



多種類特徴統合による動作認識手法の提案

PRMU研究会

電気通信大学 情報工学専攻

野口 顕嗣

2010年3月16日

概要

■ 動作認識手法の提案

- 特にWeb動画に着目
- 新しい時空間特徴抽出手法を利用
- MKLに基づく特徴統合による精度向上について議論

■ 評価実験

- 3種類のデータセットを利用
 - 標準データセットであるKTHデータセット
 - 2種類のYoutubeデータセット
- 最新手法に匹敵する結果

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
 - 特徴統合による動作認識手法の提案
- **評価実験**
 - データセット
 - 動作認識に関する実験
- **おわりに**
 - 考察, 今後の課題



背景

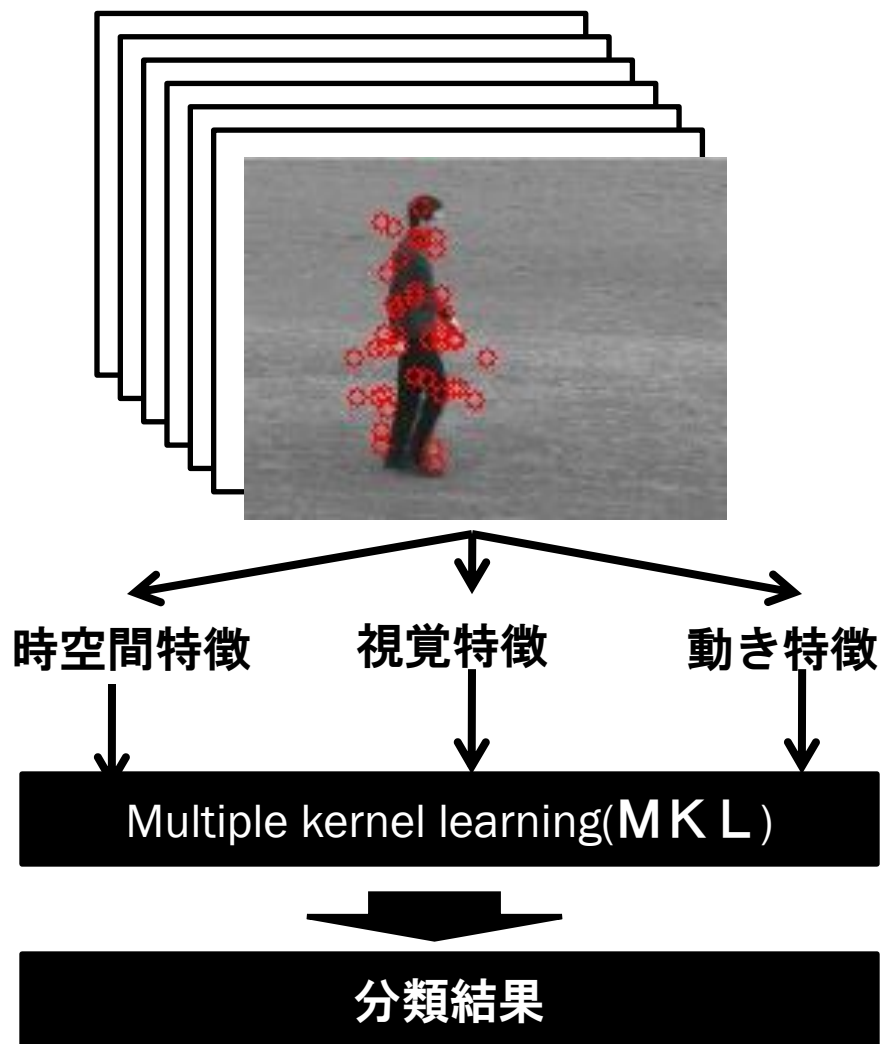
- **Web上には大量の動画が存在**
 - Youtube
 - ニコニコ動画
- **見たい動画を探すためにはどうすれば良い？**
 - 現状ではテキストベーストな検索手法
 - 動画を完全に特定することは困難
- **コンテンツベーストなアプリケーションの必要性**
 - 映像中における動作認識を行うことは極めて有意義
 - Web動画における動作認識を行った研究は少ない
- **Web動画分類の困難さ**
 - 背景ノイズ, 低解像度, カメラモーション
 - 新しい時空間特徴抽出手法を提案
 - 特徴を統合することによる精度向上

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**
 - SURFによる特徴点の検出とその点の動き
 - 局所的な特徴を記述
 - **特徴統合による動作認識フレームワークの提案**
 - **時空間特徴，視覚特徴，動き特徴**
 - 時空間特徴は局所的な特徴
 - 動き特徴は全体的な特徴を記述
 - 視覚特徴は背景からも特徴を抽出

本研究では動画をシーンチェンジの境界で分割したショットを単位として利用

特徴統合による動作認識





関連研究

■ Web動画における動作認識の研究

- CinbisらはWeb上から動作を自動学習する手法を提案[Cinbis et al CVPR2009]
 - 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 ICCV2009]
 - 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

動作認識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
 - 特徴統合による動作認識手法の提案
- **評価実験**
 - データセット
 - 動作認識に関する実験
- 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
- カメラモーションを含む
- 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
- カメラモ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)

Step1

- カメラモーション検出

Step2

- 視覚特徴抽出

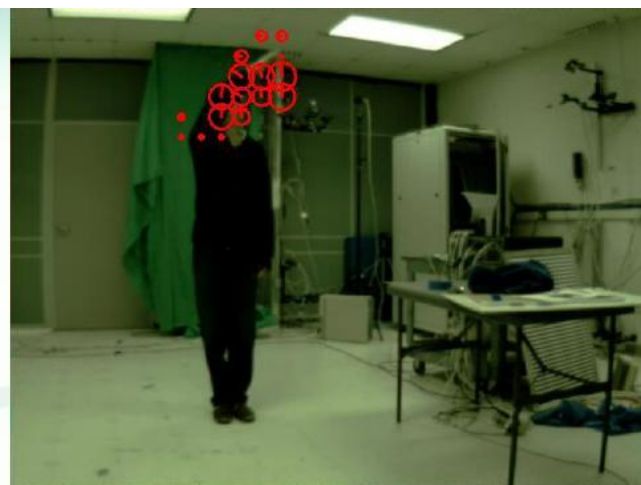
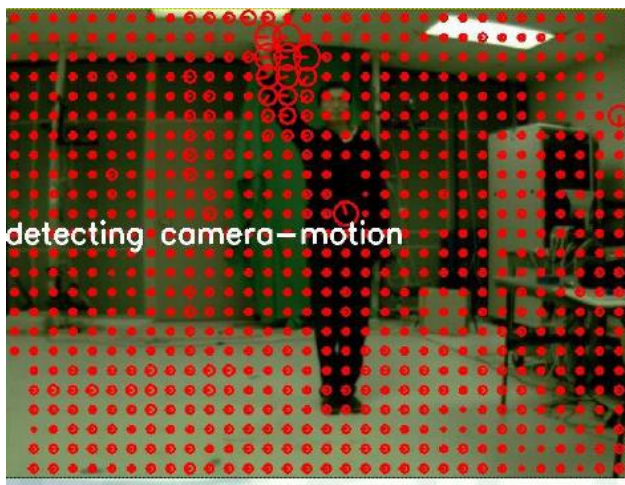
Step3

- 動き特徴抽出

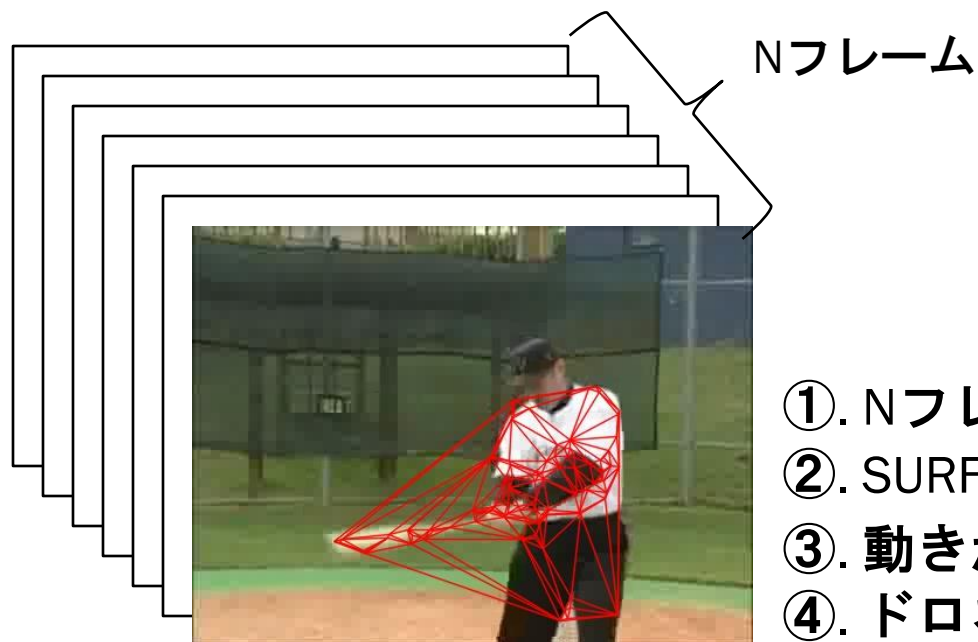
(Step 1) カメラモーション検出

■ グリッドで動きを計算

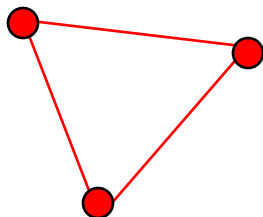
- 動いていた領域が一定割合以上ならカメラモーション
- カメラモーションが検出されたフレームは破棄
- Web動画における正確なカメラモーション検出は困難
 - 低い解像度
 - 手振れなどの意図しないカメラモーション



(Step 2) 視覚特徴抽出

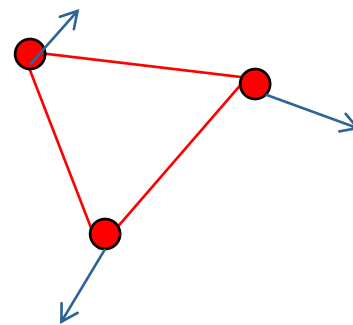
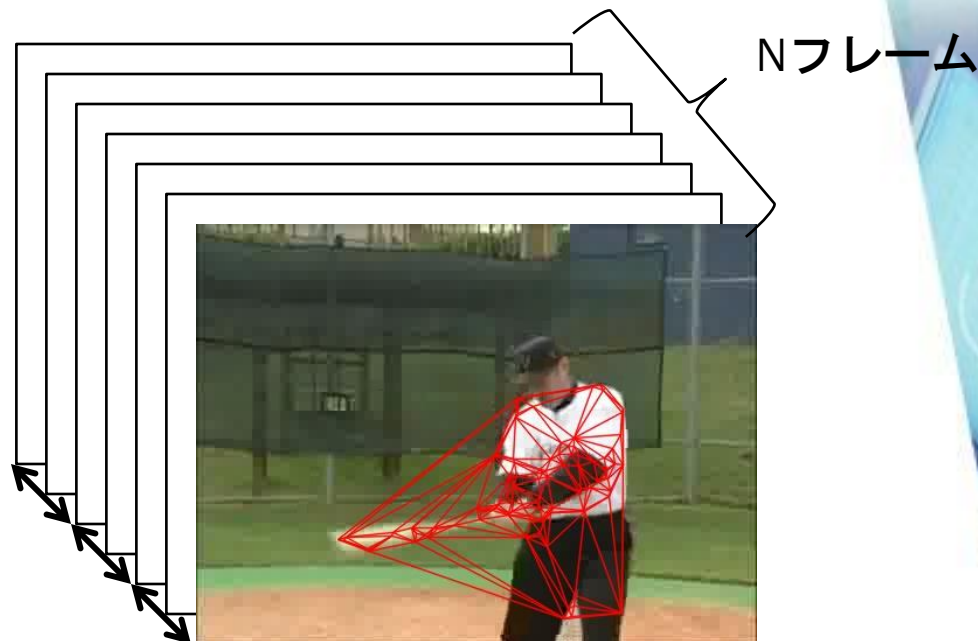


- ①. Nフレームを1ユニットとする
- ②. SURFを抽出
- ③. 動きがない点を削除
- ④. ドロネー分割
以降三点で一組の特徴と考える



Step 3 動き特徴抽出

- ① 選ばれたNフレームをM分割する
- ② それぞれの区間における時空間特徴点の動き情報を計算
- ③ 視覚特徴と同様に動きも三点の組として考える
- ④ 動き特徴としてそれぞれの点の動きと三角形の面積の変化を利用
- ⑤ 視覚特徴と動き特徴を結合
- ⑥ 抽出した特徴をベクトル量子化



特徴統合による動作認識

- 動作によって重要な特徴は異なる



Jogging



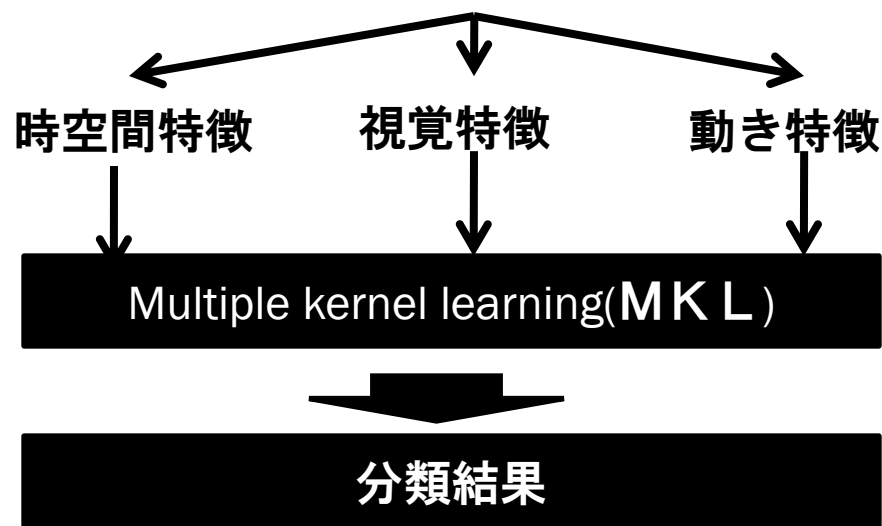
Running



Boxing

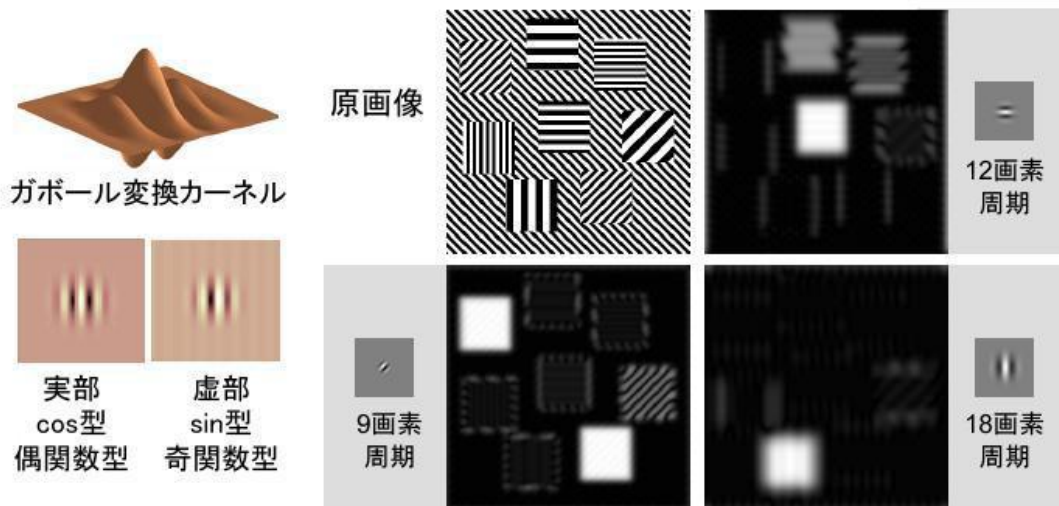


clapping



特徴統合による動作認識（視覚特徴）

- **ガボールフィルタネルを用い、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
 - すべてのフレームから特徴を抽出
 - 抽出した特徴はベクトル量子化される





特徴統合による動作認識(動き特徴)

- 抽出前にカメラモーションを検出
- 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
 - 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





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)

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

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

- 画像認識でよく使われている

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
 - 特徴統合による動作認識手法の提案
- **評価実験**
 - データセット
 - 動作認識に関する実験
- **おわりに**
 - 考察, 今後の課題



評価実験

■ KTHデータセット

- 最も広く使われているデータセット
- Leave-one-outで学習・分類

■ Our Youtubeデータセット

- 本研究で独自に構築
- 5-fold cross validationで学習・分類

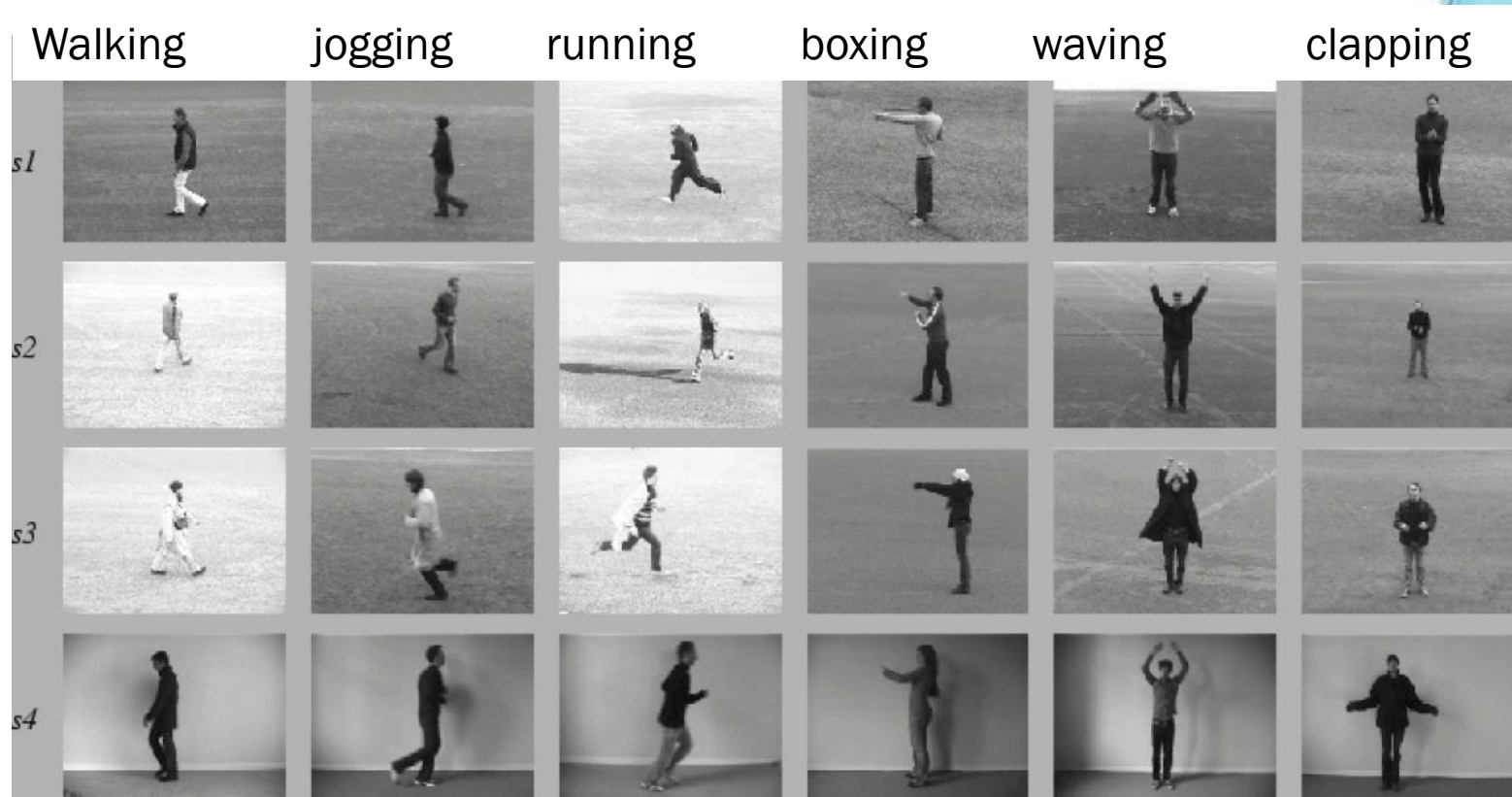
■ Wild Youtubeデータセット

- Liuらによって構築
- 5-fold cross validationで学習・分類

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(動作認識)

■ Our Youtube dataset

- 本研究で独自に構築したデータセット
- 6種類の動作, 合計600の動作を含む

batting



walking



running



shoot



jumping



eating



データセット(動作認識)

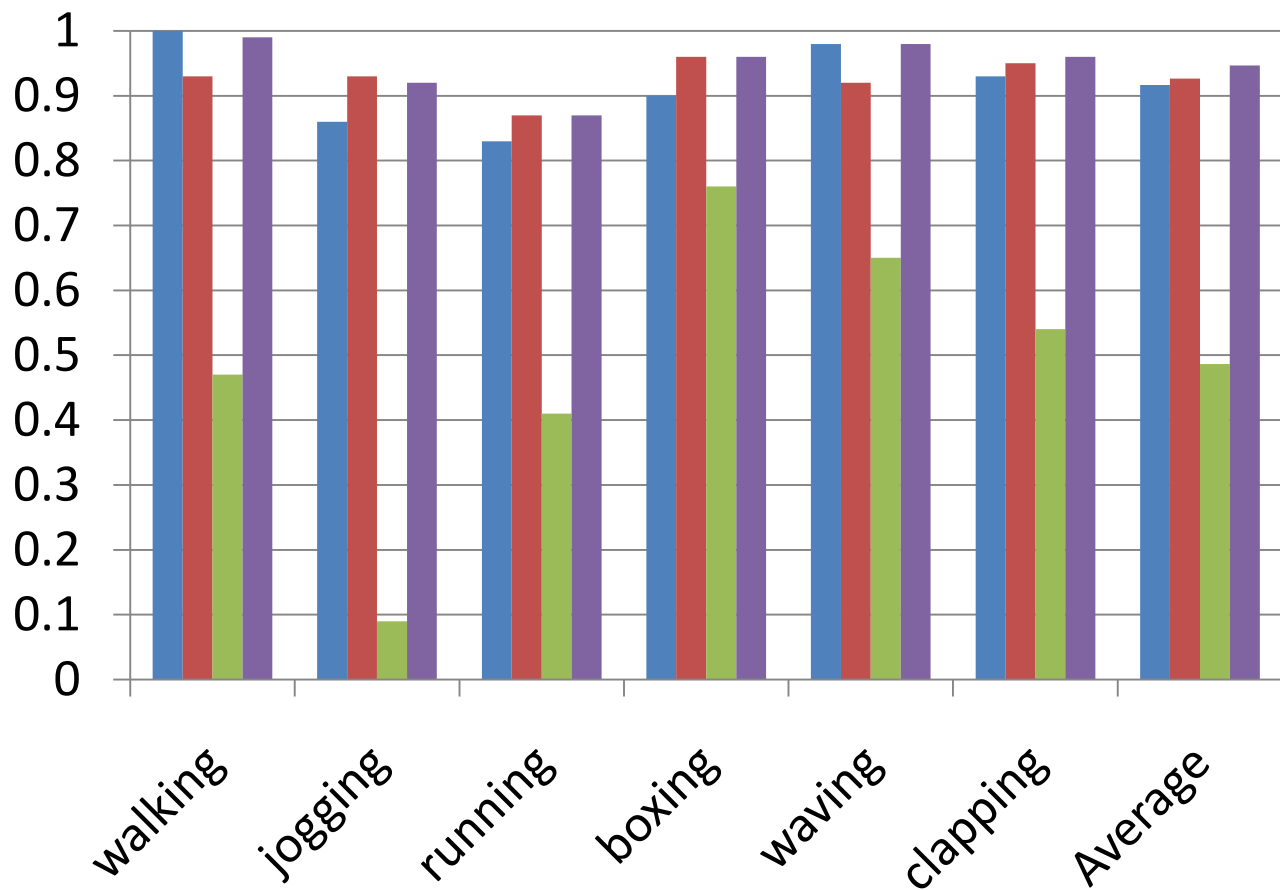
■ Wild Youtubeデータセット

- Liuらによって構築されたデータセット[Liu et al 2009]
- 11の動作, 合計1595のショットを含む



実験結果

■ KTHデータセット



時空間特徴	91.7%
動き特徴	92.7%
視覚特徴	48.7%
MKL	94.7%

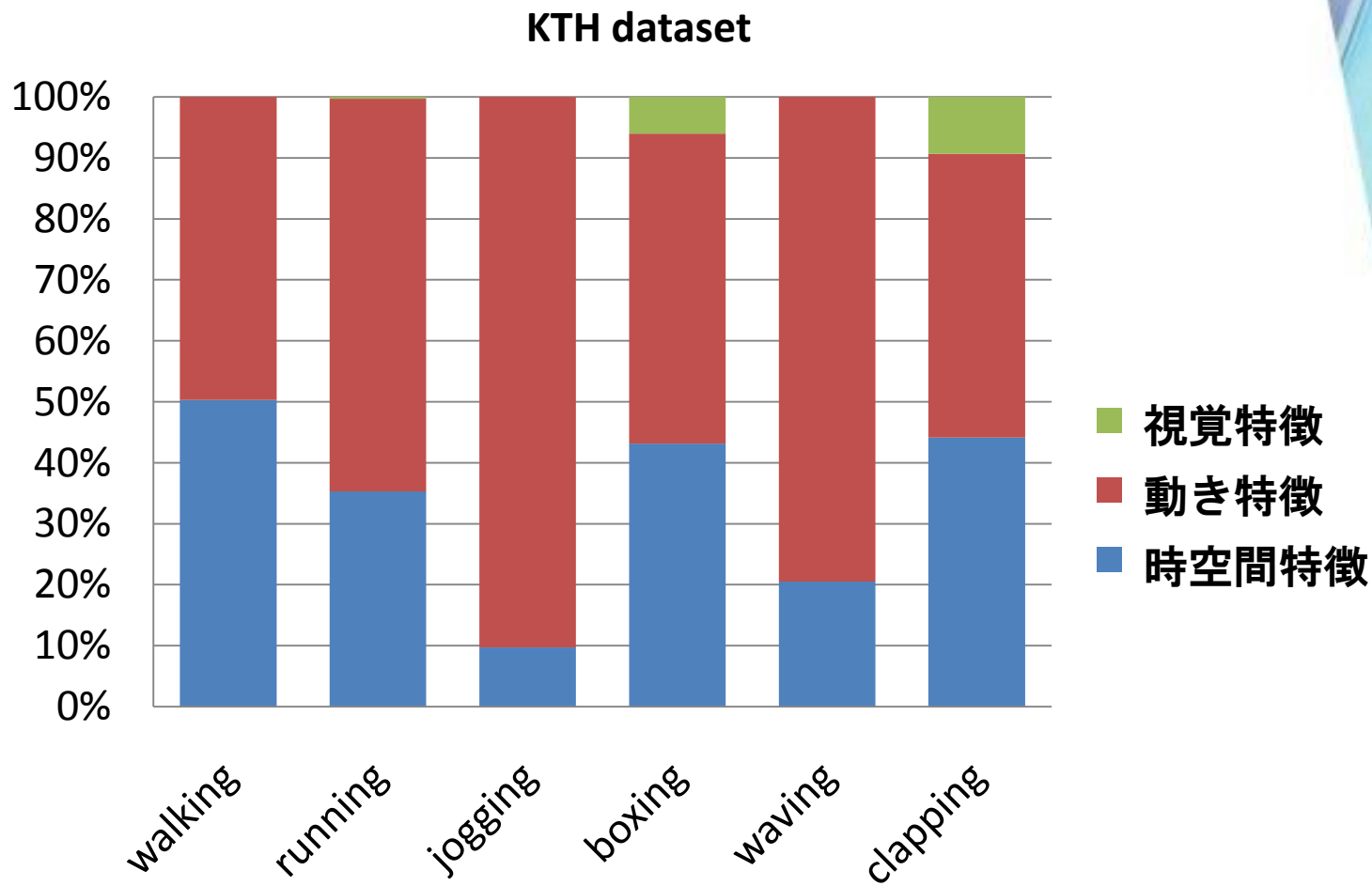
実験結果

■ 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



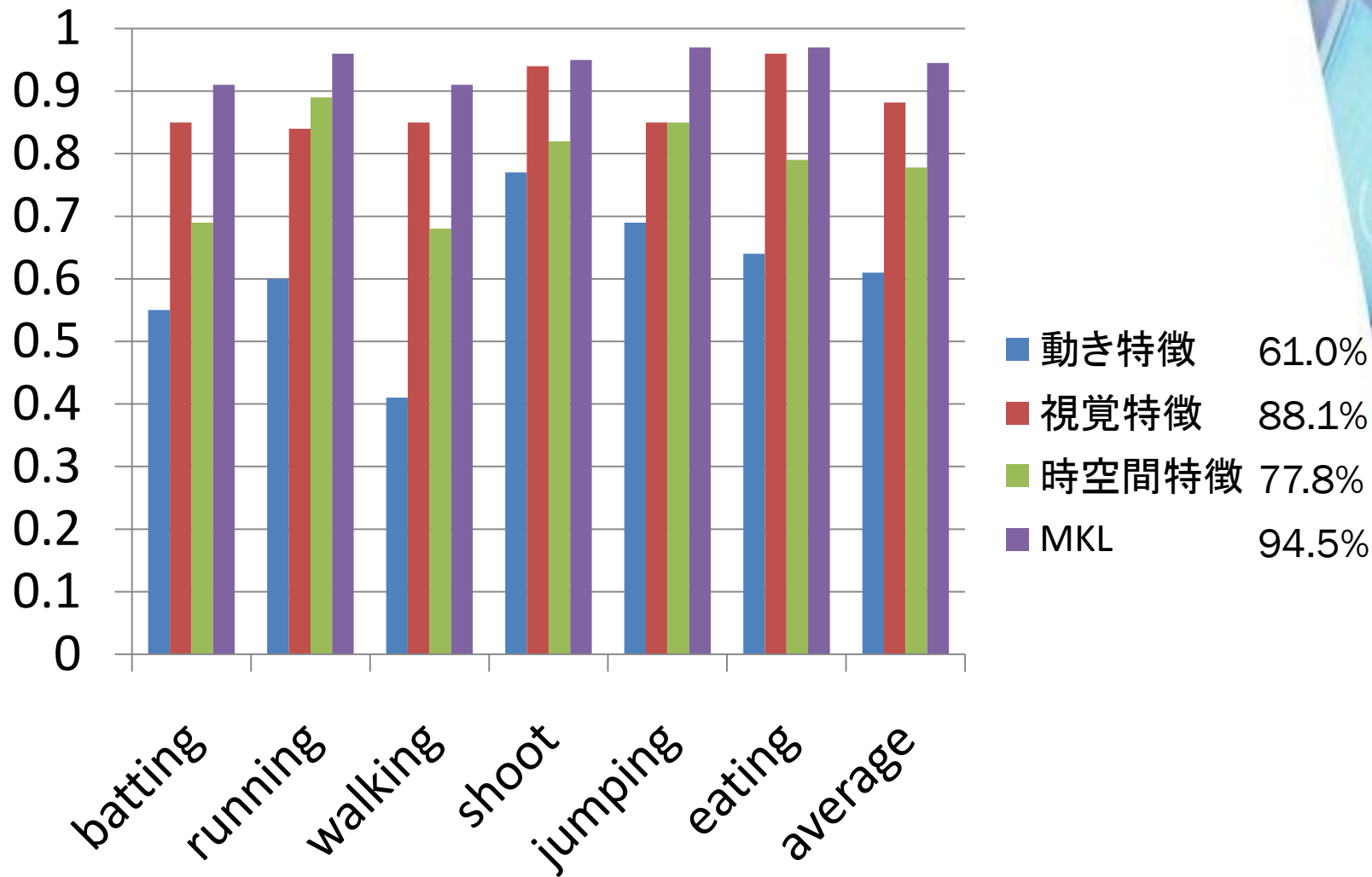
実験結果

■ KTHデータセット

分類結果 (Leave-one-out)	
Ours	94.7%
Liu et al. [CVPR 2009]	91.8%
Gilbert et al. [ICCV2009]	96.2%
Dollar et al. [2005]	81.2%

実験結果

■ Our Youtube dataset



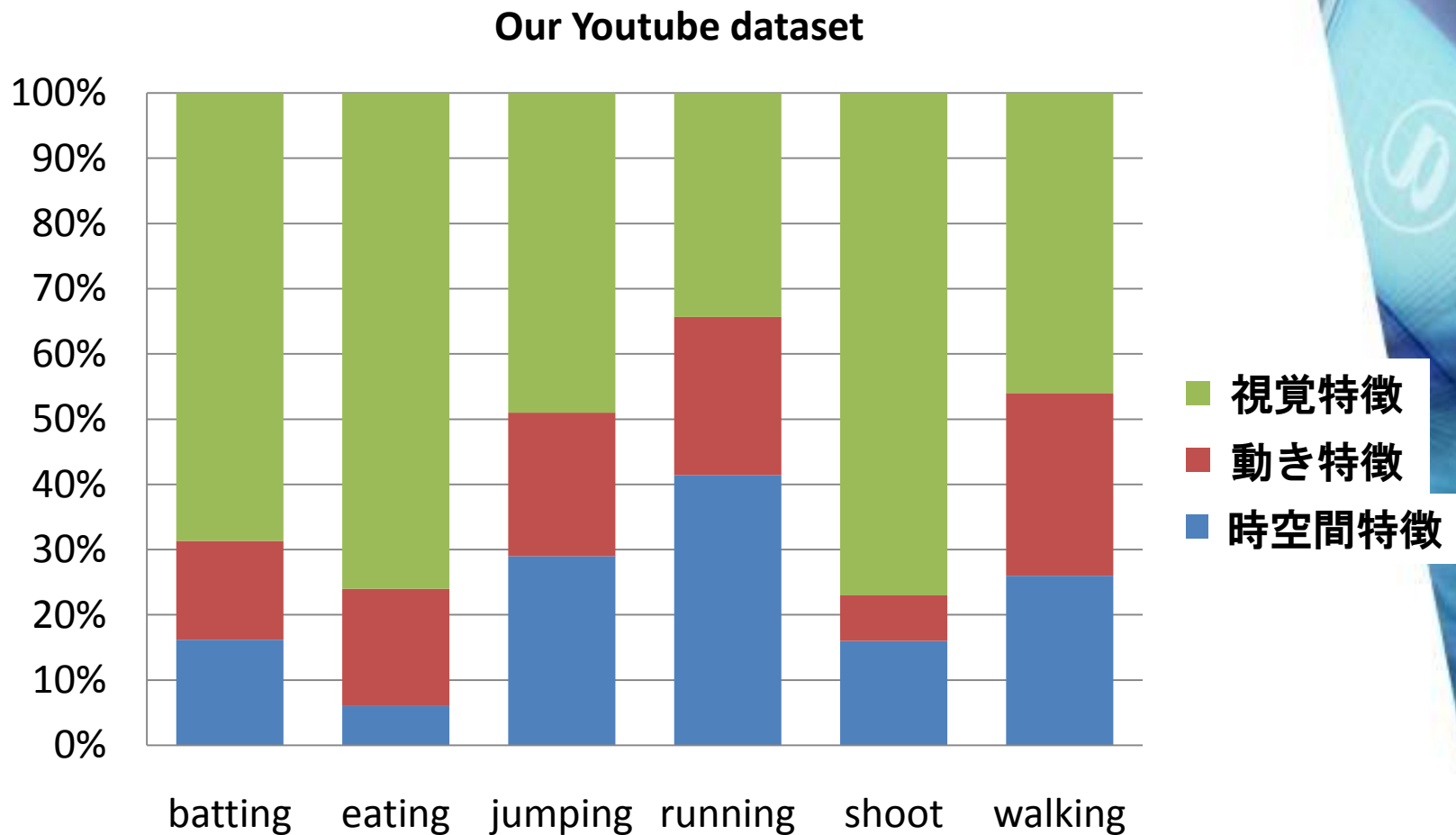
実験結果

■ Our Youtubeデータセット

	batting	running	walking	shoot	jumping	eating
batting	0.91	0	0.03	0.02	0.02	0.02
running	0	0.96	0.04	0	0	0
walking	0.02	0.04	0.91	0.01	0	0.02
shoot	0.02	0.02	0	0.95	0.01	0
jumping	0.01	0.01	0	0	0.97	0.01
eating	0.01	0	0.01	0	0.01	0.97

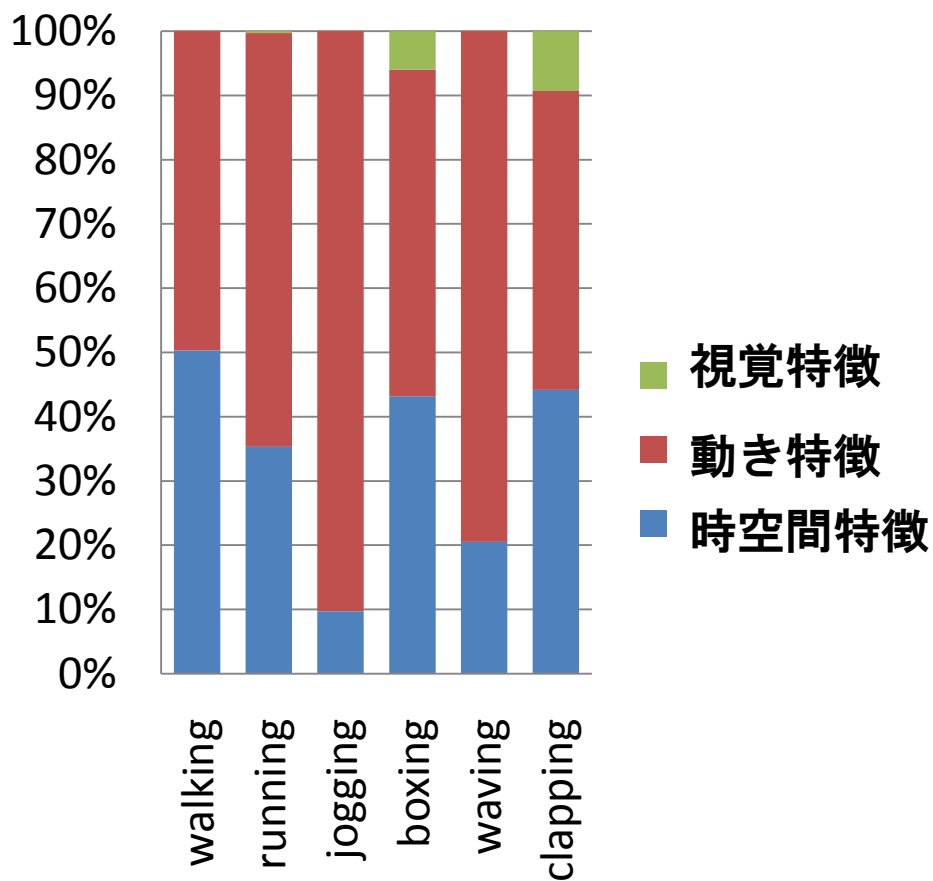
実験結果

■ Our Youtube dataset

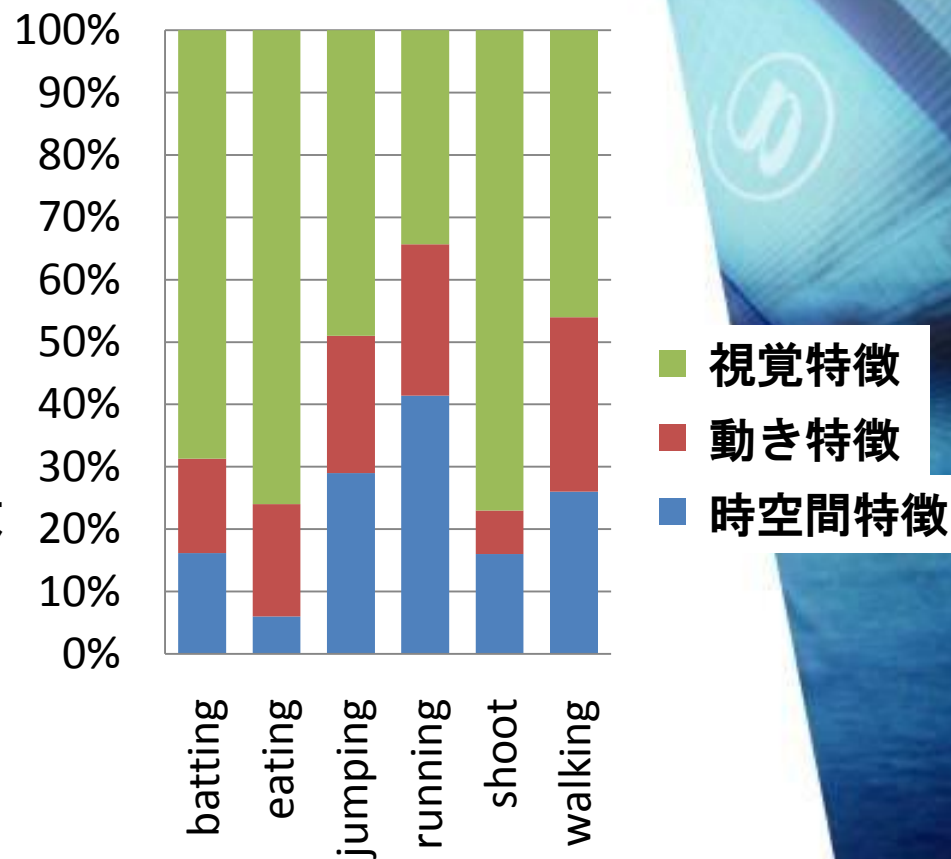


KTHとYoutubeデータの重みの比較

KTH dataset

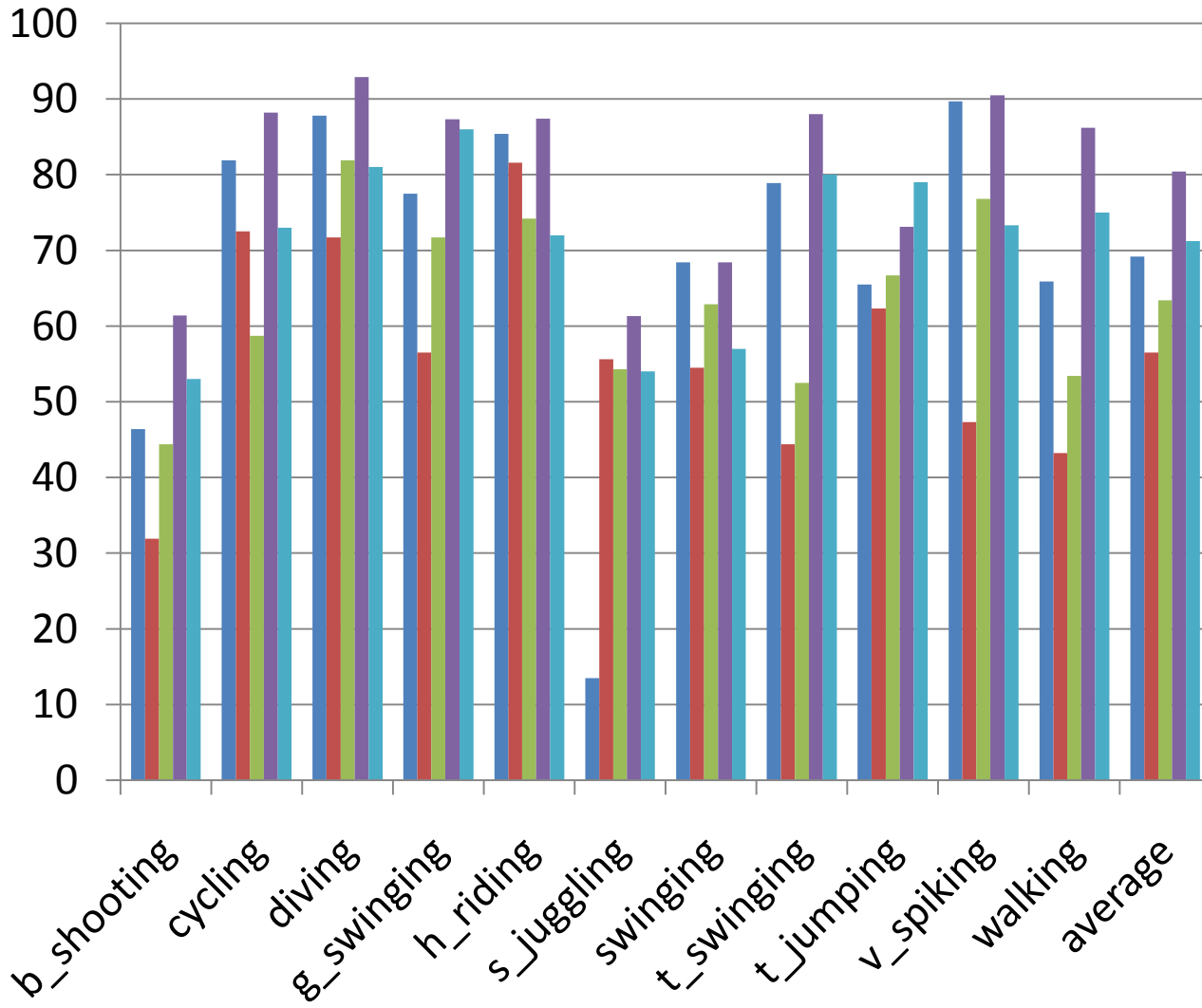


Our Youtube dataset



実験結果

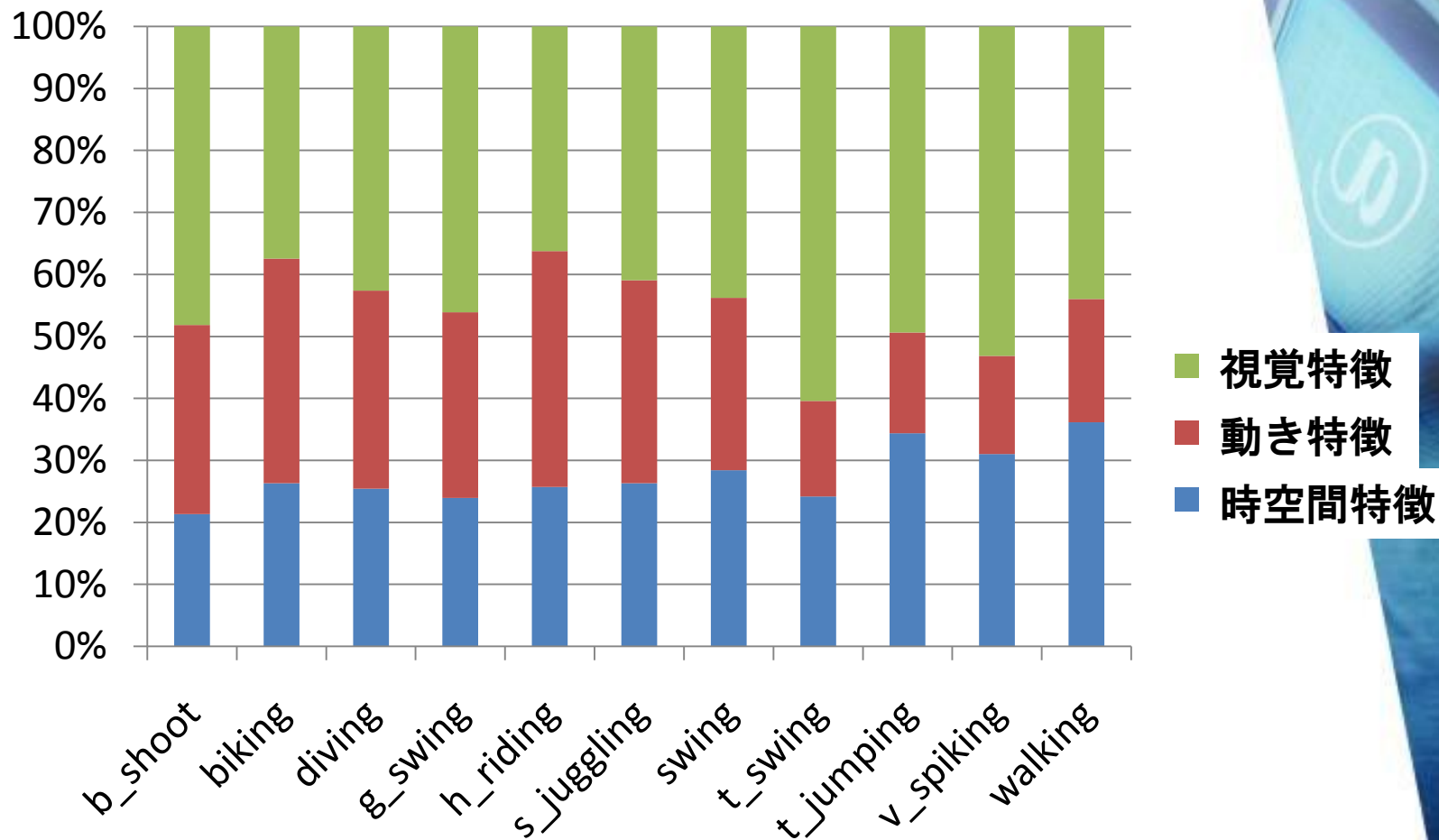
■ 動作認識 (Wild Youtube dataset)



視覚特徴	69.1%
動き特徴	56.5%
時空間特徴	63.4%
MKL	80.4%
Liu [CVPR2009]	71.2%

実験結果

Youtube dataet





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
 - 特徴統合による動作認識手法の提案
- **評価実験**
 - データセット
 - 動作認識に関する実験
- **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
 - Youtubeデータセットにおいて最新手法を上回る結果

■ 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
- 有益な特徴の選択を行う

ご清聴ありがとうございました