



OS5-O2

TRECVID Semantic Indexing Taskと Multimedia Event Detection Taskへの 取り組み

樋爪 和也，柳井 啓司（電気通信大学）

背景

- コンテンツベースの映像検索手法が盛んに
 - TRECVID : 国際的な評価ワークショップ
- 膨大なデータ処理の必要性
 - 4000時間の大規模データ



本発表について

- TRECVIDの概要
 - SINタスク
 - MEDタスク
- 参加タスクに対する実行手法

TRECVIDとは

映像検索技術に関連する競争型ワークショップ

- NISTとDTOが主催
- 東工大、NTT、IBM、コロンビア大、アムステルダム大など各国の研究チームが参加

TRECVID2012の6つのタスク

-  ➤ Semantic indexing (SIN)
- Known-item search (KIS)
- Interactive surveillance event detection (SED)
- Instance search (INS)
-  ➤ Multimedia event detection (MED)
- Multimedia event recounting (MER)

本研究はSINタスク、MEDタスクに関連

Semantic INdexing task

ショット (=シーンごとの短い動画)の認識

✓ 1ショットあたり、**最長で3.5分**

✓ 主に対象となるのは

物体、人 : Chair、George_Bush、Skier...

動作 : Singing、Sitting_down、Walking__Running...

風景 : Hill、Kitchen、Forest...



Airplane



Bus



Hand

Semantic INdexing task

- カテゴリ数：346

- ✓ 20(2008, 09)、130(2010)と増加
- ✓ カテゴリ数を減らしたlight、組み合わせたpairも
 - ・ 最終結果を提出できないチームも多い

- データ

- ✓ 学習データ：403,800ショット(600時間)
- ✓ テストデータ：145,634ショット(200時間)

テストデータは翌年の学習データに

- 評価形式：平均適合率

- ✓ ランキング付けした上位結果からサンプリングして計算
- ✓ カテゴリごとに上位2000ショットが対象

2011、2012は東工大チームがTOP

Multimedia Event Detection task

複雑なイベントの判定

✓ SINタスクよりも**複合的な高次概念**を検出

動作：Repairing an appliance、Rock climbing...

状態：Birthday party、Winning a race without a vehicle...



Working on a
sewing project



Parade



Birthday party

Multimedia Event Detection task

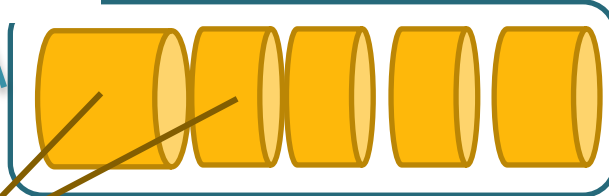
- イベント数：20+5

- ✓ 事前に告知されるイベント20、追加のAdHocイベント5
- ✓ Ad Hocイベント
 - 締め切りまでの期間が短く
 - メタデータも十分でない状態

- データ

- ✓ SINデータと異なり、ショット分割が行われていない
ビデオクリップが対象

SIN



MED



ビデオ

ショット

- ✓ 学習データ : 47430クリップ(1474時間)
- ✓ テストデータ : 98117クリップ(3722時間)

Multimedia Event Detection task

- 評価形式

- ✓ 各チームが独自にThresholdと、スコアを決定
- ✓ 正解、非正解データの**エラー率**が対象

未検出率

$$FN = \frac{\text{誤分類した正解データ数}}{\text{全正解データ数}}$$

誤検出率

$$FP = \frac{\text{誤分類した非正解データ数}}{\text{全非正解データ数}}$$

- ✓ Normalized Detection Cost(NDC)
 - MEDでのシステム全体の評価指針

$$NDC = \frac{FN * 0.08 + FP * 0.999}{0.08}$$

NDCが**低い**=FN、FPが**低い**ほど**高性能**

認識手法：SINタスク

- 特徴抽出

- ✓ 画像特徴：全フレームから

- SURF

- RGB色特徴



- ✓ 時空間特徴

- [Noguchi et al. 2010]



認識手法：SINタスク

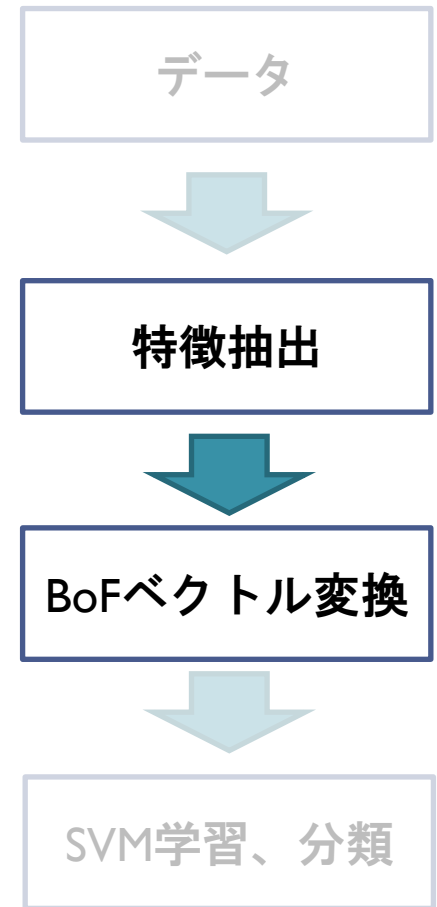
- Bag-of-Features表現に変換

- ✓ soft assignment

- 複数のコードワードへの割り当て

- ✓ 空間ピラミッド表現
(1 + 2 × 2)

- ✓ 画像特徴はコードワード1000
時空間特徴はコードワード
5000

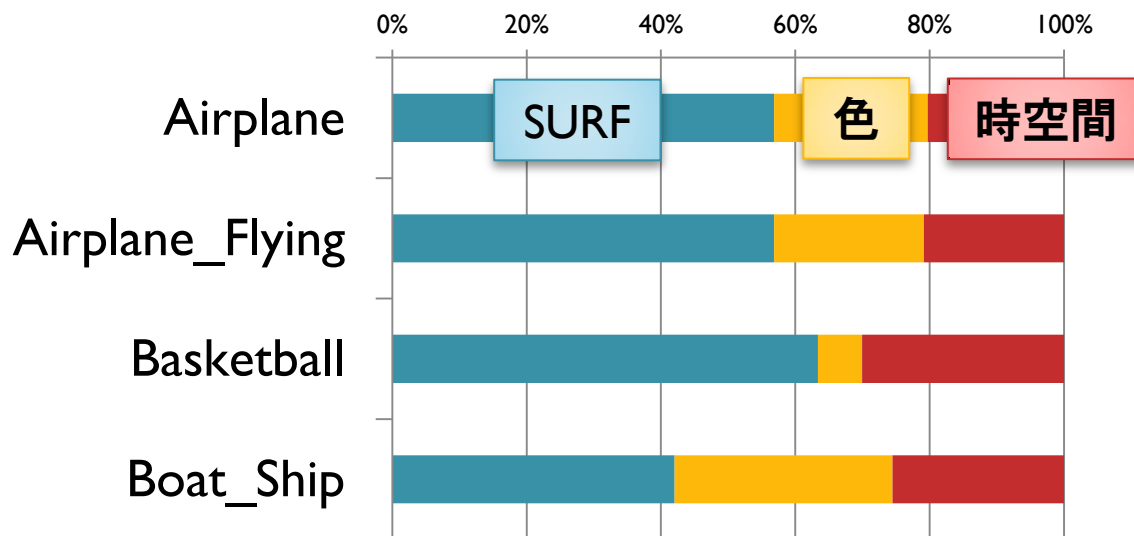


認識手法：SINタスク

- Multiple Kernel Learning-SVM

- ✓ 複数の特徴を統合
- ✓ 出力値をソートしてランキング付け

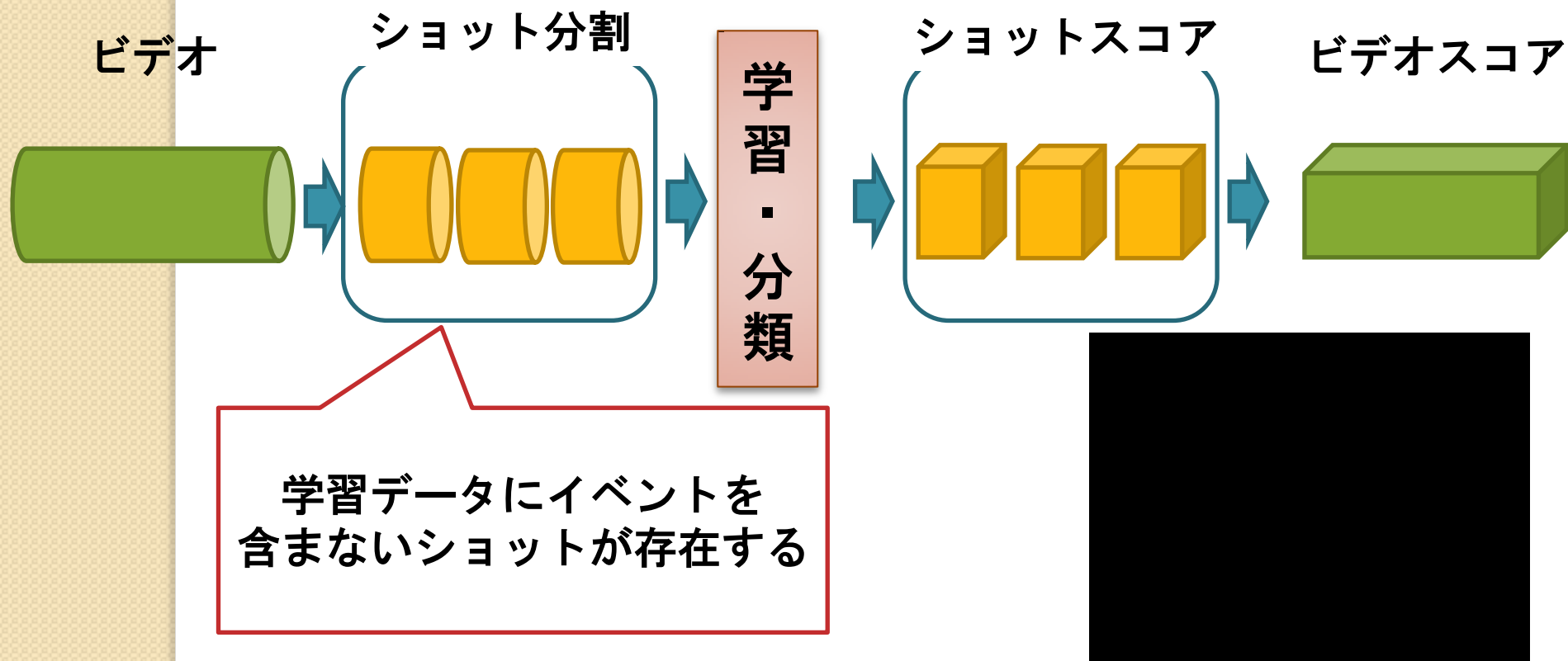
$$K(x, x') = \sum_{j=1}^K \beta_j k_j(x, x') \quad \beta_j \geq 0, \sum_{j=1}^K \beta_j = 1$$



認識手法：MEDタスク

- SINタスクの手法をMEDに適用

MEDデータはショット分割がされていない

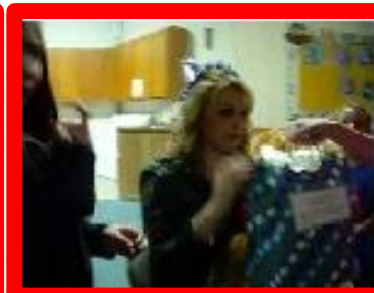


認識手法：MEDタスク

- 学習データの中からショットを選択

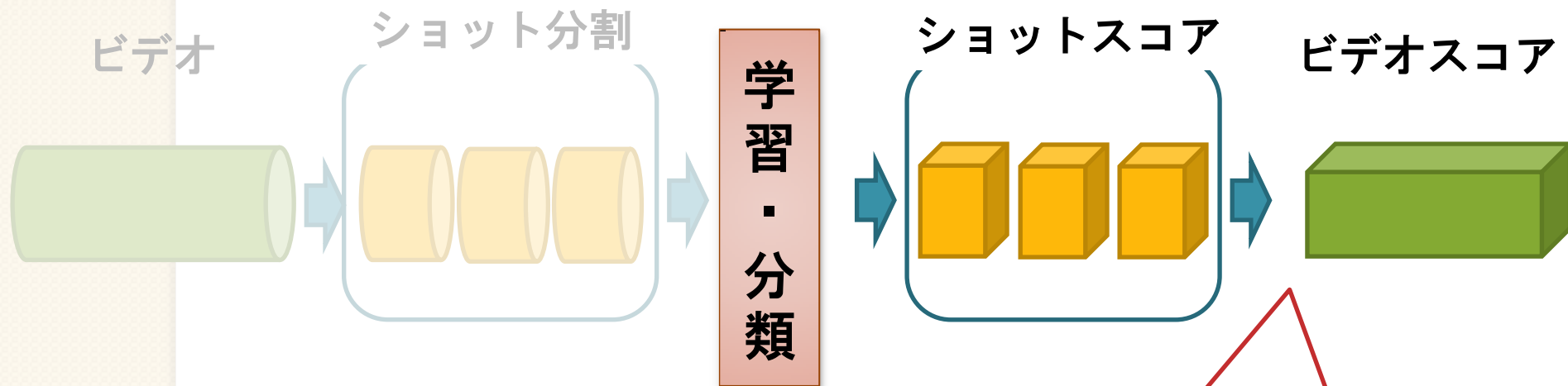
- ✓ VisualRank：教師なしショットランキング
- ✓ ランキング上位のショットを学習に使用

Birthday party



認識手法：MEDタスク

- 元のビデオクリップのスコア決定



$$\text{元の動画のスコア} = \frac{S_1 + S_2 + S_3}{3}$$

S_i : 上位 i 番目のショットスコア

認識手法：MEDタスク

- ショット分割結果

- ✓ 最大3000フレーム(100秒分)
- ✓ ランキング上位500ショットをポジティブに

	学習データ数 (25クラス)	テストデータ
元動画	4225	98118
分割後ショット数	48792	733764
使用した ポジティブデータ	12500	—

実験環境

- 80台のクラスタマシン

- ✓ 4コアCPU、8GBメモリ

- AMD : Phenom II X4 (3.00GHz)

- Intel : Xeon X3230 (2.66GHz)、Xeon X3363 (2.83GHz)、
Core2 Quad Q9650(3.00GHz)、Core i7 870 (2.93GHz)、
Core i7-2600 (3.40GHz)

- ✓ 計320コア、計算に使用したのは平均3コア／マシン

- ✓ 全行程の計算時間

- SINタスク : 約4日

- MEDタスク : 約6日

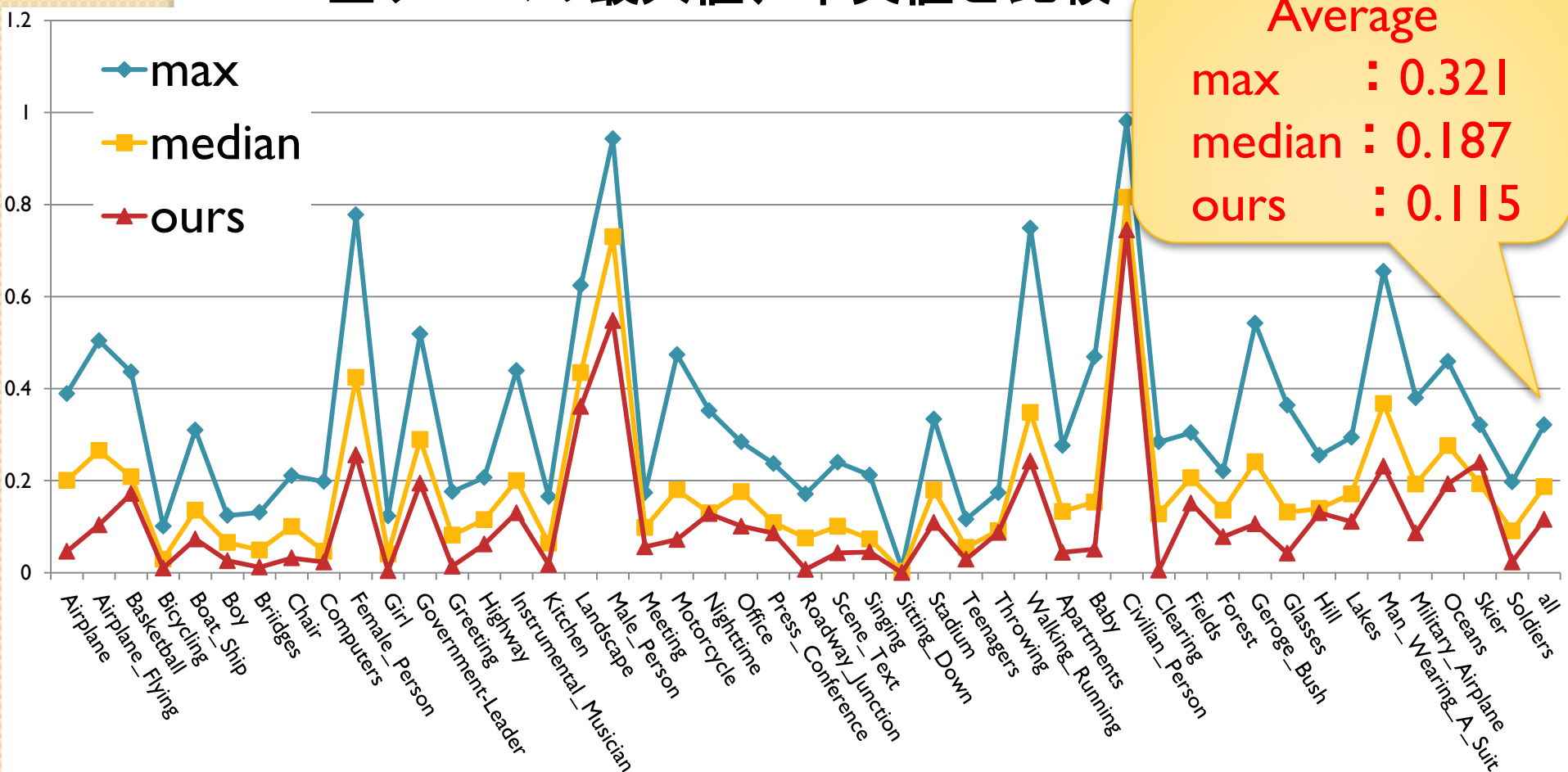
特徴抽出の時間は約80%

ただし、実時間は表記以上

マシンの不具合、エラーファイルチェック、etc...

SINタスク認識結果

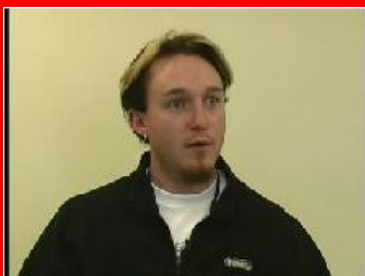
- 346カテゴリ中46カテゴリが評価対象
- 全チームの最大値、中央値と比較



1.0



アノ認識社用



al M

on

der

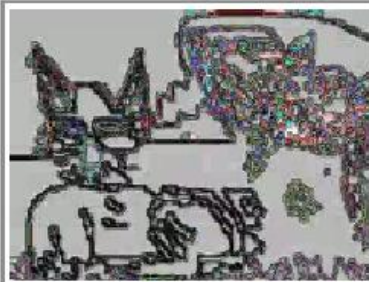
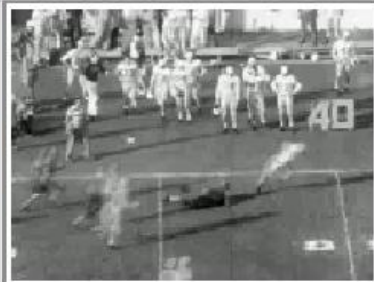
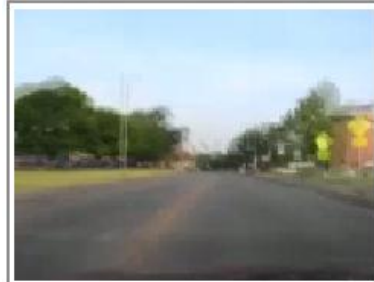
sician

e n

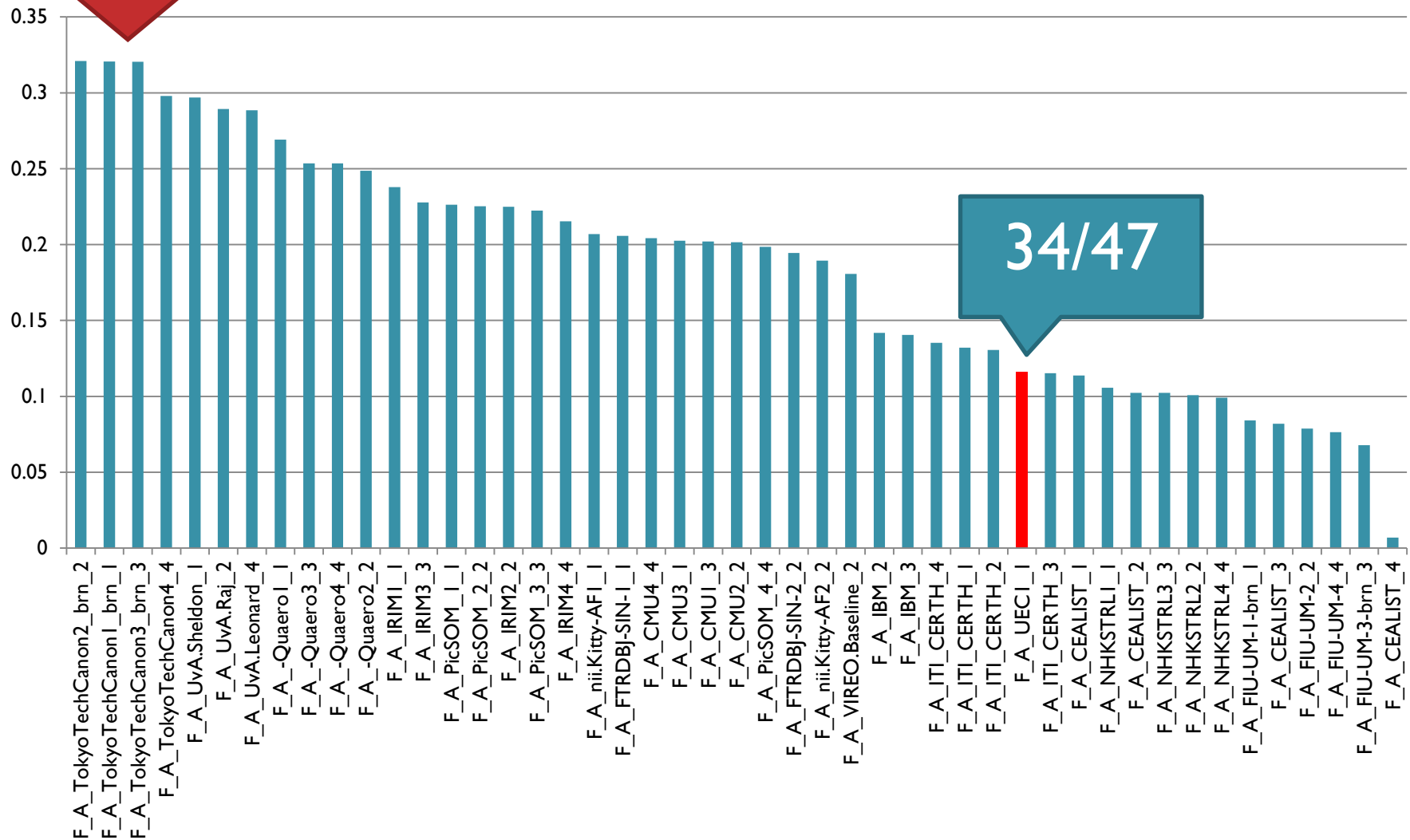
s

Suit

識結果



SINタスク全チーム結果



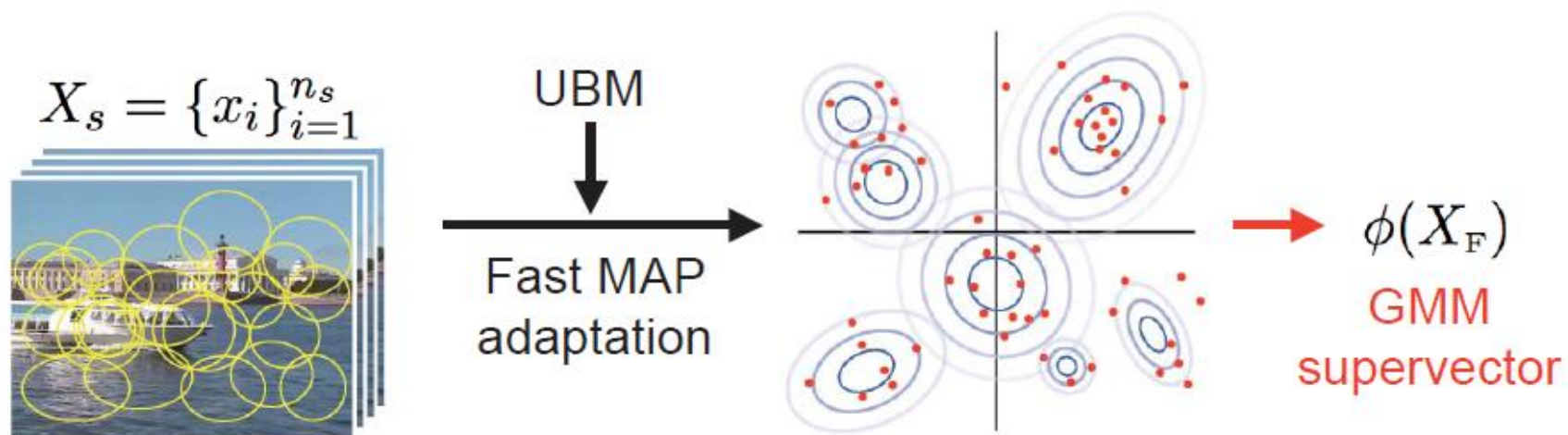
SINタスク上位手法：東工大

- **特徴量**(画像特徴5+音響特徴1)

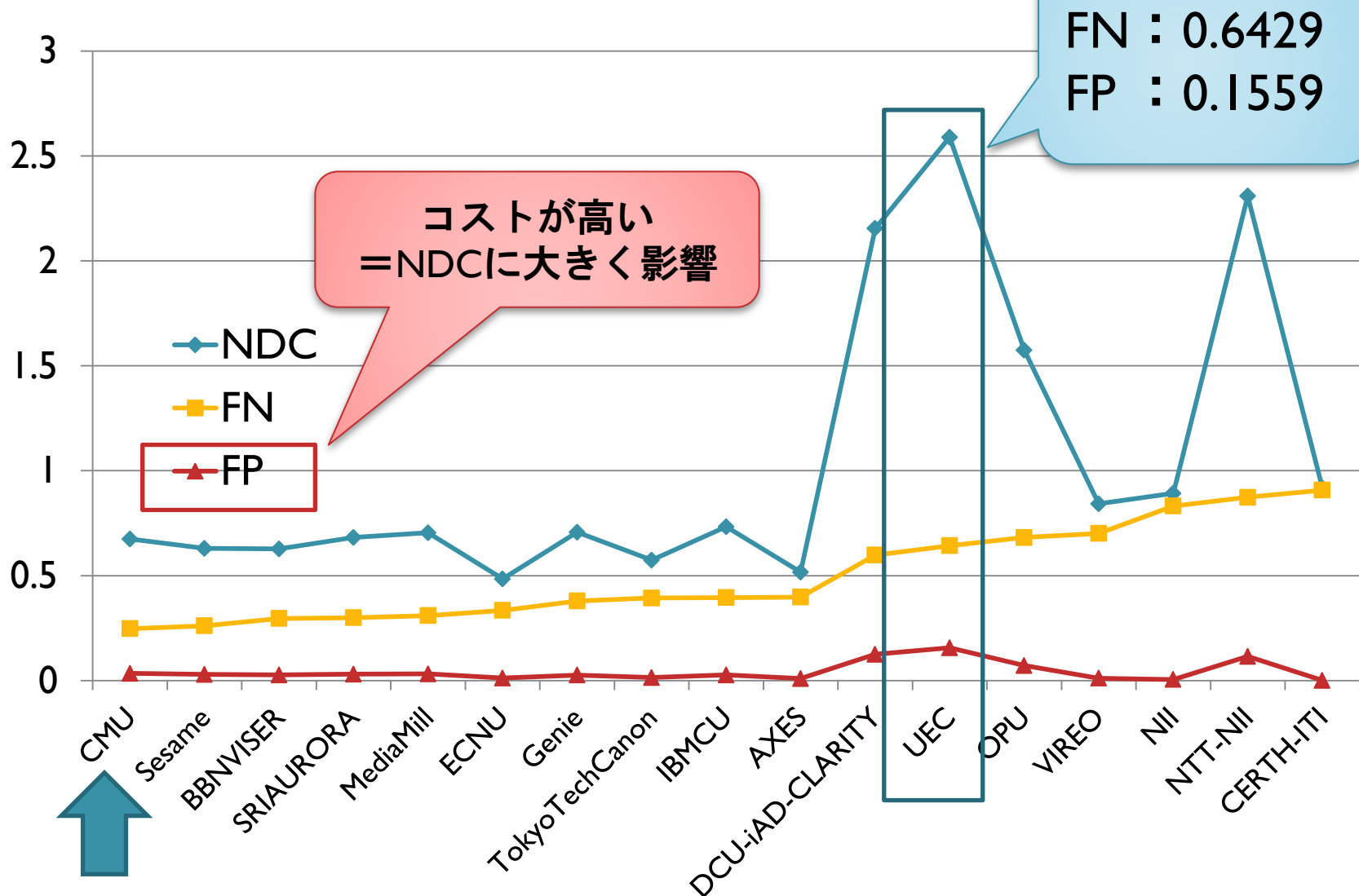
SIFT-Har, SIFT-Hes, SIFTH-Dense, HOG-Dense,
LBP-Dense, MFCC

- **GMM Supervector SVM**

木構造GMMを用いたFast MAP Adaptationにより高速化



MEDタスク全チーム結果



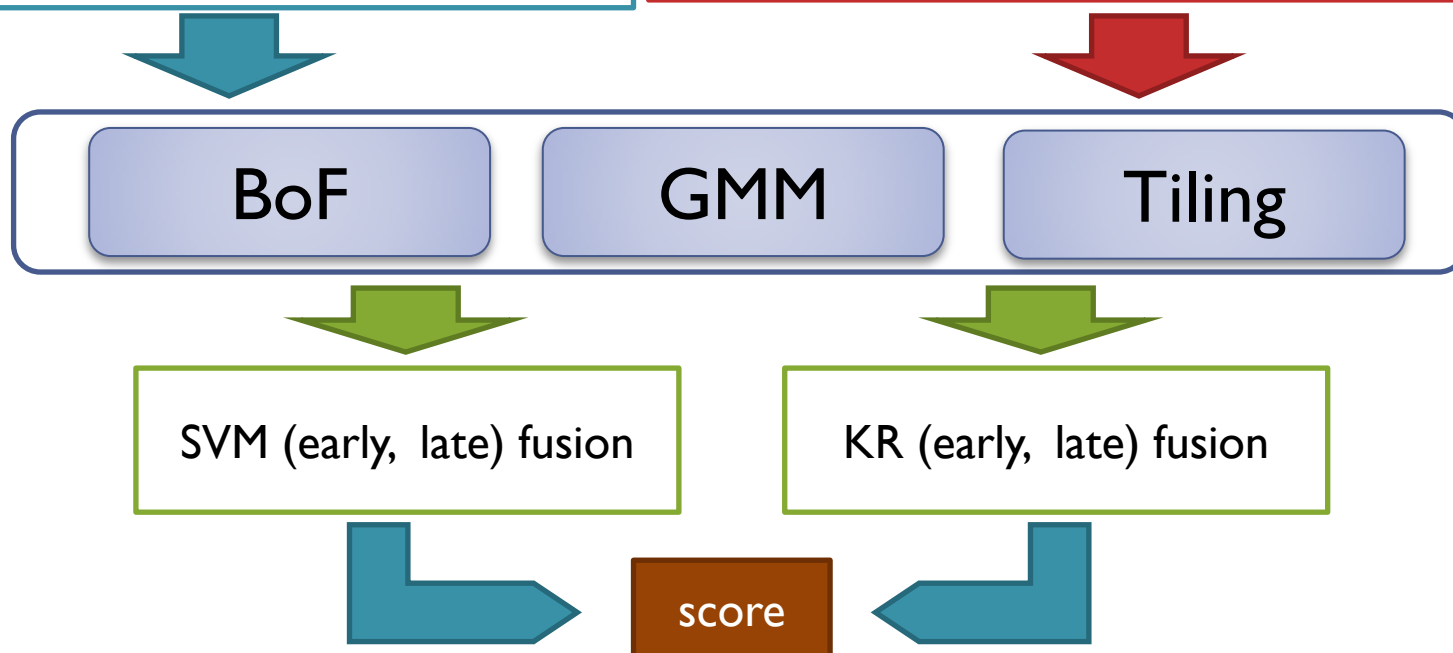
MEDタスク上位手法：CMU

画像特徴

- SIFT, C-SIFT, Transformed Color Histogram
- Semantic Indexing Concept, Object Bank
- Optical Character Recognition

動画画像特徴

- Motion SIFT, STIP, Dense Trajectory
- MFCC, AUDs
- Acoustic Scene Analysis
- Automatic Speech Recognition



まとめ

- 映像検索のワークショップ
TRECVIDの概要
 - ✓ Semantic Indexingタスク
 - ✓ Multimedia Event Detectionタスク
- 実行手法の紹介
- TRECVID全チームの手法は公開済み
<http://trecvid.nist.gov/>



ご清聴ありがとうございました

MEDタスク全チーム結果

- random : 全ショットからランダムに100ショット
- top500 : VisualRank上位500からランダムに100ショット(今回の手法)
- top100 : VisualRank上位100

