Detecting "In-play" Photos in Sports News Photo Database

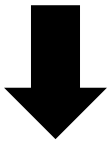
Akio Kitahara and Keiji Yanai
Univ. of Electro-Communications, Tokyo, JAPAN

PCM 2009
Dec. 16th 2009

Background

- There are many news photos on the Web
- By collecting them regularly, we can build our own news photo DB.

  We have been crawling Web photo news since 2005.
  - About half million news photos in our lab file server.

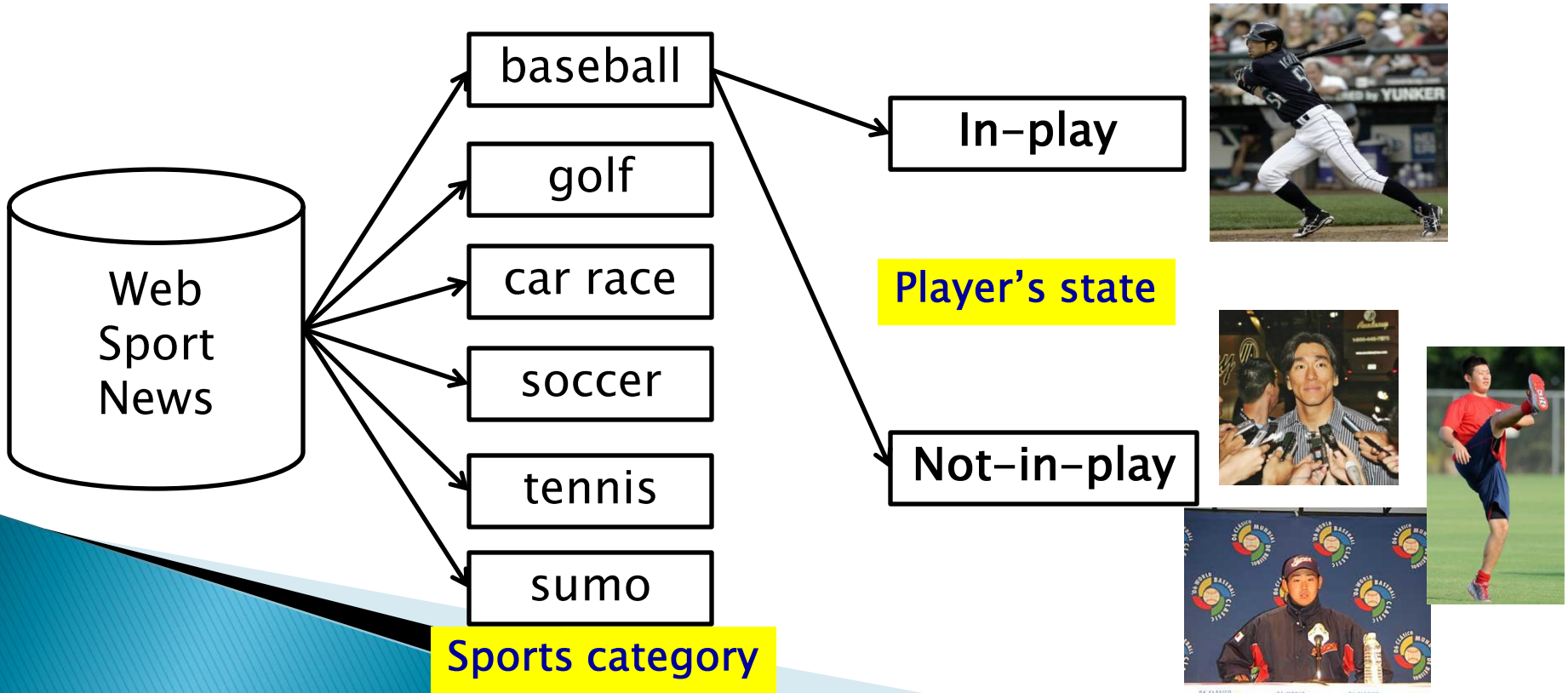


however

- Not classified precisely
  - Just classified in terms of broad genre such as "domestic", "international" "economy" and "sports"
  - So not easy to find out news photos a user want to see

Objective

- **Focus on sports photo news in this work**
- **Classify news photos in terms of sports categories and player's state ("in-play" or not)**
  - **"In-play" = a player are playing in the game**




YAHOO! JAPAN ニュース
ログイン
IDでもっと便利に[新規取得]
ニュース トピックス 写真 動画 地域 リサーチ 雑誌/ブログ ランキング
国内 海外 経済 エンターテインメント スポーツ テクノロジー 地域 ニュース提供社
[PR] ≪おすすめカードローン≫⇒来店不要・24時間受付！
スポーツ
前の写真
次の写真
“Baseball” is not shown !
photo
title
大リーグ・イチロー、日本記録へ6本
9月26日15時49分配信 時事通信
article
エンゼルス戦の7回、内野安打を放つイチロー。日米通算3079安打とし、張本勲の日本最高記録まで残り3試合であと6本とした。2得点を加えて今季99得点となり、200安打に続く8年連続100得点にも迫った

Related work

- Quattoni et al. (CVPR 2007)
  News classification of Web photo news
  using visual features and textual features

- Jain et al. (CVPR 2008)
  - Sports classification of sports photos taken
    by normal people (not for news photo taken by professional photographers)

No work to classify photos regarding sports categories and "in-play" simultaneously so far

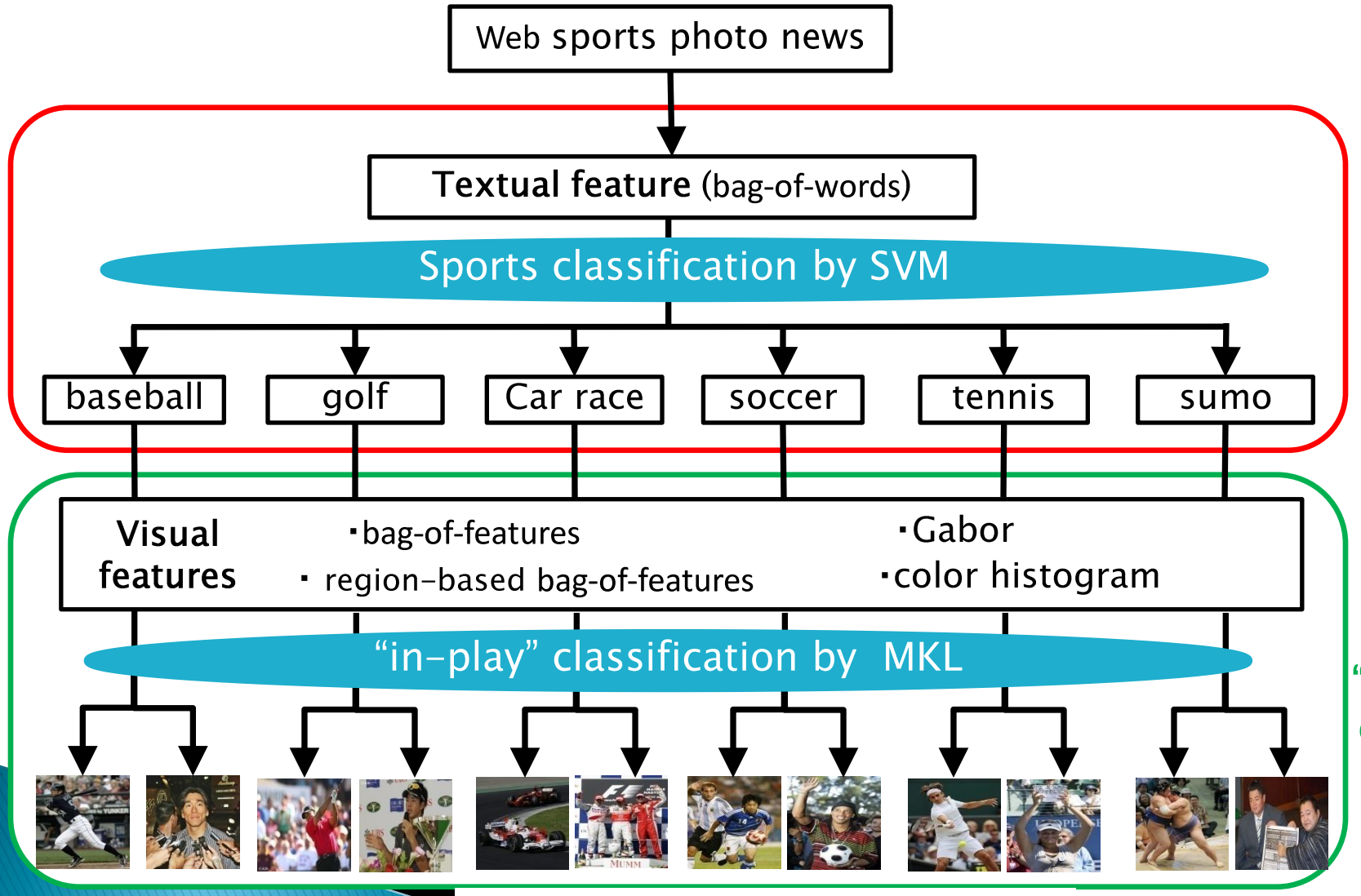
Overview of this work

- Propose two methods to estimate sports genre and player's state for Web news photos
  - Use supervised learning
  - Use textual features and visual features

- [1] two-step method
  - [1st step] sport classification with textual info.
  - [2nd step] "in-play" classification with visual info.

- [2] one-step method
  - Simultaneous sport-and-"in-play" classification with both textual and visual information

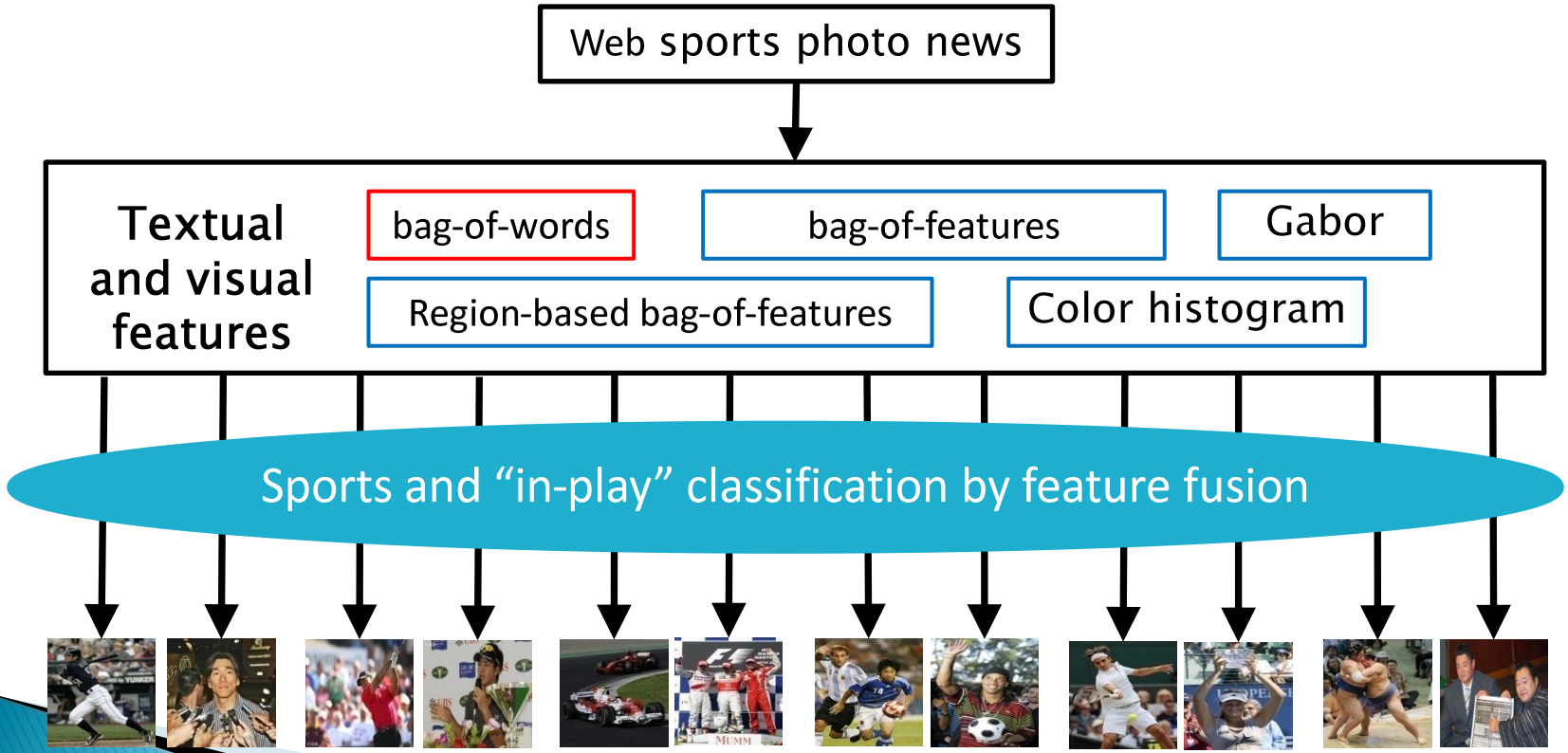
Two-step method



MKL…Multiple Kernel Learning

One-step method

- ***Simultaneous sport-and-"in-play" classification using both textual and visual information***

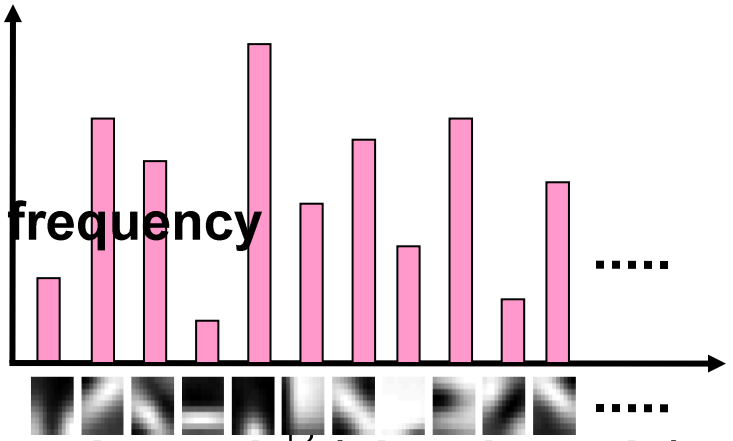


Features

- **Textual features** (from news articles)
  - Bag-of-words
    - Word histogram with top N frequent words
- **Visual features** (from news photos)
  - (normal) Bag-of-featuress
  - Region-based bag-of-features
  - Gabor features
  - Color histogram

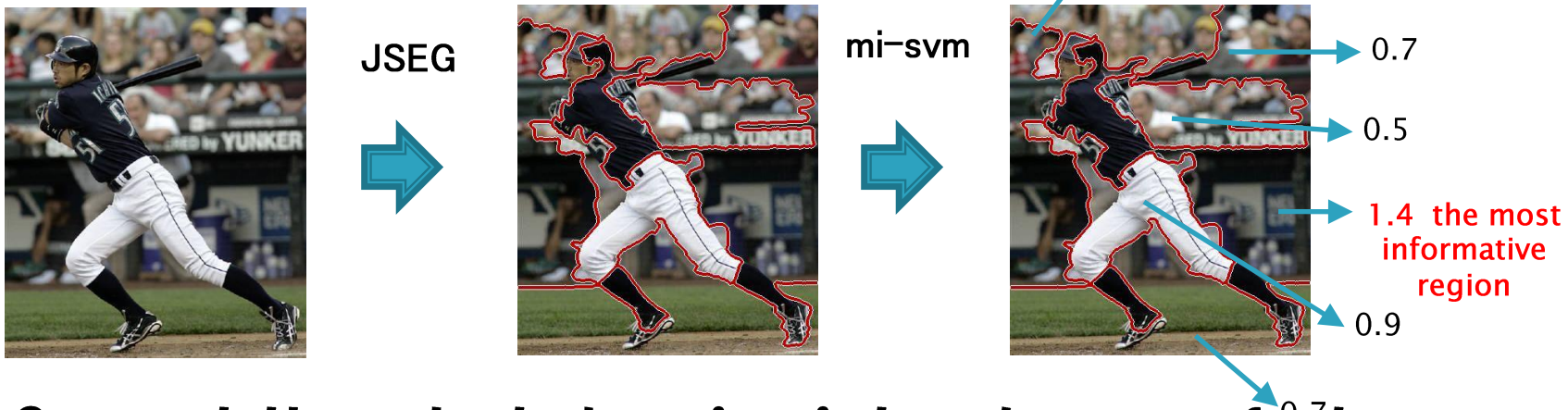
Bag-of-features

- Represent an image as a set of local features
  1. Sample 2000 points randomly
  2. Describe local patterns around sampled points with the SIFT descriptor [Lowe 2004]
  3. Generate codebooks by k-means (k=1000)
  4. Convert images into BoF vectors by voting to the nearest codewords



Region-based bag-of-keypoints

1. Region segmentation (JSEG [Deng 2001])
2. Select the most informative region (mi-SVM [Andrew 03])
   - Estimate most informative regions using multiple instance learning (mi-SVM)
   - "in-play" photos are positive, and "not-in-play" are negative samples.
   - The region with the largest SVM value is regarded as "most informative" region.



3. Convert the selected region into a bag-og-features (BoF) vector.

Not foreground detection but informative (background) detection

Feature fusion by Multiple Kernel Learning (MKL) [Sonnenburg et al. 2006]

- Is an extenstion of a SVM .
- Can handle "a combined kernel" which is a linear combination of kernels.
- Can estimate kernel weights and SVM model parameters simultaneously.
- Can integrate features by assigning one feature to one kernel.

Combined Kernel

BoF reg-BoF Gabor Color

$$\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{K} \beta_k \mathbf{k}_k(\mathbf{x}_i, \mathbf{x}_j)$$

kernel weights

Experiments: data set

- We built an Web photo news dataset.
- Data source: Yahoo! Japan Photo News
- 6-sports: baseball, golf, car race, soccer, tennis, sumo
- We gave ground truth label (sports, in-play) by hand.

Training data (2007)		
Test data (2008年)		
	Not-in-play	100 photos / each sports

- Evaluation: classification rate=

$$\frac{\#\text{ correctly classified photos}}{\#\text{ all the classified photos}}$$

(Equivalent to the average of diagonal of confusion matrix)

Experiment: two-step method

- [1st step] sports category classification
  - **Classifier:** standard SVM with chi-square kernel.
  - **Feature:** textual features (2800-dim bag-of-words)

- [2nd step] "in-play" classification
  - **Classifier:** Multi Kernel Learning (MKL) with chi-square (χ^2) kernels.

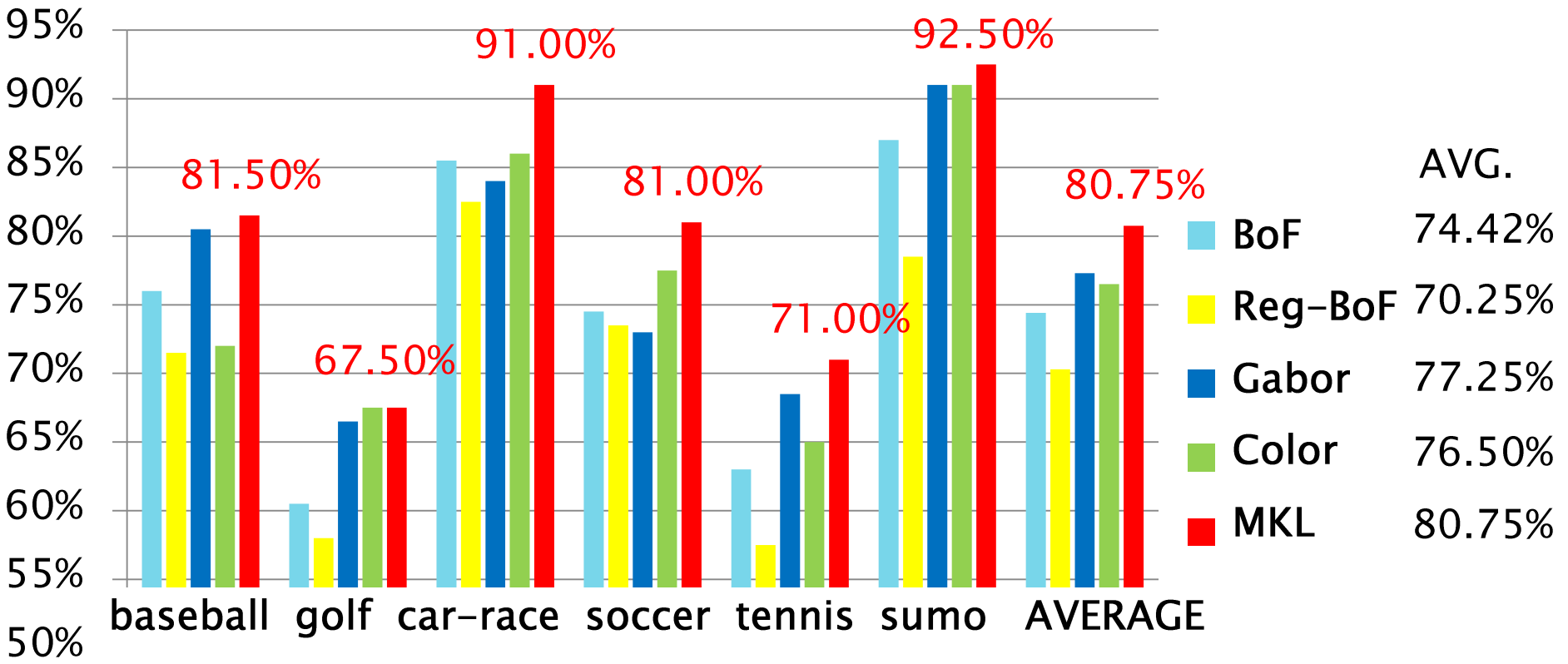
$$k(x, x') = \exp(-\gamma \sum_i \frac{\|x - x'\|^2}{x + x'})$$

  - **Features:** BoF, region-based BoF, Gabor, Color

In all the experiments,
we optimized kernel param. (γ) and cost param. (C) by grid search.

Results: [1st] sports classification

Results: [2nd] in-play classification



"Car race" photos classified as "in-play"

Correctly classified





Failed



"Car race" photos classified as "not-in-play"

Success






Failue






"Golf" photos classified as "in-play"

Success






Failure






"Golf" photos classified as "not-in-play"

Success



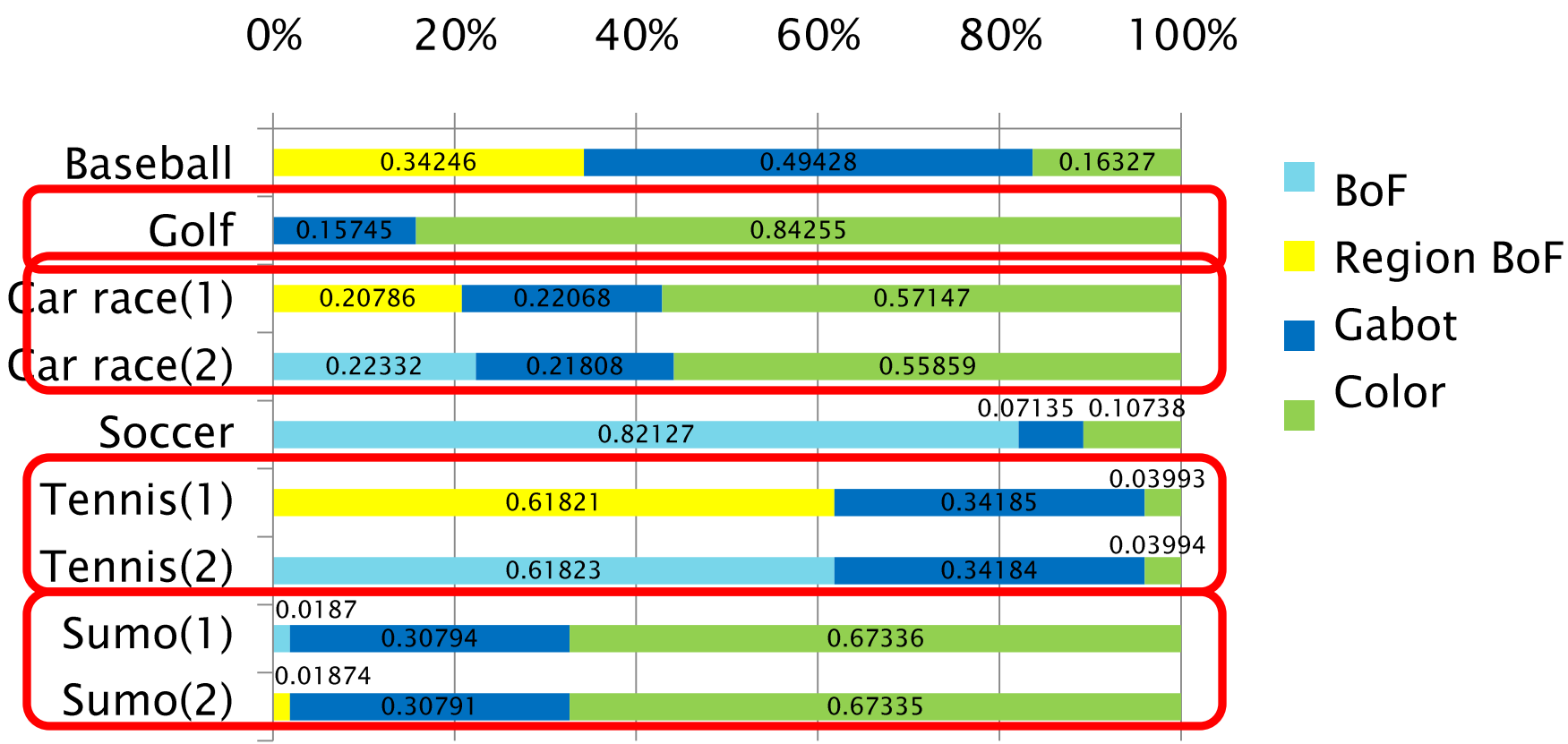


Failue





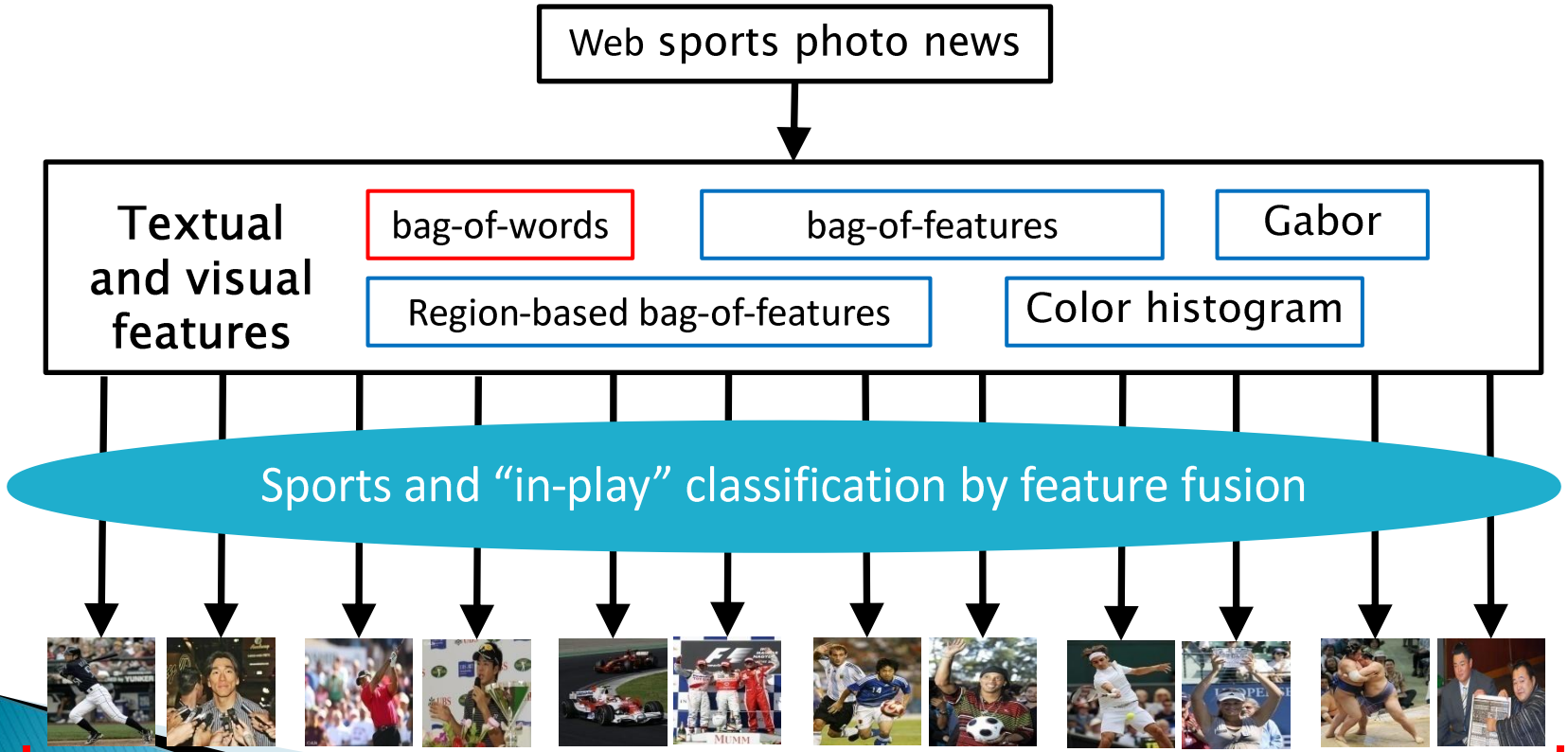

Weights estimated by MKL



- In case of “car race”, “tennis” and “sumo”, two kinds of weights were obtained, because we have the two best paremeter settings in the CV of training data.

Experiment : one-step method

- **12-class classification by MKL :**
  - **6 sports × {in-play, not-in-play}**



Results of the one-step method by single features and fusion

Features	Classification rate
Text (BoW)	10.50%
(whole image) BoF	72.08%
Region-based BoF	67.67%
Gabor	39.58%
Color histogram	49.17%
Fusion by MKL	77.08%

Confusion matrix: one-step

		Baseball		Golf		Car race		Soccer		Tennis		Sumo	
		Play	Not	Play	Not	Play	Not	Play	Not	Play	Not	Play	Not
Base ball	Play	68	32	0	0	0	0	0	0	0	0	0	0
	Not	7	93	0	0	0	0	0	0	0	0	0	0
Golf	Play	0	0	56	44	0	0	0	0	0	0	0	0
	Not	0	0	22	78	0	0	0	0	0	0	0	0
Car race	Play	0	0	0	0	81	19	0	0	0	0	0	0
	Not	0	0	0	0	7	93	0	0	0	0	0	0
Soccer	Play	0	0	0	0	0	0	85	15	0	0	0	0
	Not	0	0	0	0	0	0	22	78	0	0	0	0
Ten nis	Play	0	0	0	0	0	0	0	0	17	83	0	0
	Not	0	0	0	0	0	0	0	0	7	93	0	0
Sumo	Play	0	0	0	0	0	0	0	0	0	0	87	13
	Not	0	0	0	0	0	0	0	0	0	0	4	96

Failure results of "tennis in-play"

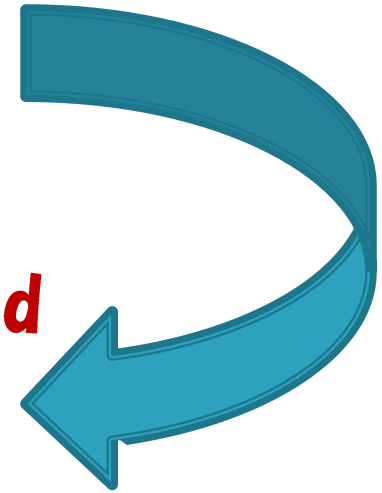
- Incorrectly classified into "not-in-play"

Conclusions

- Propose two methods to classify Web sports news photos into "in-play" or "not-in-play"

- Results by the two-step method
  - Sports classification: 99.33 %
  - "in-play" classification: 80.75%

- Results by the one-step method
  - Sports classification : 100.0 %
  - "in-play" classification : 77.08 %



Future works

- Add new sports categories.
  - Use Amazon Mechanical Turk ?
  - Transfer learning might be possible. (e.g. baseball vs softball)
- Add new player's states such as "interview" and "close-up faces"
- Add new visual features
- Recognize "pose" of players
- Introduce sport-category-dependent methods.
  - e.g. "golf club/bat/tennis racket detection"

Thank you for your attention.