Visual Tracking with Online Multiple Instance Learning

Boris Babenko, Ming-Hsuan Yang, Serge Belongie

Kelsie Zhao
Content

• Goal
• Background: Tracking by Detection
• Previous Work
• New Tracking Solution
  • MILTrack
  • Online MILBoost
• Experiments & Results
Goal

Track one arbitrary object in video, given its location in first frame
Background: Tracking by detection

• Frame 1 is labeled, tracker location known
Background: Tracking by detection

- Crop one positive and some negative patches near tracker
Background: Tracking by detection

• Use patches to train the classifier
Background: Tracking by detection

- Frame 2 comes
Background: Tracking by detection

• Calculate classifier response within a range of the old tracker location
Background: Tracking by detection

- Find the maximum response location
Background: Tracking by detection

- Move tracker
Background: Tracking by detection

• Repeat

Frame 2
Background: Tracking by detection

- Problem: If tracker location is not precise, might select bad training examples
Background: Tracking by detection

• Problem: If tracker location is not precise, might select bad training examples

Model start to degrade!
Background: Tracking by detection

- Problem: If tracker location is not precise, might select bad training examples

Model start to degrade!

- How to select good training examples?
Previous Work

- Solution 1: multiple positive examples around tracker location

\{(x_1, 1), (x_2, 1), (x_3, 1), (x_4, 0), (x_5, 0)\}

Classifier
Previous Work

• Solution 1
  Might confuse classifier!

\{ (x_1, 1), (x_2, 1), (x_3, 1), (x_4, 0), (x_5, 0) \}
Previous Work

• Solution 2: Multiple Instance Learning (MIL)

[Keeler ‘90, Dietterich et al. ‘97]
Previous Work: Multiple Instance Learning

- Multiple examples in one bag

[Keeler ‘90, Dietterich et al. ‘97]
Previous Work: Multiple Instance Learning

- Multiple examples in one bag

[Keeler ‘90, Dietterich et al. ‘97]
Previous Work: Multiple Instance Learning

- Multiple examples in one bag
- One bag one label

[Keeler ‘90, Dietterich et al. ‘97]
Multiple examples in one bag
One bag one label
Bag **Positive** if at least one example is Positive

[Keeler ‘90, Dietterich et al. ‘97]
Previous Work: Multiple Instance Learning

\{(X_1, 1), (X_2, 0), (X_3, 0)\}

[Keeler '90, Dietterich et al. '97]
Previous Work: Multiple Instance Learning

MIL training input:

\[\{(X_1, y_1) \ldots (X_n, y_n)\},\]

where,

\[X_i = \{x_{i1} \ldots x_{im}\},\]

\[y_i = \max_j y_{ij}\]

[Keeler ‘90, Dietterich et al. ‘97]
Previous Work: Multiple Instance Learning

MIL training input:

\[ \{(X_1, y_1) \ldots (X_n, y_n)\}, \]

where,

\[ X_i = \{x_{i1} \ldots x_{im}\}, \]

\[ y_i = \max_j y_{ij} \]

Bag babel is 1 if at least one instance is 1

[Keeler '90, Dietterich et al. '97]
Now we have training examples!

How to train the classifier?
Previous Work: MILBoost

• MIL + boosting

Train a boosting classifier that maximizes log likelihood of bags

\[ \log L = \sum_{i} \log(p(y_i | X_i)) \]

where,

\[ p(y_i | X_i) = 1 - \prod_{j} (1 - p(y_i | x_{ij})) \]

[Viola et al. '05]
Previous Work: MILBoost

- MIL + boosting

Train a boosting classifier that maximizes log likelihood of bags

$$\log L = \sum_i \log(p(y_i | X_i))$$

where,

$$p(y_i | X_i) = 1 - \prod_j (1 - p(y_i | x_{ij}))$$

[Viola et al. '05]
Previous Work: MILBoost

- Problem: need all training examples

[Viola et al. '05]
Previous Work: MILBoost

But in tracking, only current frame available

[Viola et al. ‘05]
Previous Work: MILBoost

But in tracking, only current frame available

Need an online training algorithm for MIL

[Viola et al. '05]
Main Contribution of this paper

• Online-MILBoost:
  Online training for MIL-based classifier

• MILTrack
  New tracking solution using Online-MILBoost
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches

$$X^s = \{x \mid < s^1 \text{ pixels from tracker location}\}$$

1: $s = 35$ in authors' experiment
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
   \[ X^s = \{ x \mid < s^1 \text{ pixels from tracker location} \} \]

1: \( s = 35 \) in authors’ experiment
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches

\[ X^s = \{ x \mid < s^1 \text{ pixels from tracker location} \} \]

1: \( s = 35 \) in authors’ experiment

Babenko et al., 09
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
   \[ X^s = \{ x \mid < s^1 \text{ pixels from tracker location} \} \]

1: \( s = 35 \) in authors' experiment
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
   \[ X^s = \{ x \mid < s^1 \text{ pixels from tracker location} \} \]

1: \( s = 35 \) in authors’ experiment
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
   \[ X^s = \{ x \mid < s \text{ pixels from tracker location} \} \]

2. Use MIL classifier to find new tracker location
   \[ l_{new} = l(\arg\max_{x \in X^s} P(y = 1|x)) \]

Babenko et al., 09
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
   \[ X^s = \{x| < s \text{ pixels from tracker location} \} \]

2. Use MIL classifier to find new tracker location
   \[ l_{\text{new}} = l(\text{argmax}_{x \in X^s} p(y = 1|x)) \]

3. 1) Crop positive examples
    \[ X^r = \{x| < r^1 \text{ pixels from tracker location} \} \]

1: \( r = 5 \) in authors’ experiment
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
   \[ X^s = \{ x | < s \text{ pixels from tracker location} \} \]

2. Use MIL classifier to find new tracker location
   \[ l_{\text{new}} = l(\arg\max_{x \in X^s} p(y = 1| x)) \]

3. 1) Crop positive examples
   \[ X^r = \{ x | < r^1 \text{ pixels from tracker location} \} \]

1: \( r = 5 \) in authors’ experiment
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
   \[ X^s = \{ x | < s \text{ pixels from tracker location} \} \]

2. Use MIL classifier to find new tracker location
   \[ l_{new} = l(\arg \max_{x \in X^s} p(y = 1|x)) \]

3. 1) Crop positive examples
   \[ X^r = \{ x | < r \text{ pixels from tracker location} \} \]

3. 2) Crop Negative examples
   \[ X^{r,\beta} = \{ x | r \text{ to } \beta^1 \text{ pixels away from tracker location} \} \]

1: \( \beta = 50 \) in authors’ experiment

Babenko et al., 09
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
   \[ X^s = \{ x \mid < s \text{ pixels from tracker location} \} \]

2. Use MIL classifier to find new tracker location
   \[ l_{\text{new}} = l(\arg\max_{x \in X^s} p(y = 1 \mid x)) \]

3. 1) Crop positive examples
   \[ X^r = \{ x \mid < r \text{ pixels from tracker location} \} \]

3. 2) Crop Negative examples
   \[ X^{r,\beta} = \{ x \mid r \text{ to } \beta \text{ pixels away from tracker location} \} \]

1: \( \beta = 50 \) in authors’ experiment

Babenko et al., 09
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
   \[ X^s = \{ x | < s \text{ pixels from tracker location} \} \]

2. Use MIL classifier to find new tracker location
   \[ l_{\text{new}} = l(\arg\max_{x \in X^s} p(y = 1 | x)) \]

3. Crop positive and negative examples near new object location

4. Online MILBoost:
   Update MIL classifier with positive and negative example bags

Babenko et al., 09
Online-MILBoost:

Image patch $x$

\[ f_1, f_2, f_3, \ldots \]

Babenko et al., 09
Online-MILBoost:

- $h_k$: a weak classifier using one feature

Image patch $x$

$$
\begin{align*}
&h_1(x) \\
&h_2(x) \\
&\vdots
\end{align*}
$$

Babenko et al., 09
Online-MILBoost:

- $h_k$: a weak classifier using one feature

$$h_k(x) = \log \left[ \frac{p(y = 1|f_k(x))}{p(y = 0|f_k(x))} \right]$$
Online-MILBoost:

- $h_k$: a weak classifier using one feature

\[
h_k(x) = \log \left[ \frac{p(y = 1|f_k(x))}{p(y = 0|f_k(x))} \right]
\]

with,

\[
\begin{align*}
p(f_k(x)|y = 1) & \sim \mathcal{N}(\mu_1, \sigma_1) \\
p(f_k(x)|y = 0) & \sim \mathcal{N}(\mu_0, \sigma_0) \\
p(y = 1) & = p(y = 0)
\end{align*}
\]
Online-MILBoost:

- $H(x)$: the MIL classifier made from weak classifiers

\[ H(x) = \sum_{k=1}^{K} h_k(x) \]

K = 50 in authors’ experiment

Babenko et al., 09
Online-MILBoost:

- Always keep a pool of $M >> K$ weak classifier candidates

$M = 250 \& K = 50$ in authors' experiment
Online-MILBoost:

- Update all M weak classifiers with positive and negative bags

\[(X_1, 1), (X_2, 0), (X_3, 0)\]

{Classifier \(h_1\), Classifier \(h_2\), ..., Classifier \(h_M\)}

Babenko et al., 09
Online-MILBoost:

- Pick best $K$ weak classifiers to form $H(x)$, where

$$h_k = \underset{h \in \{h_1 \ldots h_M\}}{\text{argmax}} \log L(H_{k-1} + h)$$

$$H(x) = \sum_{k=1}^{K} h_k(x)$$

where $H_{k-1}$ is the classifier made up of the first $k-1$ weak classifiers.
Online-MILBoost:

- Prediction: $p(y = 1|x) = \sigma(H(x))$

\[ \sigma(x) = \frac{1}{1 + e^{-x}} \]
Online-MILBoost:

- Prediction: $p(y = 1|x) = \sigma(H(x))$

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$h_1(x) = 2, \ h_2(x) = 1.8, \ h_3(x) = 0.6$

Babenko et al., 09
Online-MILBoost:

- Prediction: \( p(y = 1|x) = \sigma(H(x)) \)

\[
\begin{align*}
\sigma(t) &= \frac{1}{1 + e^{-t}} \\

H(x) &= \sum_{k=1}^{K} h_k(x) = 4.4 \\
p(y = 1|x) &= \sigma(H(x)) = 0.99
\end{align*}
\]

Babenko et al., 09

\[
\begin{align*}
h_1(x) &= 2 \\
h_2(x) &= 1.8 \\
h_3(x) &= 0.6
\end{align*}
\]
MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
   \[ X^s = \{ x | < s \text{ pixels from tracker location} \} \]

2. Use MIL classifier to find new tracker location
   \[ l_{new} = l(\arg\max_{x \in X^s} p(y = 1|x)) \]

3. Crop positive and negative examples near new object location

4. Online MILBoost:
   Update MIL classifier with positive and negative example bags

Babenko et al., 09
Experiments

Datasets: 8 publicly available videos,
- Grayscale, 320 x 240 pixels
- Ground truth labeled every 5 frames by hand

Babenko et al., 2009
Experiments

Compared with:

- **OAB1**
  Online AdaBoost w/ 1 positive example per frame

- **OAB5**
  Online AdaBoost w/ 45 positive examples per frame

- **SemiBoost**
  Label in 1st frame only.

- **FragTrack**
  Static appearance model

Babenko et al., 2009
Experiments

Evaluation criterion:

Tracker position error (pixels) w.r.t. Ground truth
Results

Video David
Results

Position error versus Frame #, Video David

Babenko et al., 2009
Results

Video Occluded Face
Results

Position error versus Frame #, Video Occluded Face

Babenko et al., 2009
## Results

### Average Center location errors (pixels)

<table>
<thead>
<tr>
<th>Video Clip</th>
<th>OAB1</th>
<th>OAB5</th>
<th>SemiBoost</th>
<th>Frag</th>
<th>MILTrack</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Indoor</td>
<td>49</td>
<td>72</td>
<td>59</td>
<td>46</td>
<td>23</td>
</tr>
<tr>
<td>Sylvester</td>
<td>25</td>
<td>79</td>
<td>22</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Occluded Face</td>
<td>44</td>
<td>105</td>
<td>41</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>Occluded Face 2</td>
<td>21</td>
<td>93</td>
<td>43</td>
<td>45</td>
<td>20</td>
</tr>
<tr>
<td>Girl</td>
<td>48</td>
<td>68</td>
<td>52</td>
<td>27</td>
<td>32</td>
</tr>
<tr>
<td>Tiger 1</td>
<td>35</td>
<td>58</td>
<td>46</td>
<td>40</td>
<td>15</td>
</tr>
<tr>
<td>Tiger 2</td>
<td>34</td>
<td>33</td>
<td>53</td>
<td>38</td>
<td>17</td>
</tr>
<tr>
<td>Coke Can</td>
<td>25</td>
<td>57</td>
<td>85</td>
<td>63</td>
<td>21</td>
</tr>
</tbody>
</table>

Babenko et al., 2009
Conclusion

• Online MILBoost:
  Online algorithm to update MIL-based classifier

• Performance of “MILTrack” is stable
Discussion

• Why it can handle occlusion?

• Possible improvements
  • Motion Model
  • Features
  • Part based representation
Note…

Wu et al., 2013
Wu et al., 2013
Note…

Project 2 use this!

Wu et al., 2013
Thank You!

Q&A