されたフレームは破棄





時空間特徴抽出手法



①. Nフレームを1ユニットとする

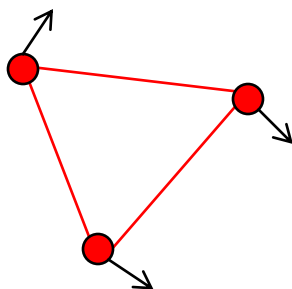
②. SURFを抽出

③. 動きがない点を削除

④. ドロネー三角形を作成

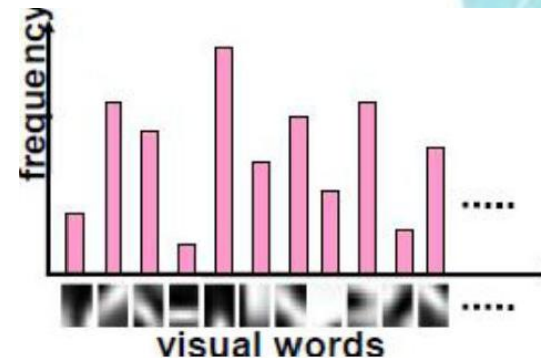
以降三点で一組の特徴と考える

⑤. ユニットを更に区切り,それぞれの
インターバルから動き特徴を抽出



Bag-of-Spatio-Temporal-Features

- **Bag-of-Features (BoF) を動画に拡張したもの**
 - 画像を特徴の出現頻度で表現したもの





特徴統合による分類手法

- 重要な特徴は異なる



MKLで自動で重みを算出



Multiple Kernel Learning(MKL)

■ 複数のサブカーネルを線形結合

- 最適な重み β を求める (MKL問題)

$$K_{combined}(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^K \beta_j k_j(\mathbf{x}, \mathbf{x}')$$

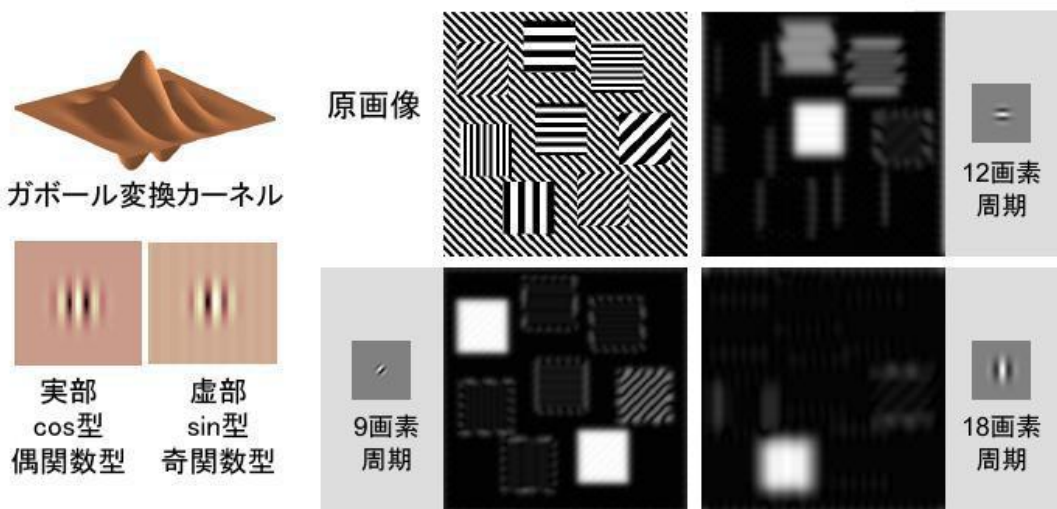
$$\text{with } \beta_j \geq 0, \quad \sum_{j=1}^K \beta_j = 1.$$

- 凸面最適化問題として解く



特徴統合による分類手法（視覚特徴）

- ガボールフィルタカーネルを用い、画像の局所的な濃淡情報を表現
 - フィルタカーネル：6方向，4周期
- 画像を 20×20 グリッドに分割
 - 一つのフレームから200の24次元の特徴を抽出





特徴統合による分類手法(動き特徴)

- **動き情報をグリッドごとに抽出**
 - 8方向7段階の大きさのヒストグラムに投票





特徴統合による分類手法

- キーフレームの選択は重要だが難しい
 - 選ばれたフレームによって特徴は異なる

そこで

本研究ではBag-of-Framesという考えを導入

- ◆すべてのフレームから特徴を抽出
- ◆抽出された特徴をベクトル量子化
- ◆フレームから抽出される特徴の出現頻度で動画を表現

キーフレームのみでなく，動画全体の特徴を考慮可能

時空間特徴，視覚特徴，動き特徴の3つをMKLで統合



アウトライン

- はじめに
 - 背景, 研究の目的, 関連研究
- 提案手法
 - 時空間特徴抽出手法の提案
 - 特徴統合による分類手法の提案
- 評価実験
 - データセット
 - 動作認識に関する実験
 - Web動画分類に関する実験
- おわりに
 - 考察, 今後の課題



評価実験

■ 動作認識

- KTHデータセット Leave-one-outで学習

■ Web動画分類

- 教師信号ありのランキング付け
- 教師信号なしのクラスタリング

データセット(動作認識)

■ KTHデータセット

- 6種類の動作, 合計599ショット





データセット(Web動画分類)

■ 教師信号ありランキング付け

動作	動画数	ショット数	学習データ	
			positive	negative
batting	174	8,980	31	75
running	170	7,342	28	66
walking	174	6,567	23	63
shoot	164	7,718	14	75
eating	142	3,442	22	64
jumping	160	3,130	27	40
合計	948	37,179	145	385

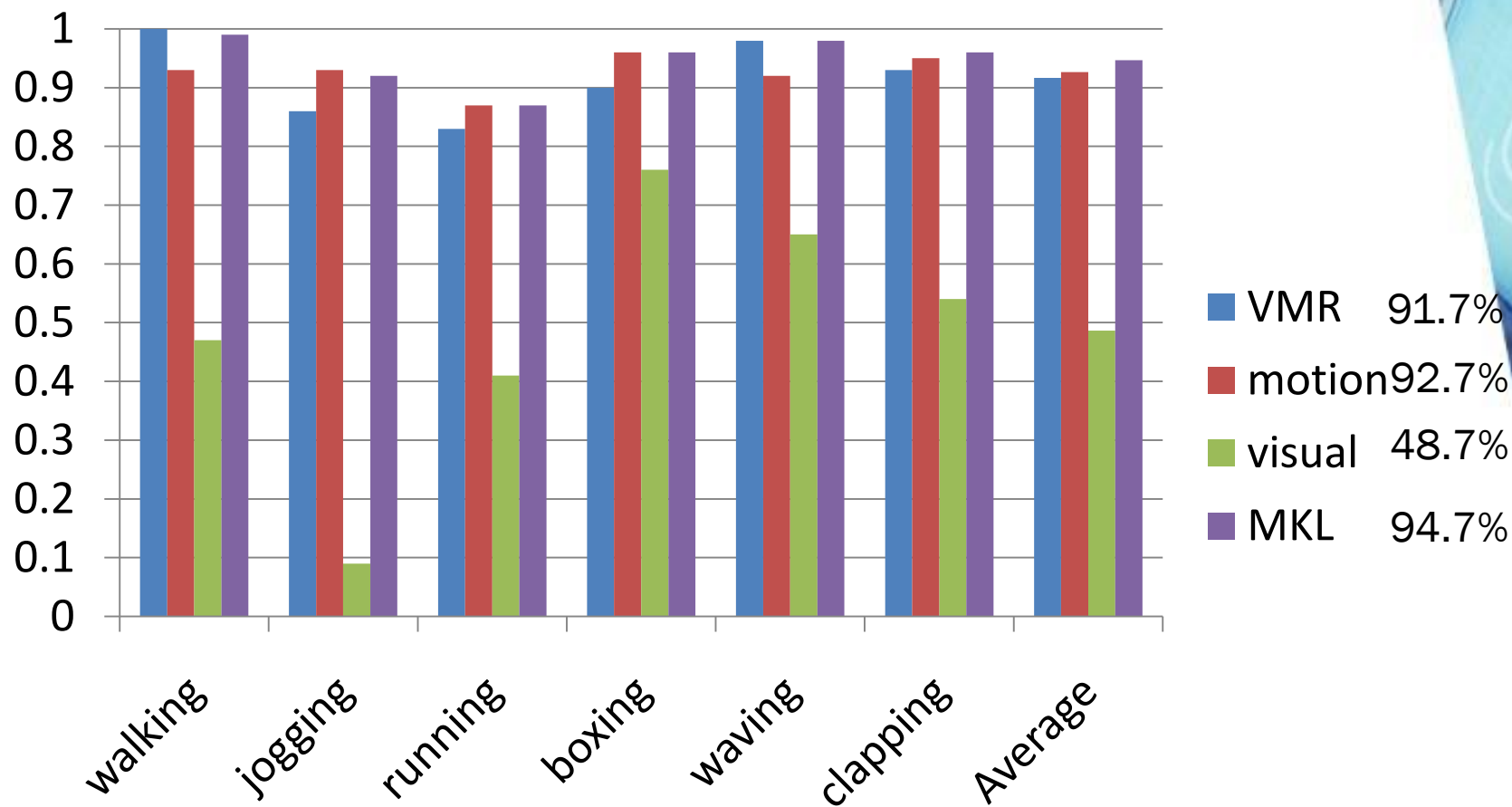
■ 教師信号なしのクラスタリング

- クラスタ数は200に設定

dancing	185	8,235
soccer	178	10,430

実験結果

■ KTHデータセット





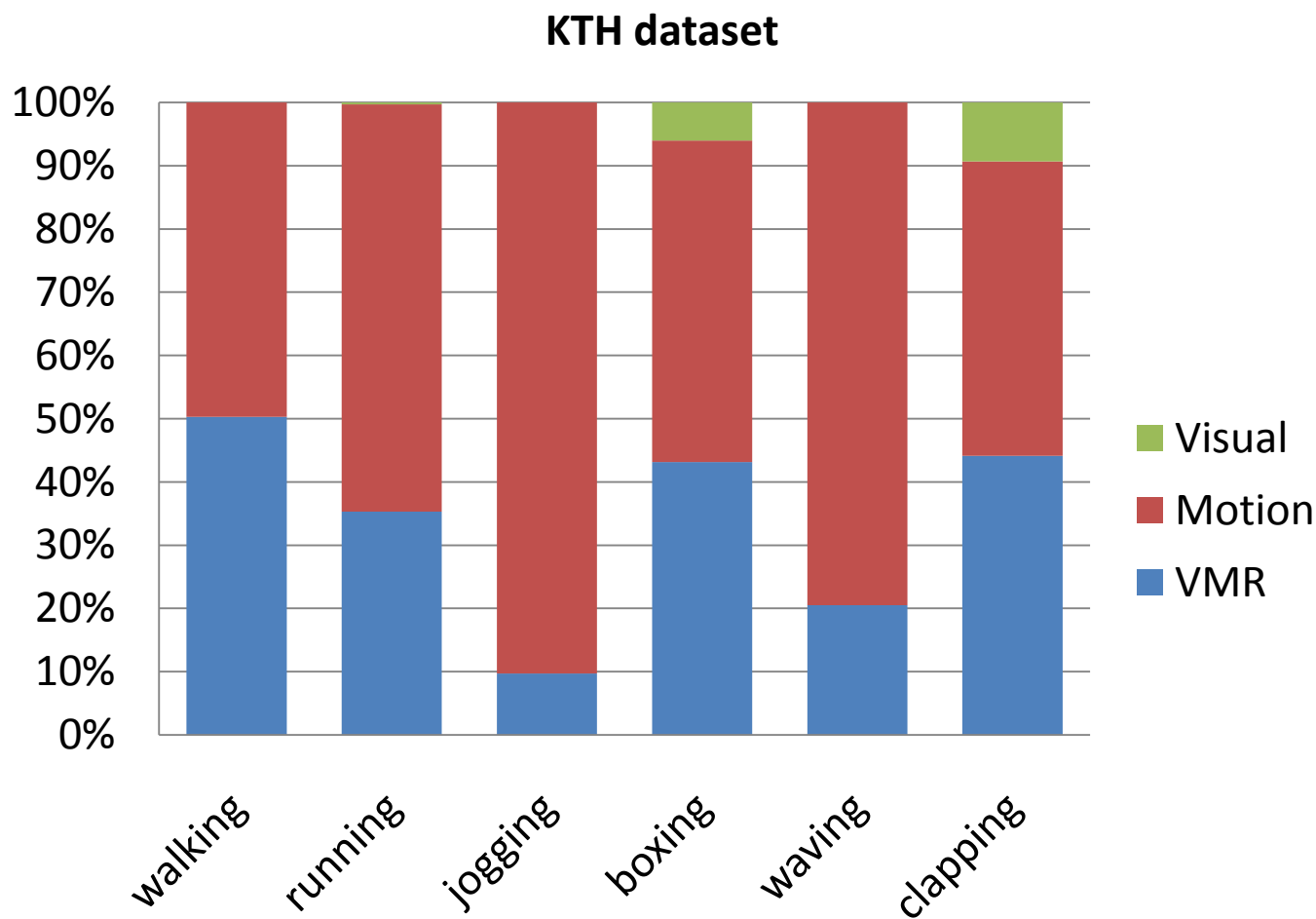
実験結果

■ KTHデータセット

	walking	jogging	running	boxing	waving	clapping
walking	0.99	0.01	0	0	0	0
jogging	0.04	0.92	0.04	0	0	0
running	0	0.13	0.87	0	0	0
boxing	0.01	0	0	0.96	0	0.03
waving	0	0	0	0	0.98	0.02
clapping	0	0	0	0.04	0	0.96

実験結果

■ KTHデータセット





実験結果

■ KTHデータセット

分類結果 (Leave-one-out)	
Ours	94.7%
Liu et al.	91.8%
Gilbert et al.	96.2%



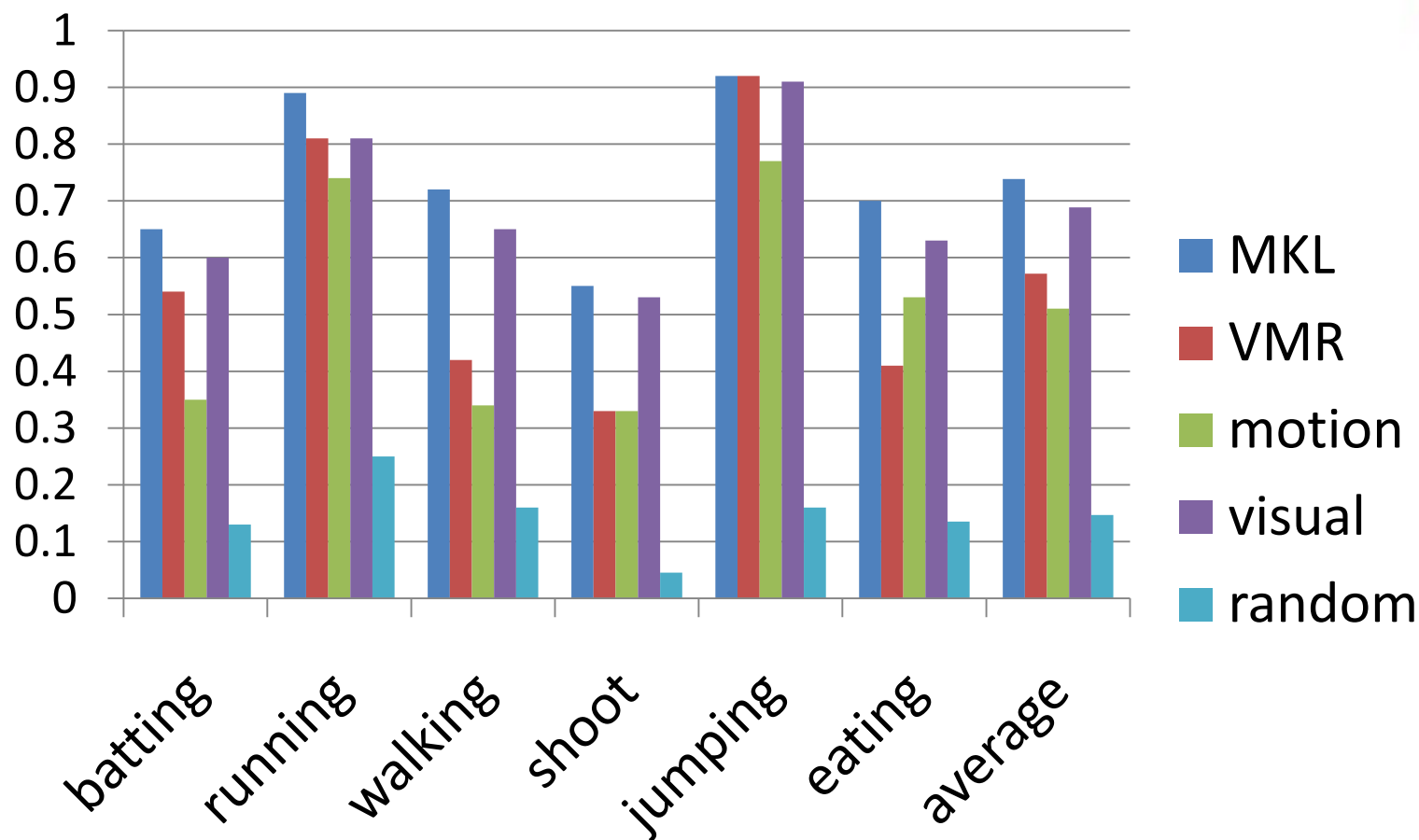
Web動画分類

- 教師信号ありランキング付け

結果のデモ

実験結果

■ 教師信号ありランキング付け





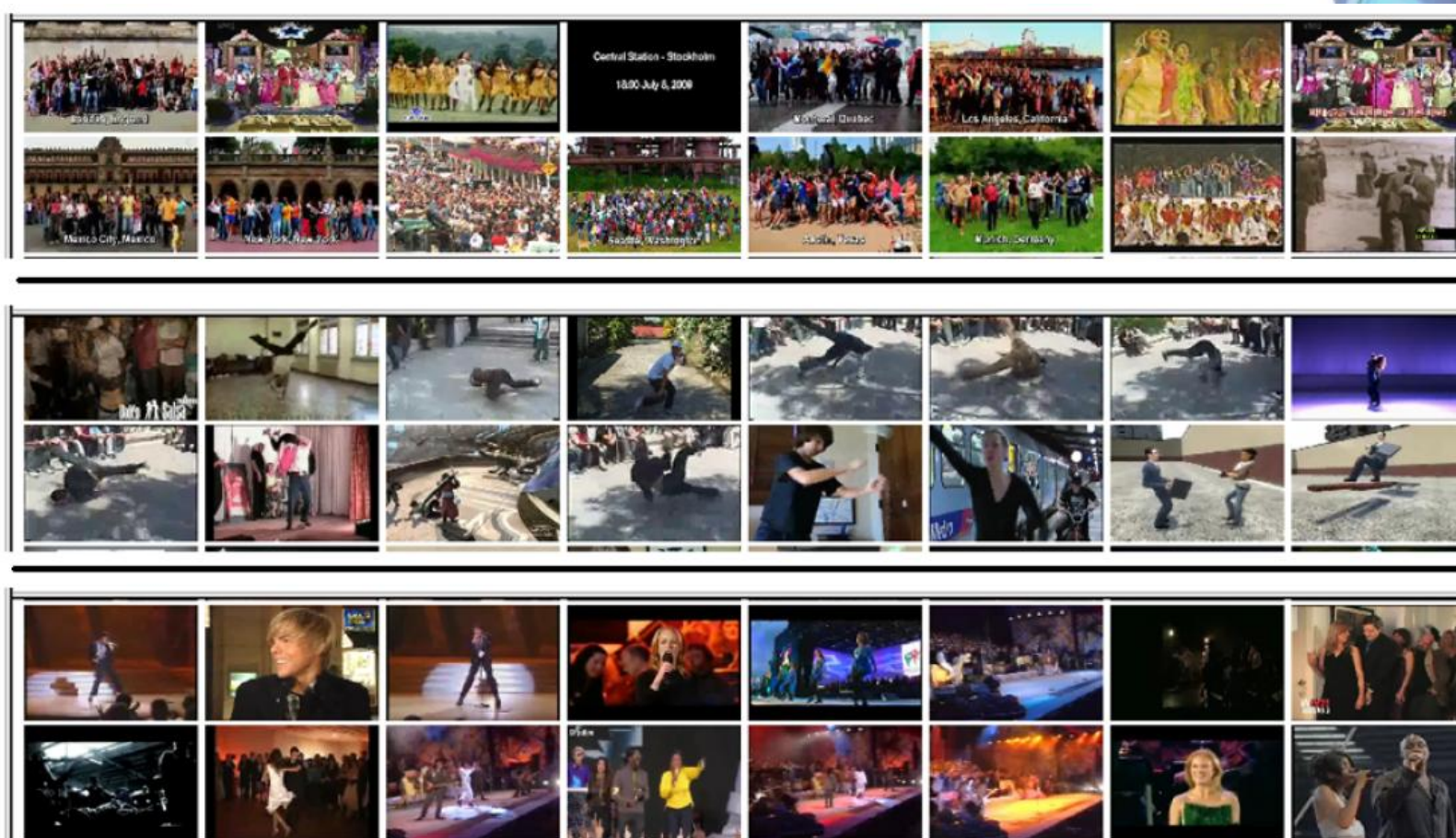
Web動画分類

- 教師信号なしクラスタリング

結果のデモ

Web動画分類

■ 教師信号なしクラスタリング





アウトライン

- はじめに
 - 背景, 研究の目的, 関連研究
- 提案手法
 - 時空間特徴抽出手法の提案
 - 特徴統合による分類手法の提案
- 評価実験
 - データセット
 - 動作認識に関する実験
 - Web動画分類に関する実験
- おわりに
 - まとめ, 今後の課題



おわりに

■ まとめ

- Web動画分類のための時空間特徴抽出手法を提案
- 特徴統合による動作認識手法の提案
 - KTHデータセットにおいて最新手法と同等
 - Web動画分類において高い精度
 - MKLによる特徴統合はWeb動画において有効

■ 今後の課題

- 時空間特徴抽出手法に関する課題
- カメラモーションに動き補正
- 有益な特徴の選択

ご清聴ありがとうございました