Randomized Ensemble Tracking



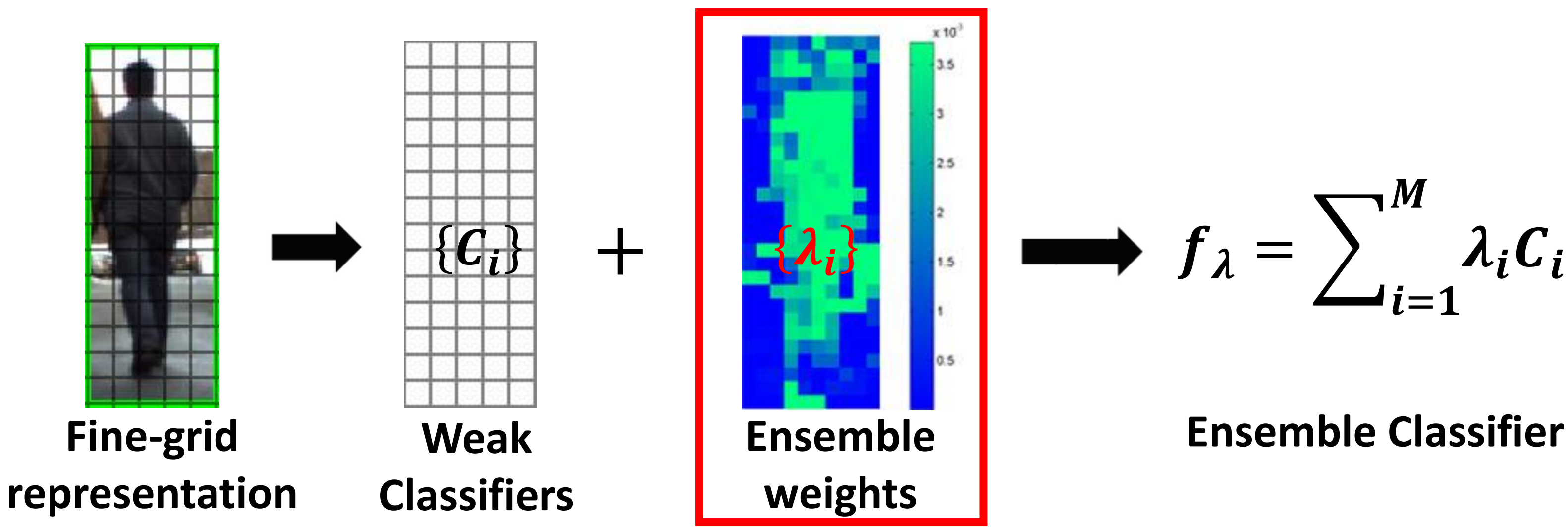

Qinxun Bai   Zheng Wu   Stan Sclaroff   Margrit Betke

IMAGE AND VIDEO COMPUTING GROUP

BOSTON UNIVERSITY

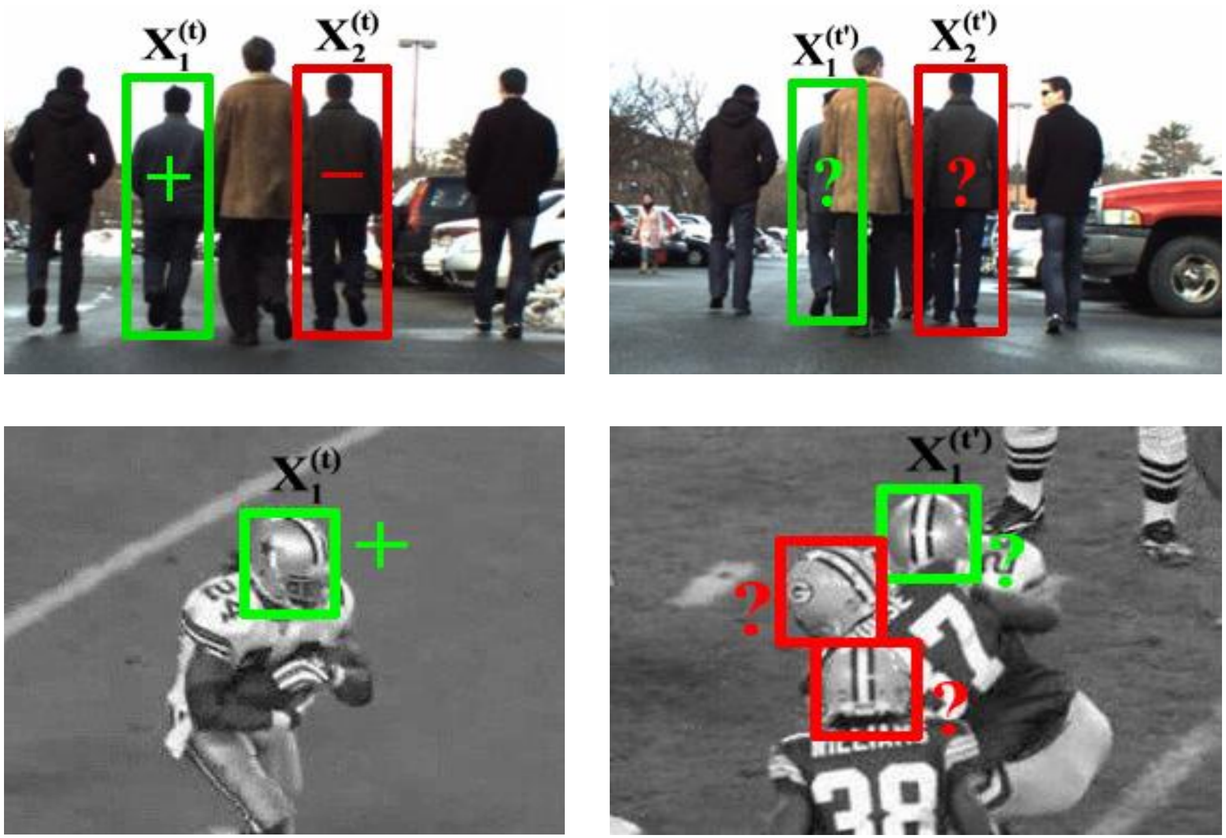
Tracking by Detection

- Object detector via classifier ensemble
  linearly combine weak classifiers with associated weights
- Online learning
  online update both weak classifiers and ensemble weights



Our focus !

Challenging!



Motivation

Limitations of previous method: online boosting

- Lack of strong *theoretical guarantee*

  converges to its off-line counterpart under restrictive conditions

- Classifier ensemble: *uncertainty* in ensemble weights

  existing methods don't model this, but adopt deterministic estimate

- Relies on *importance weights* of training data

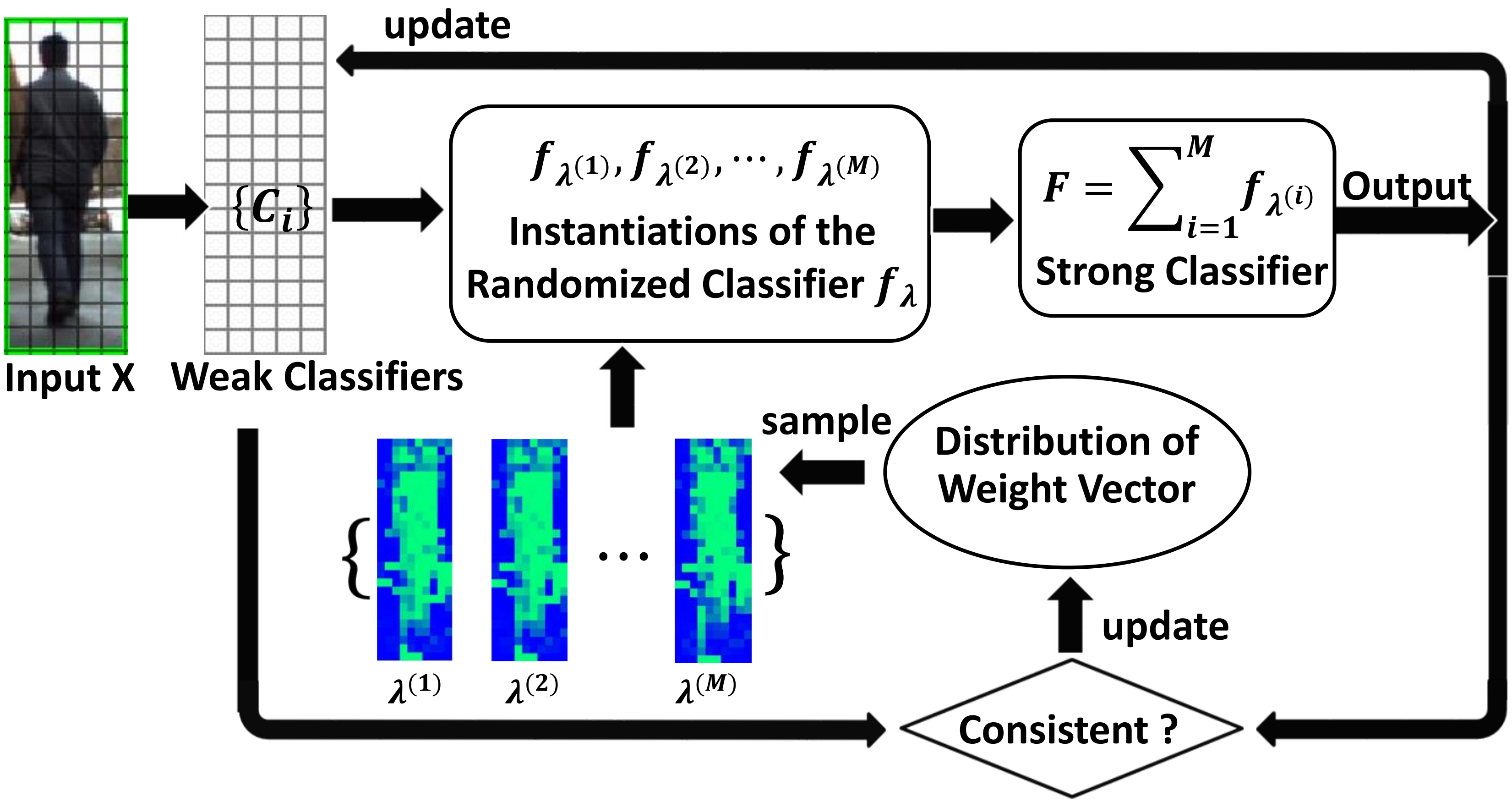
  inherited from offline version, difficult to estimate in online environment
  even more challenging given the non-stationary $p(x, y)$

Our Idea

Ensemble weight vector as a random variable

- A probabilistic interpretation of which features of object are relatively more discriminative
- Characterize the distribution of weight vector online
- Bayesian filtering to recursively estimate its posterior distribution
- Expected output of the randomized ensemble
- Theoretical guarantees on asymptotic properties (Bai, et al. ICML 14)

Method Overview



Classification

Randomized ensemble

$$f_{\boldsymbol{\lambda}}(x) = \begin{cases} 1 & if \sum_{i=1}^{N} \lambda_i c_i(x) \geq \tau \\ 0 & otherwise \end{cases}$$

Expected output

$$y^* = \int f_{\boldsymbol{\lambda}}(x) p\left(\boldsymbol{\lambda} \middle| S^{(0)} \cdots S^{(t-1)}\right) d\boldsymbol{\lambda}$$

Approximate by sampling and voting

$$F(x) = \begin{cases} 1 & if \frac{1}{M} \sum_{j=1}^{M} f_{\boldsymbol{\lambda}^{(j)}}(x) \geq \frac{1}{2} \\ 0 & otherwise \end{cases}$$

where $f_{\boldsymbol{\lambda}^{(1)}}, f_{\boldsymbol{\lambda}^{(2)}}, \cdots, f_{\boldsymbol{\lambda}^{(M)}}$ are instantiations of randomized ensemble $f_{\boldsymbol{\lambda}}$

Model Update

Multinomial-Dirichlet conjugacy

$$p(\boldsymbol{\lambda}|\alpha, H, g_{1:N}) \propto p(g_1 \cdots g_N|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\alpha, H) \propto Dir(\boldsymbol{\lambda}; \alpha', H')$$

Observation model $p(g_1 \cdots g_N|\boldsymbol{\lambda}) = k \prod_{i=1}^{N} (\lambda_i)^{g_i}$

- λ as multinomial parameters

  expectation of "relative reliability" of weak classifiers

- Observations of "relative reliability of each weak classifier"

  performance measure $g: \{1,2,\cdots,N\} \rightarrow [0,2]$

- g_i: "occurrences of the i^{th} weak classifier being reliable"

$$g_i = g(i) = \frac{2}{1 + e^{-s_i w_i}}$$

Experiments Setup

- Test on 28 video sequences
- 2 baseline methods
  - B1(SVM): concatenation of local features + SVM
  - B2(OB): same representation & same weak classifiers + online boosting
- Compare with 8 state-of-art object trackers

Experimental Results

Table: Tracking performance on datasets with fixed scale objects. Each entry in the table reports the ACLE and TA performance measure as ACLE (TA).

	TLD	PROST	CT	DF	Frag	MIL	Struck	SVM (Baseline 1)	OB (Baseline 2)	DET (Ours)	RET (Ours)
Coke	11 (.68)	-	16 (.30)	7 (.76)	61 (.06)	21 (.21)	7 (.76)	12 (.24)	20 (.12)	14 (.22)	13 (.23)
David	4 (1)	15 (.80)	16 (.89)	10 (1)	46 (.47)	23 (.60)	7 (.98)	4 (1)	11 (1)	7 (1)	6 (1)
Dollar	6 (1)	-	20 (.92)	5 (1)	33 (.66)	15 (.93)	14 (1)	5 (1)	7 (1)	5 (1)	4 (1)
Face1	15 (.99)	7 (1)	19 (.89)	5 (1)	7 (1)	27 (.78)	9 (1)	7 (1)	24 (.81)	8 (.99)	7 (1)
Face2	13 (.97)	17 (.82)	10 (1)	11 (.99)	45 (.48)	20 (.82)	7 (.98)	7 (1)	26 (.60)	10 (1)	9 (1)
Girl	18 (.93)	19 (.89)	21 (.78)	22 (.73)	27 (.70)	32 (.56)	10 (1)	56 (.26)	25 (.89)	34 (.72)	19 (.84)
Sylv	6 (.97)	11 (.67)	9 (.75)	16 (.67)	11 (.73)	11 (.74)	10 (.87)	22 (.60)	8 (.88)	10 (.82)	12 (.80)
Tiger1	6 (.89)	7 (.79)	10 (.78)	7 (.89)	20 (.40)	15 (.57)	7 (.85)	5 (.97)	34 (.35)	4 (.97)	4 (.92)
Tiger2	29 (.26)	-	13 (.60)	7 (.82)	39 (.09)	17 (.63)	12 (.60)	5 (.90)	6 (.86)	4 (.96)	4 (.96)
Twinings	16 (.52)	-	9 (.89)	11 (.77)	15 (.69)	10 (.85)	7 (.98)	24 (.45)	28 (.43)	15 (.57)	21 (.63)
Surfer	4 (.97)	-	19 (.13)	5 (.95)	139 (.20)	9 (.76)	8 (.74)	3 (.97)	3 (.99)	3 (.99)	3 (.99)
Board	11 (.87)	39 (.75)	62(.53)	-	90 (.68)	51 (.68)	37 (.78)	59 (.70)	244 (.11)	39 (.84)	38 (.86)
Box	17 (.92)	13 (.91)	14 (.89)	-	57 (.61)	105 (.25)	140 (.37)	106 (.40)	13 (.90)	13 (.96)	10 (.97)
Lemming	16 (.86)	25 (.71)	63 (.31)	-	83 (.55)	15 (84)	31 (.69)	82 (.46)	88 (.26)	80 (.47)	16 (.82)
Liquor	7 (.92)	22 (.85)	180 (.21)	-	31 (.80)	165 (21)	74 (.60)	82 (.52)	26 (.24)	13 (.95)	13 (.96)

Experimental Results (continue)

Table: Tracking performance (AOR (TA)) on datasets with varying, sometimes significant changes in object scales.

	VTD	TLD	SVM (B1)	OB (B2)	DET (Ours)	DET* (Ours)	RET (Ours)
Animal	.65 (.92)	.48 (.76)	.73 (1)	.62 (.94)	.72 (1)	.7 (1)	.72 (1)
Basketball	.72 (.98)	-	.43 (.36)	.51 (.50)	.53 (.63)	.62 (.92)	.54 (.64)
Football	.66 (.78)	.55 (.77)	.56 (.78)	.69 (.93)	.61 (.74)	.66 (.96)	.62 (.82)
Shaking	.75 (.99)	.12 (.16)	.20 (.21)	.03 (.04)	.55 (.64)	.60 (.80)	.44 (.53)
Singer1a	.82 (1)	.66 (.93)	.70 (.98)	.46 (.37)	.70 (.90)	.70 (.89)	.73 (.97)
Singer1b	.59 (.63)	.11 (.10)	.20 (.12)	.20 (.12)	.70 (.89)	.70 (.93)	.69 (.93)
Singer2	.74 (.97)	-	.29 (.23)	.69 (.93)	.07 (.06)	.08 (.06)	.38 (.50)
Skating1a	.68 (.92)	.39 (.43)	.48 (.39)	.48 (.38)	.56 (.55)	.58 (.64)	.48 (.52)
Skating1b	.67 (.90)	.42 (.58)	.34 (.42)	.44 (.27)	.43 (.45)	.54 (.58)	.46 (.52)
Skating2	.57 (.68)	-	.54 (.63)	.40 (.39)	.45 (.48)	.55 (.71)	.61 (.75)
Soccer	.39 (.32)	-	.15 (.17)	.34 (.25)	.12 (.14)	.40 (.35)	.27 (.30)
ETH	.34 (.31)	.51 (.63)	.56 (.62)	.57 (.61)	.50 (.39)	-	.65 (.92)
walking	.33 (.22)	.19 (.20)	.59 (.67)	.28 (.08)	.54 (.68)	-	.77 (1)

Examples of Tracking Result

Play the Demo!

Success and Failure Snapshots

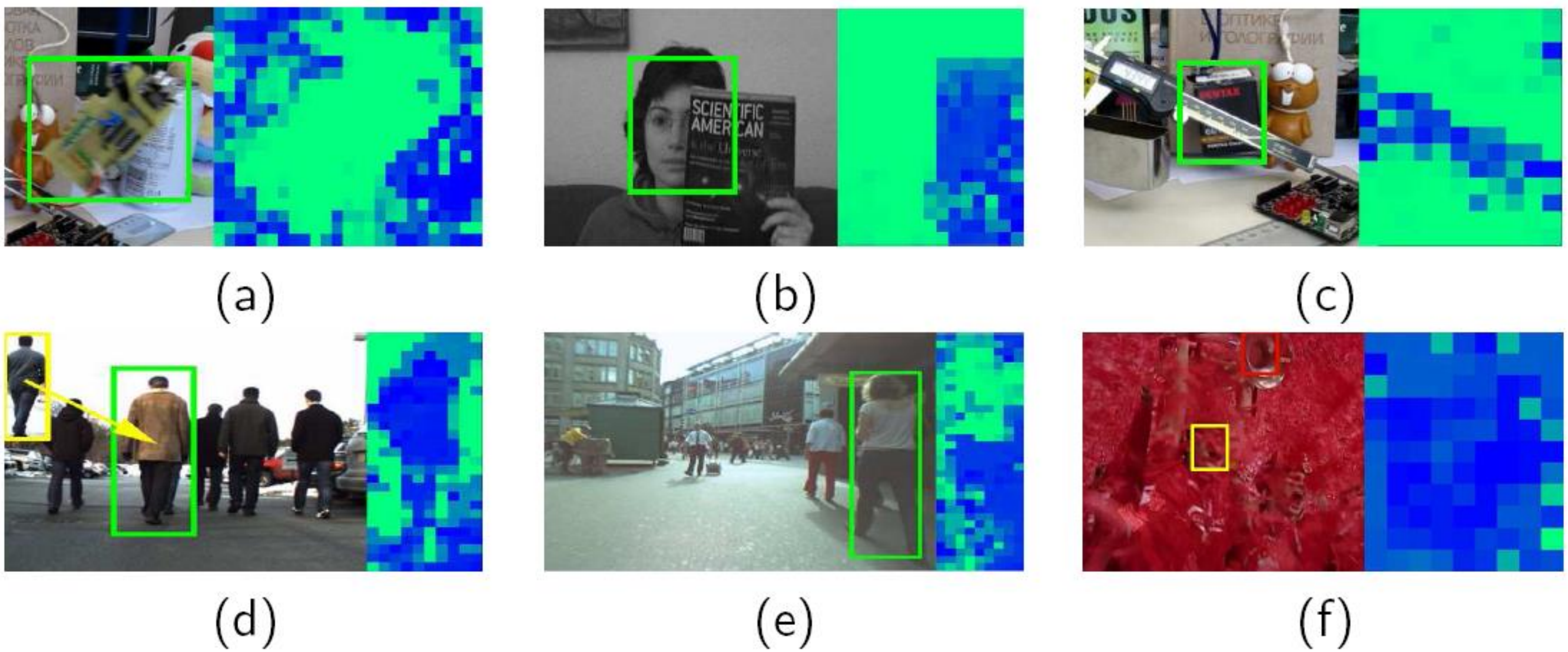


Figure: Sample images with true detections (green), false alarms (red), and ground truth (yellow) and snapshots of base distribution H of Dirichlet distribution (greener means higher weight of the associated weak classifier and its higher discriminate ability, bluer means lower weight).