IS2-68

動きの連続性を考慮した動画からの局所的な時空間特徴の抽出

電気通信大学大学院　情報工学専攻　野口 顕嗣　柳井 啓司

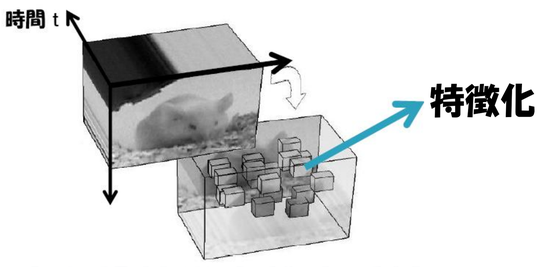
はじめに

◆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
- ユーザのニーズに完全に応えることは困難

◆映像を解析するアプリケーションの必要性
- 時空間特徴の抽出は現在活発に研究が行われている
- 大量のデータを調べるためには高速化が必要不可欠

従来手法



◆時空間特徴点検出手法
- 3D Harris corner検出手法(Laptev et al. 2008)
- 2D Gaussian filter と 1D Gabor filterによる検出手法(Dollar et al. 2005)

◆問題点
- Cuboid全体から特徴を抽出するのは計算コストが高く大規模な処理には向いていない
- Cuboid全部を特徴化する必要はあるのか？

そこで

◆本研究では
- 計算コストを低くしたい
- 特徴的な点と，その点の局所的な動きによって映像を記述する
- その高速性を利用して大量の映像に対して特徴を抽出していく

提案手法

◆提案手法の流れ

視覚特徴抽出
- SURFを抽出

追跡点の決定
- 視覚特徴の座標における動き特徴を計算

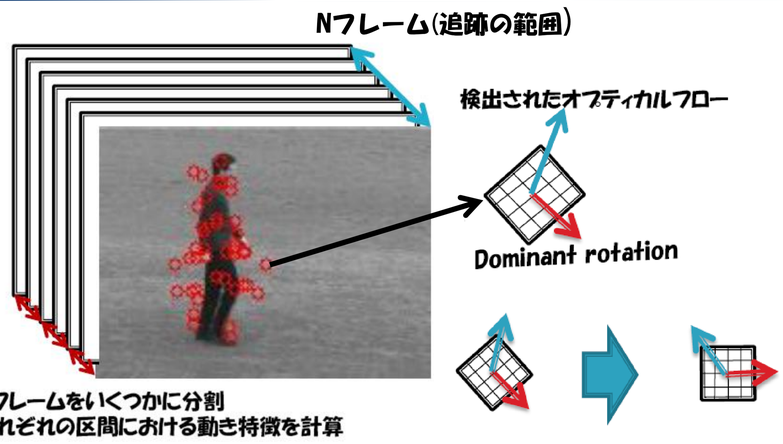
動き特徴の抽出
- 局所的な追跡

ベクトル化
- 視覚特徴と動き特徴を重みを付けて線形結合

◆視覚特徴抽出
- SURFによって特徴を抽出
  右は実際にSURFで抽出された点を示している

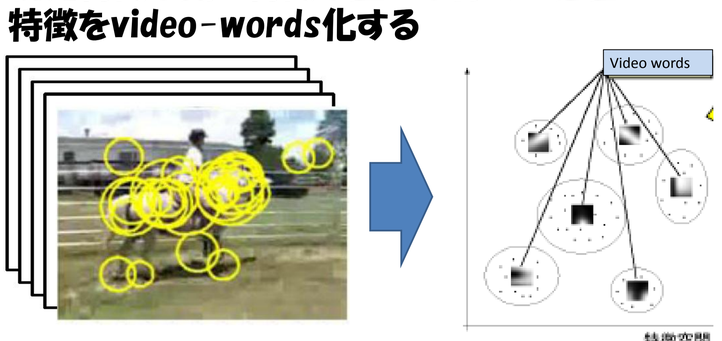
◆追跡点の決定
- 抽出された点に関してLucas-Kanadeアルゴリズムで動き情報を計算する
- 動きのあった点を時空間特徴とする
  以降この点に関して特徴を抽出する

◆動き特徴抽出部

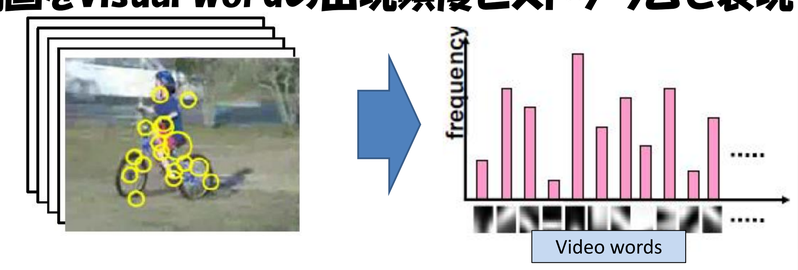


候補点に対して局所的な追跡を行う
① フレームからN先のフレームを抽出このNが追跡の範囲となる
② 抽出されたNフレームをいくつかに分割しそれぞれの区間で動き情報を計算する
③ 回転に関して頑健にするために
視覚特徴の回転に沿って，動き情報も回転させる
抽出された特徴はvideo-words化して利用

◆ベクトル化
視覚特徴と動き特徴を重みを付けて結合
特徴をvideo-words化する



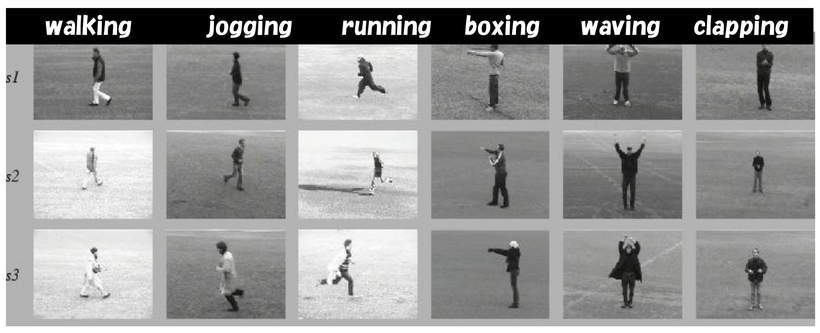
動画をvisual wordの出現頻度ヒストグラムで表現



動画をvisual wordsの集合とみなす

実験

◆人間の動作分類
分類はSVMを使用する
データセット
- KTHデータセットを使用
- 6つの動作があるデータセット
- 一つの動作につき100のデータが存在
- 5-fold cross validation によって学習・分類を行う
- 学習，分類にはSVMを利用する



分類に用いた特徴
①動き＋視覚＋回転(VMR)　②動き＋視覚(VM)
③動きのみ(M)　④視覚のみ(V)

分類結果

表 1　視覚特徴＋動き特徴＋回転あり

	walking	running	jogging	boxing	waving	clapping
walking	0.91	0.01	0.06	0.01	0.01	0
running	0.02	0.77	0.21	0	0	0
jogging	0.03	0.15	0.82	0	0	0
boxing	0	0	0	0.92	0.02	0.06
waving	0	0	0	0.05	0.87	0.08
clapping	0	0	0	0.07	0.06	0.86

表 2　視覚特徴のみ

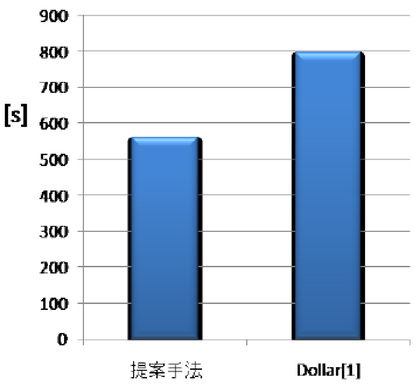
	walking	running	jogging	boxing	waving	clapping
walking	0.7	0.13	0.16	0.01	0	0
running	0.1	0.59	0.21	0	0	0
jogging	0.12	0.29	0.58	0	0	0.01
boxing	0.13	0.13	0.1	0.6	0.03	0.01
waving	0.03	0.09	0.01	0.05	0.74	0.08
clapping	0.04	0.05	0.02	0.06	0.25	0.58

表 3　動き特徴のみ

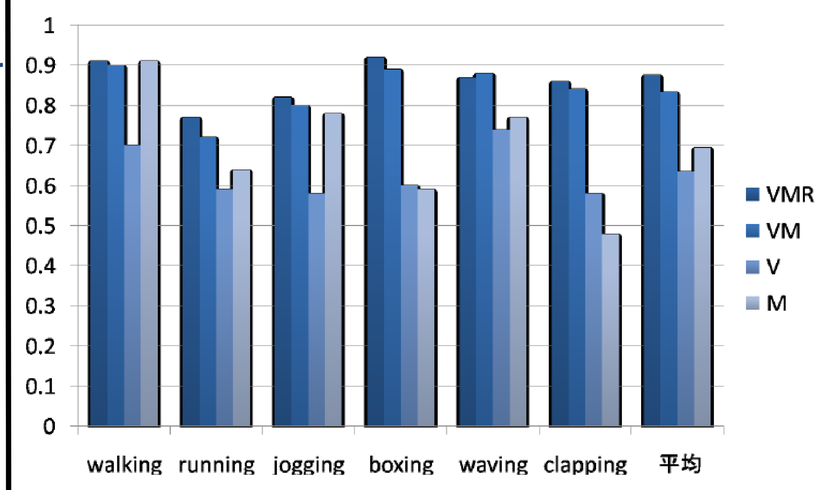
	walking	running	jogging	boxing	waving	clapping
walking	0.91	0	0.06	0.03	0	0
running	0	0.64	0.3	0	0.02	0.04
jogging	0.04	0.13	0.78	0.02	0.03	0
boxing	0.01	0	0	0.59	0.32	0.08
waving	0	0	0.01	0.17	0.77	0.05
clapping	0	0	0	0.18	0.33	0.48

表 4　視覚特徴＋動き特徴＋回転無し

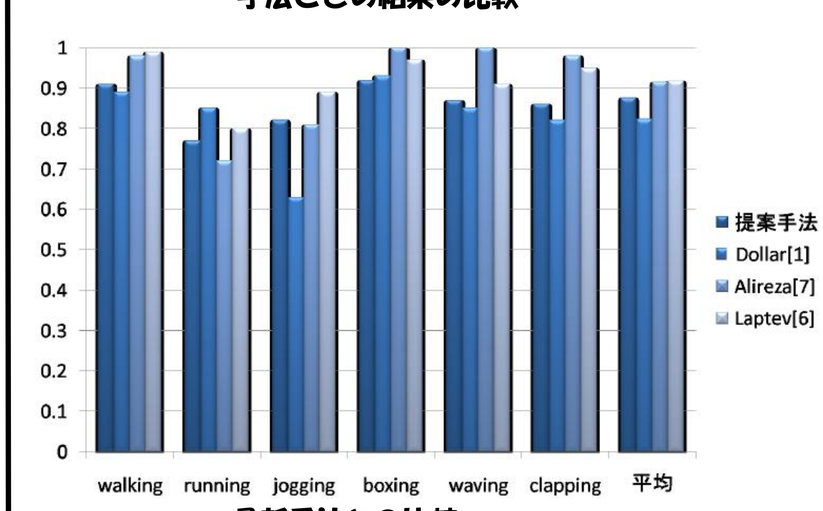
	walking	running	jogging	boxing	waving	clapping
walking	0.9	0.01	0.07	0.01	0	0
running	0.01	0.72	0.27	0	0	0
jogging	0.01	0.18	0.8	0.01	0	0
boxing	0	0	0	0.89	0	0.11
waving	0	0	0	0.06	0.88	0.06
clapping	0	0	0	0.13	0.02	0.84



計算時間の比較



手法ごとの結果の比較



最新手法との比較

◆Web動画のクラスタリング
実験手順
- Youtubeからサッカーの動画を収集
- 収集した動画をショット分割する
- 各ショットから本特徴を抽出
- bag-of-video-words化した後K-meansでクラスタリング

動画１本に対するショットクラスタリング結果の一部(k=8)

動画100本に対するショットクラスタリング結果の一部(k=50)

◆考察
- 動作分類から，最新の手法と比べると精度は劣るが高速化に成功している
- 回転を考慮に入れることで精度向上に貢献することを確かめた
- カメラモーションがあった場合，大量の特徴が抽出されてしまい，計算に時間がかかることがあった

おわりに

◆まとめ
- 時空間特徴抽出の高速化をはかる手法を提案
- 特徴点と，その点の追跡に基づく手法
- 最新の手法と比較して精度は若干劣るが，高速化することができた
- カメラモーションを含む動画に対して時間がかかる

◆これからの課題
- 時空間特徴抽出手法の改良
  - カメラモーション検出の追加
  - 精度と速度向上への検討
- 大量のデータにおけるWeb動画のマイニングを行う