マルチフレーム認識を用いた動画像認識の分析

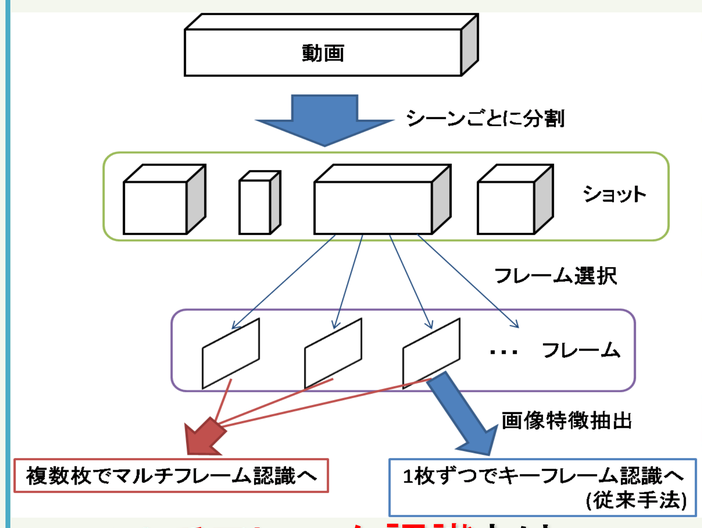
電気通信大学大学院　情報理工学研究科 総合情報学専攻　樋爪和也　柳井啓司

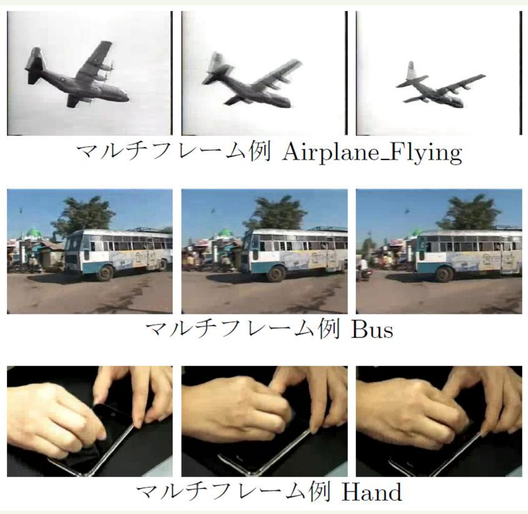
1. 背景と目的

◆コンテンツベースの動画認識研究が盛んに行われている
- ➢TRECVID（映像認識技術促進のワークショップ）の存在
  - – 動画中の物体や動作の認識、クラス分類を行う
  - –Web動画データセットを使用、全130クラス
- ➢動画の認識手法としてマルチフレーム認識が登場

◆フレームの選択枚数および選択方法を変化させ、マルチフレーム認識の有効性を検証

2.システム概要



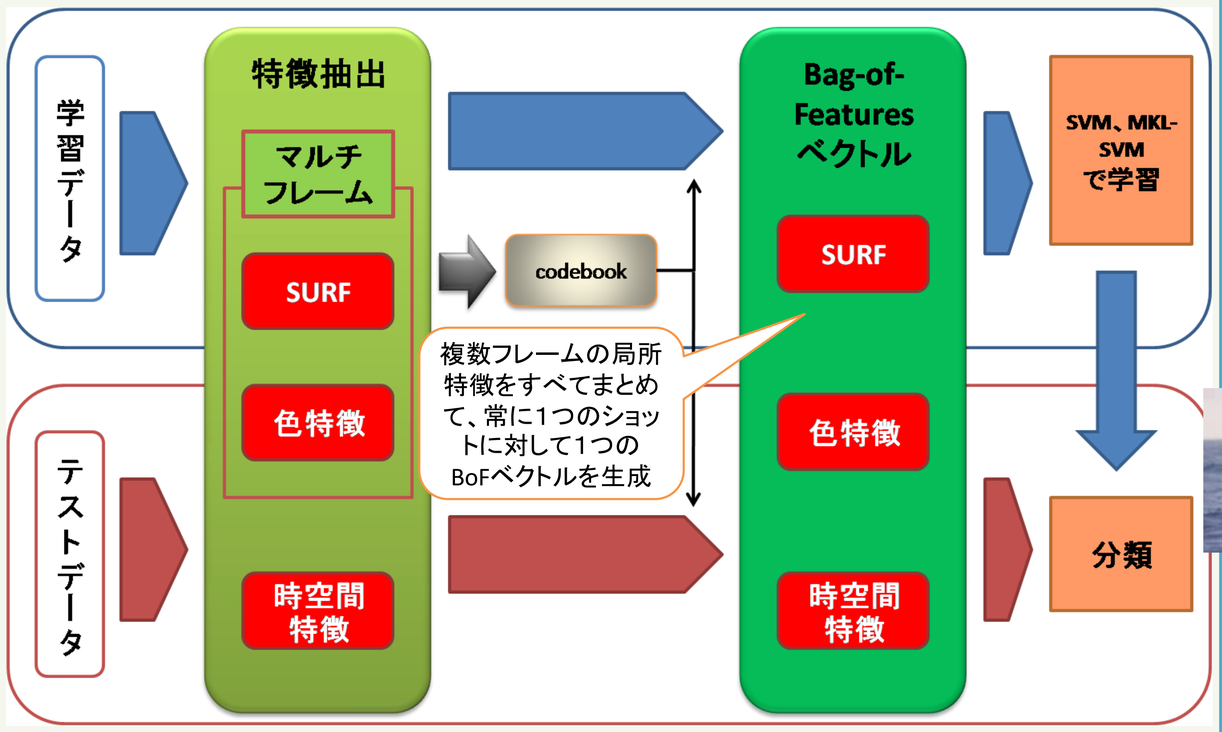


マルチフレーム認識とは
- ◆ 一つのショットから複数のフレームを選択し、特徴を抽出
  - ➢ 取得できる特徴量が増加
  - ➢ 複数のアングルから特徴量を検出可能
- ◆ 従来はシングルフレーム認識が主流

◆ 3種類のフレーム取得方法を採用
- a.ショットからN枚のフレーム (N=3,5,10)
- b.ショットからMフレーム置き (M=30,15,10)
- c.ショットのすべてのフレーム

Mフレーム置き　平均選択枚数

	M=30	M=15	M=10
Airplane_Flying	4.0	7.6	11.1
Boat_Ship	5.4	10.6	15.8
Bus	19.1	37.7	56.6
City_scape	4.8	9.2	13.6
Classrom	14.1	28.0	41.9
テストデータ	5.0	9.5	14.0



1. ショットからSURF、色特徴、時空間特徴を抽出
   - ➢ SURF、色特徴は複数のフレームから抽出
2. 学習データからcodebookを作成し、各特徴量をBag-of-Features表現に変換
   - ➢ codebookは各500次元
   - ➢ SURF、色特徴に2*2の空間ピラミッドマッチングを適用
     ⇒ 全4500次元
3. SVMまたはMKL-SVMで学習
   - ➢ RBF-χ^2カーネルを使用
4. テストデータからも同様に特徴抽出・BoF変換をし、SVMまたはMKL-SVMで分類
   - ➢ 出力値を使用してショットをランキング付け

3.実験

◆TRECVID2010 Semantic Indexingのルールで実験
- ➢ 5種類の概念を使用
  - – 学習データ：12,966ショット
  - – テストデータ：144,988ショット
- ➢ 推定平均適合率(infAP)で評価
  - – ランキング上位2000枚が対象
- ➢ シングルフレーム認識（従来手法）と比較

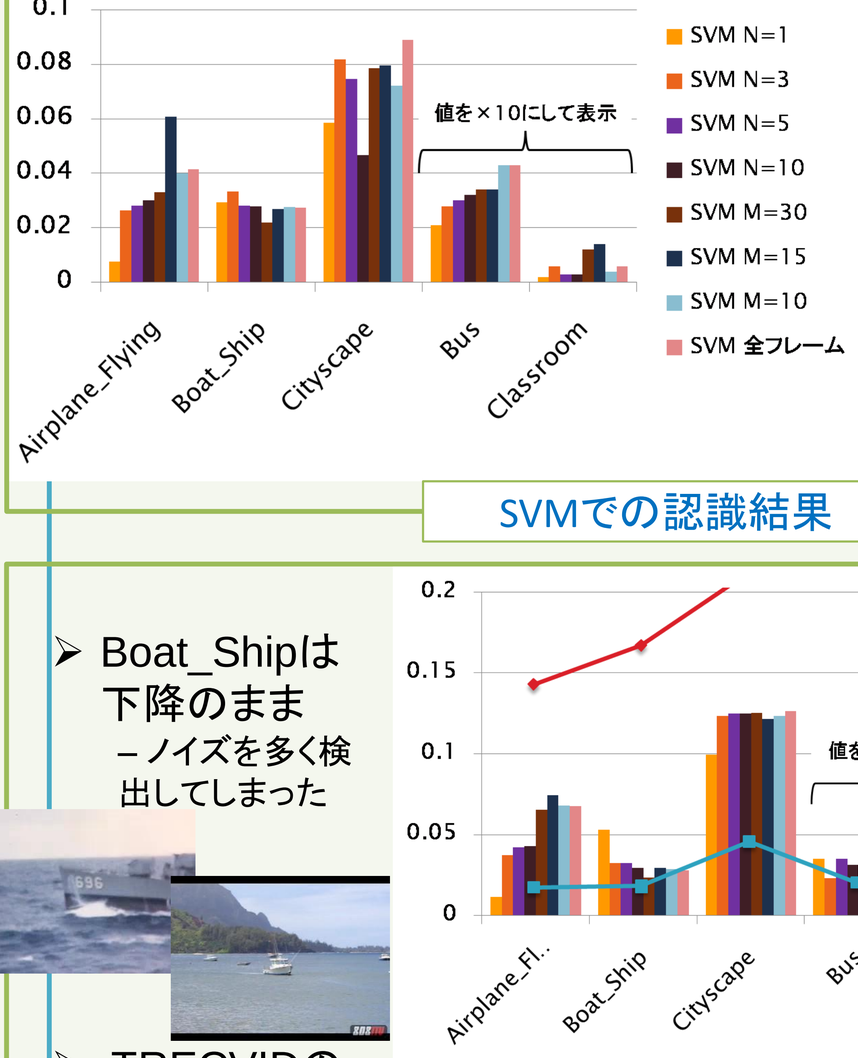
MKL-SVMでの認識結果

全フレームより少ない枚数で良い結果も

		N枚抽出			Mフレーム置きに抽出			
	シングルフレーム (N=1)	N=3	N=5	N=10	M=30	M=15	M=10	全フレーム
Airplane_Fying	.0114	.0373	.0420	.0429	.0654	**.0742**	.0678	.0675
Boat_Ship	**.0528**	.0324	.0322	.0294	.0231	.0291	.0286	.0276
Bus	.0035	.0023	.0035	.0031	.0024	.0029	**.0036**	.0030
City-Scape	.0996	.1236	.1250	.1251	.1255	.1216	.1235	**.1264**
Class room	.0006	.0017	.0027	.0031	.0048	**.0059**	.0029	.0032

枚数が増えると値も上昇

最大883%上昇

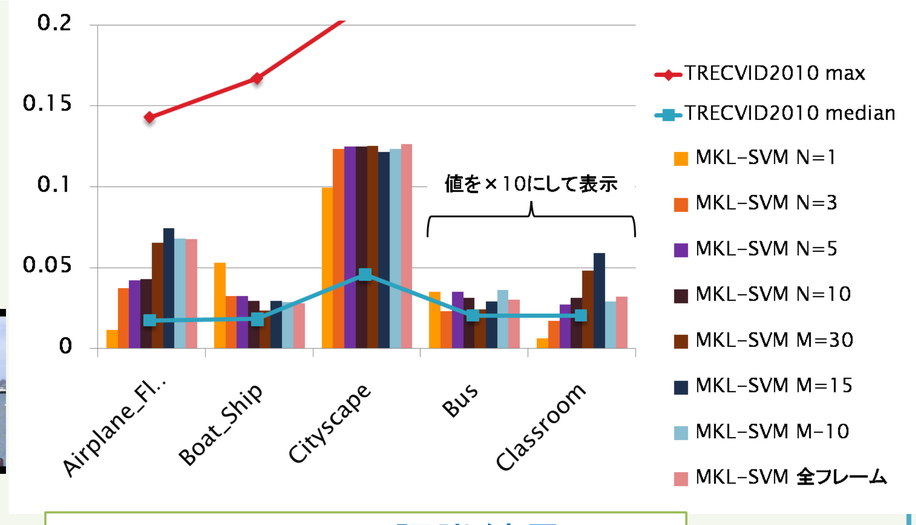


- ➢「Mフレーム置き」の方がよい結果
  - – 長いショットではより多くのフレームを取得できる
- ➢ 枚数に比例しない例もある
  - – 枚数以外の要因も考えられる

SVMでの認識結果

- ➢ Boat_Shipは下降のまま
  - – ノイズを多く検出してしまった

- ➢ TRECVIDの中央値を上回る



MKL-SVMでの認識結果、TRECVID全チーム結果との比較

4. まとめ、今後の課題

◆ キーフレームのみの認識を上回る結果を得られた
- ➢ ショット認識におけるマルチフレームの有効性を確認
- ➢ 全フレーム使用に比べると、クラスによっては中間フレームを使わないことでノイズの検出を回避できている

◆ フレーム枚数を増やしても、精度が悪くなる場合がある
- ➢ より有用なフレームの選択
  - – 既に抽出した前のフレームとの差異を計算
  - – フレームの変化によって取得間隔をショット毎に変更