ROBUST MIL BASED FEATURE TEMPLATE LEARNING FOR OBJECT TRACKING

Shivam Agrawal (20172074)

Kshitij Nagar (20172046)

Sagar Thakur (20172103)

Vachaspathi Ramakrishnan (20172011)

PROBLEM STATEMENT

- Because of appearance variations, online trackers requires updating of the tracking model.
- This often leads to tracking drift problem.
  - Contaminated samples, and
  - Misaligned samples caused by tracking inaccuracy.

PROPOSED MODEL

The proposed feature template learning framework is capable of:

- Adaptively learning uncontaminated feature templates by separating out contaminated samples, and
- Resolving label ambiguities caused by misaligned samples via MIL.

PREVIOUS WORK

Sparsity-based trackers

Models the outliers in the corrupted target's samples caused by occlusion or noise.

Advantage

Detect and prevent contaminated samples which increases robustness to tracker.

Problem

Do not explicitly handle the misaligned samples.

Most existing tracking algorithms which aim to reduce tracking drift are not effective in handling either or both scenarios.

PREVIOUS WORK

Weakly/semi-supervised learning-based methods

Aim to resolve the label ambiguities caused by misaligned samples.

Problem

More likely to update with contaminated samples and thereby deteriorates the discriminability.

ADVANTAGE OF PROPOSED MODEL

- Integrates robustness of sparse representation & the flexibility of multiple instance learning into model updating process
- Enables tracker to separate out corrupted samples & resolve label ambiguity.

Robust MIL-based Feature Template Learning

PROBABILISTIC CONTAMINATED FEATURE MODELLING

Objective of the learning model is to adaptively learn the uncontaminated feature templates for sparse representation of the tracked object while explicitly modeling the contaminated features.

$$Y = DX + E \Rightarrow E = Y - DX$$

$Y = [\, y_1,\ldots, y_n] \in R^{d\times n}$ denote the target samples.

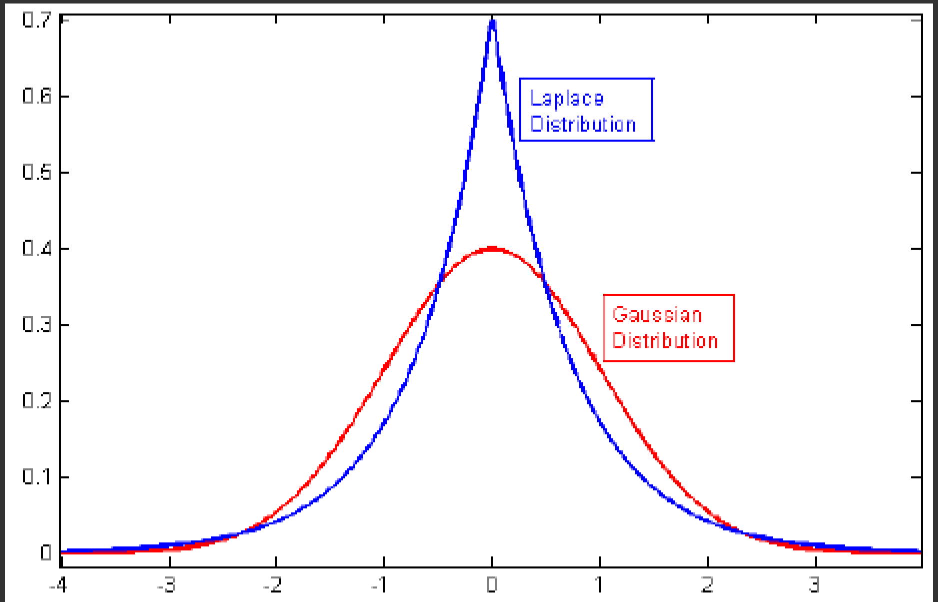
$D = [D1,\ldots,D_p] = [(D_1)^T,\ldots,(D_d)^T]^T \in R^{d\times p}$ are the set of templates.

$X = [X_1 \ldots X_n]$ is the sparse coefficient matrix of target samples.

$E = [E_1 \ldots E_n]$ are the separated contaminated features.

PROBABILISTIC CONTAMINATED FEATURE MODELLING

To explicitly model the uncertainty existing in the outliers for contaminated samples, we incorporate the probabilistic constraint that each element in E is independent and subjected to a zero-mean Laplace distribution with variance $1/b^2$, i.e.



$$P(E|\cdot;b) = (b/2)^{(nd)} \exp(-b\|E\|_1),$$

which implies

$$P(Y \mid D,X;b) = (b/2)^{(nd)} \exp(-b\|Y - DX\|_1)$$

PROBABILISTIC CONTAMINATED FEATURE MODELLING

The prior distribution for each X_j is defined as:

$P(X_j) \propto \exp(-\lambda\ \phi(X_j))$,

where $\phi(\cdot)$ is a sparsity function and λ is a constant. Here we use L1 penalty as the sparsity function,

i.e. $\phi(X_j)=||X_j||_1$.

PROBABILISTIC CONTAMINATED FEATURE MODELLING

By taking the logarithm of the joint distribution with respect to Y , D, and X, i.e. P (Y, D, X|b) = P (Y |D, X; b)P (D|; b)P (X|; b), then the maximum a posteriori (MAP) estimation turns out to be:

$$\min(D,X)\ b||Y - DX||_1 + \lambda||X||_1$$

$$\text{s.t. } ||D_{.,j}||_2 \leq c,\ j = 1,\ldots,n$$

RESOLVING LABEL AMBIGUITY VIA BAG-LEVEL BASED MAP

- Feature templates are learnt using the collection of weakly-labeled samples based on a probabilistic multiple instance learning (MIL) strategy.
- Only the label information for collection of samples called bag are available.
- A bag is positive if at least one of its samples is positive otherwise the bag is negative
- For the j-th sample in the i-th sample bag the probablility P_{ij} can be given by $(b/2)^{(d)} \exp(-b\|y_{ij} - Dx_{ij}\|_1)$

RESOLVING LABEL AMBIGUITY VIA BAG-LEVEL BASED MAP

- Therefore the probability of the i^{th} sample bag belonging to the foreground (target) can be defined as $P_i = \max_{(j)} P_{ij}$
- The likelihood at the bag level can be given as follows:

$$L(\Omega) = \prod_{i=1}^{N^+} (\max_j P^{ij}) \prod_{i=N^+ +1}^{N} \prod_{j=1}^{S^i} (1 - P^{ij})$$

- Since the samples from small neighborhood are partially overlapped, they are specially correlated.
- So it makes sense to use a generalized mean for $\max_j P^{ij}$

RESOLVING LABEL AMBIGUITY VIA BAG-LEVEL BASED MAP

Generalized mean :

$$\max_j P^i \approx \left(\frac{1}{S^i} \sum_{j=1}^{S^i} (P^{ij})^{\alpha} \right)^{1/\alpha}$$

With the same prior distribution on the feature templates and sparse coefficients by taking the logarithm of L, the bag-level-based MAP estimation of the feature template and sparse coefficients can be obtained by the following:

RESOLVING LABEL AMBIGUITY VIA BAG-LEVEL BASED MAP

$$\min_{\Omega} - \sum_{i=1}^{N^+} \log \left(\frac{1}{S^i} \sum_{j=1}^{S^i} (b/2)^d \exp(-b\|y^{ij} - Dx^{ij}\|_1) \right)$$

$$- \beta \sum_{i=N^+ +1}^{N} \log \left(1 - (b/2)^d \exp(-b\|y^{i1} - Dx^{i1}\|_1) \right)$$

$$+ \lambda \sum_{i=1}^{N} \sum_{j=1}^{S^i} \|x^{ij}\|_1$$

$$\text{s.t. } \|D_{\cdot,j}\|_2 \leq c, j = 1, \ldots, n$$

OPTIMIZATION

SMOOTHING

The ℓ_1 penalty function satisfies that

$$\|y^{ij} - Dx^{ij}\|_1 = \sum_{k=1}^{d} |y^{ij}_{k,\cdot} - D_{k,\cdot}x^{ij}|,$$

Let $u = y^{ij}_{k,\cdot} - D_{k,\cdot}x^{ij}$.

Thus we have

$$|u| = \max_s s \cdot u \quad \text{s.t.} -1 \leq s \leq 1,$$

WE GET SMOOTHENED FUNCTION AS

$$g_\theta(y^{ij}_{k,\cdot}, D_{k,\cdot}, x^{ij}) = \max_s s \cdot (y^{ij}_{k,\cdot} - D_{k,\cdot} x^{ij}) - \frac{\theta}{2} s^2$$

$$\text{s.t.} -1 \leq s \leq 1$$

FINDING VALUE OF 'S'

By taking derivative of the smothened function and setting it to zero, we get

$$s = \text{median}\left\{ \frac{y^{ij}_{k,\cdot} - D_{k,\cdot}x^{ij}}{\theta}, -1, 1 \right\}$$

FINAL SMOOTHED FUNCTION

$$g_\theta(y_{k,\cdot}^{ij}, D_{k,\cdot}, x^{ij})$$

$$= \begin{cases} |y_{k,\cdot}^{ij} - D_{k,\cdot}x^{ij}| - \frac{\theta}{2}, & |y_{k,\cdot}^{ij} - D_{k,\cdot}x^{ij}| > \theta \\ \frac{(y_{k,\cdot}^{ij} - D_{k,\cdot}x^{ij})^2}{2\theta}, & \text{else} \end{cases}$$

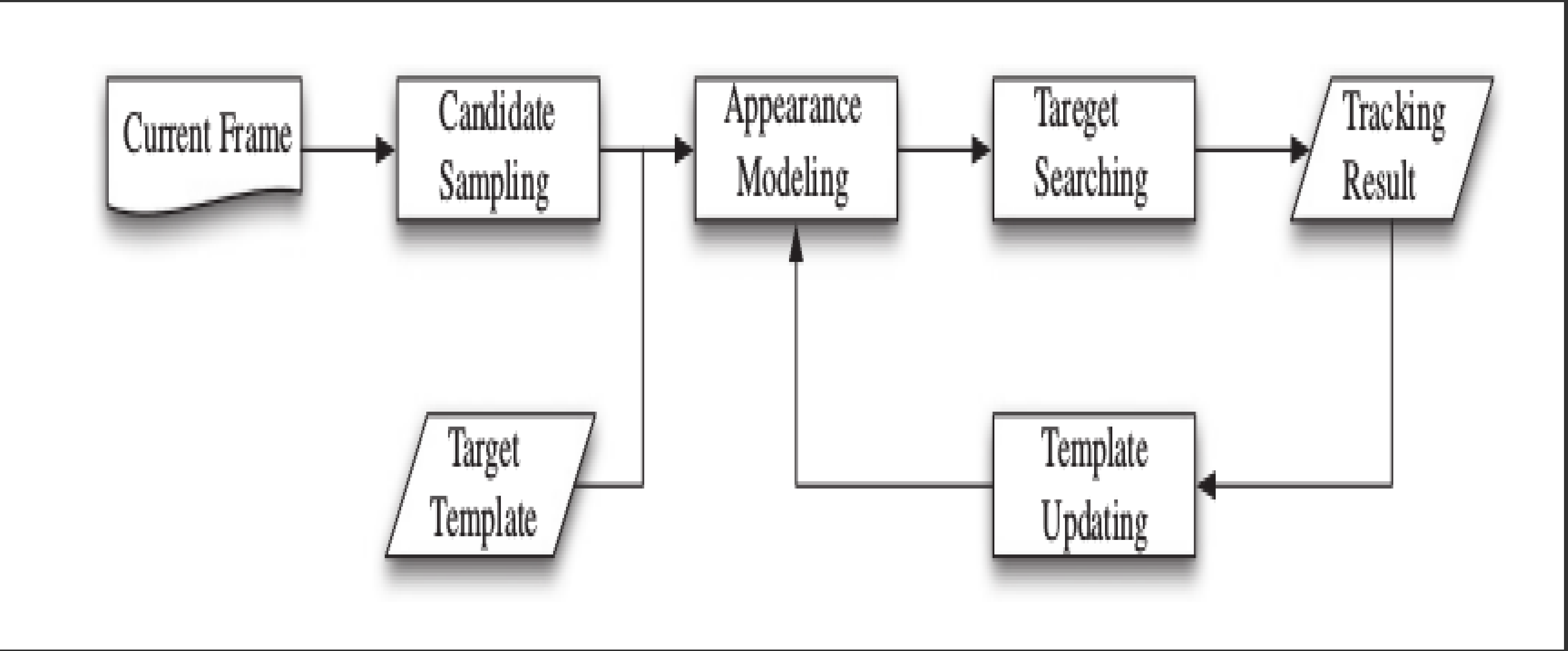
FINDING GRADIENT W.R.T D_K

$$\frac{\partial g_\theta(\cdot,\cdot,\cdot)}{\partial D_{k,\cdot}^T} = \begin{cases} \frac{D_{k,\cdot}x^{ij} - y_{k,\cdot}^{ij}}{\theta} x^{ij} & |y_{k,\cdot}^{ij} - D_{k,\cdot}x^{ij}| \leq \theta \\ \text{sign}(D_{k,\cdot}x^{ij} - y_{k,\cdot}^{ij}) x^{ij} & \text{else} \end{cases}$$

FINDING GRADIENT W.R.T X^{I,J}

$$\frac{\partial g_\theta(\cdot,\cdot,\cdot)}{\partial x^{ij}} = \begin{cases} \frac{D_{k,\cdot}x^{ij} - y^{ij}_{k,\cdot}}{\theta} x^{ij} & |y^{ij}_{k,\cdot} - D_{k,\cdot}x^{ij}| \leq \theta \\ \text{sign}(D_{k,\cdot}x^{ij} - y^{ij}_{k,\cdot})x^{ij} & \text{else} \end{cases}$$

FLOWCHART FOR ONLINE OBJECT TRACKING



HOG FEATURES