Spatio-Temporal Nonparametric Background Modeling and Subtraction

Raviteja Vemulapalli and R. Aravind
Department of Electrical engineering
Indian Institute of Technology, Madras
Background subtraction

- Technique used for separating moving objects (foreground) from rest of the scene (background) in a video.

**Figure:** Block diagram of background subtraction process
Nonparametric background model

- Current intensity of pixel \((x, y)\): \(I(x, y)\)
- Set of past intensity samples: \(\{I_1(x, y), I_2(x, y), \ldots, I_N(x, y)\}\)
- Intensity PDF of the pixel \((x, y)\) using a kernel \(K\).

\[
p' = p \left( \frac{I(x, y)}{\{I_i(x, y)\}_{i=1}^N} \right) = \frac{1}{N} \sum_{i=1}^{N} K(I(x, y) - I_i(x, y)).
\]

- Gaussian kernel

\[
p' = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{(2\pi)^{d/2} \left| \Sigma \right|^{1/2}} \exp \left\{-\frac{1}{2} [I(x, y) - I_i(x, y)]^T \Sigma^{-1} [I(x, y) - I_i(x, y)] \right\}
\]
Nonparametric background model

• For independent colour channels

\[ \Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} \]

\[ p' = \frac{1}{N} \sum_{i=1}^{N} \prod_{j=1}^{3} \frac{1}{\sqrt{2\pi \sigma_j^2}} \exp \left\{ -\frac{[I_j^i(x, y) - I_j(x, y)]^2}{2\sigma_j^2} \right\} \]

where \( I_j^i(x, y) \) denotes the \( j^{th} \) colour component of \( I_i(x, y) \).

• \( p' > thr \implies \) pixel \((x, y)\) is background.

• \( p' < thr \implies \) pixel \((x, y)\) is foreground.
Short term and Long term models

- **Short term model:**
  - Uses the most recent $N$ background samples.
  - New sample is added to the sample set only if it belongs to background.

- **Long term model:**
  - Uses $N$ samples taken from $W$ ($> N$) past samples.
  - Every new sample is added to the sample set irrespective of whether it belongs to background or foreground.
Foreground detection

- Combination of short term and long term models

Combination results (0 for background and 1 for foreground)

<table>
<thead>
<tr>
<th>Short Term Model</th>
<th>Long Term Model</th>
<th>Final Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>( O_{ST}(x, y) = 0 )</td>
<td>( O_{LT}(x, y) = 0 )</td>
<td>( O(x, y) = 0 )</td>
</tr>
<tr>
<td>( O_{ST}(x, y) = 0 )</td>
<td>( O_{LT}(x, y) = 1 )</td>
<td>( O(x, y) = 0 )</td>
</tr>
<tr>
<td>( O_{ST}(x, y) = 1 )</td>
<td>( O_{LT}(x, y) = 0 )</td>
<td>( O(x, y) = O'(x, y) )</td>
</tr>
<tr>
<td>( O_{ST}(x, y) = 1 )</td>
<td>( O_{LT}(x, y) = 1 )</td>
<td>( O(x, y) = 1 )</td>
</tr>
</tbody>
</table>

\[
O'(x, y) = \begin{cases} 
1 & \text{if } \sum_{i=-1}^{1} \sum_{j=-1}^{1} O_{ST}(x-i, y-j)O_{LT}(x-i, y-j) \neq 0 \\
0 & \text{else}
\end{cases}
\]
3 \times 3 \text{ blocks instead of individual pixels.}

\text{Block centered on } (x, y) \text{ in the current frame: } F(x, y)

\text{Set of past samples: } \{F_1(x, y), F_2(x, y), \ldots, F_N(x, y)\}

\text{Intensity PDF of this block using a kernel } K

\begin{equation}
 f = p \left( F(x, y) / \{F_i(x, y)\}_{i=1}^N \right) = \frac{1}{N} \sum_{i=1}^{N} K(F(x, y) - F_i(x, y))
\end{equation}

\text{Gaussian kernel}

\begin{equation}
 f = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} [F(x,y)-F_i(x,y)]^T \Sigma^{-1} [F(x,y)-F_i(x,y)]\right)
\end{equation}
Hyperspherical kernel

- Hyperspherical kernel instead of Gaussian to reduce computational complexity.

- Hyperspherical kernel of radius $r$

  \[
  K(u) = \begin{cases} 
  \frac{1}{V} & \text{if } |u| \leq r \\
  0 & \text{otherwise}
  \end{cases}
  \]

- Probability density becomes

  \[
  f = \frac{M}{N V}
  \]

  \[
  M = \sum_{i=1}^{N} \phi \left( \frac{\| F(x, y) - F_i(x, y) \|}{r} \right)
  \]

  \[
  \phi(u) = \begin{cases} 
  1 & \text{if } u \leq 1 \\
  0 & \text{otherwise}
  \end{cases}
  \]
Foreground detection

- \( f > thr \implies \text{pixel } (x, y) \text{ is background.} \)
- \( f < thr \implies \text{pixel } (x, y) \text{ is foreground.} \)
- Both short term and long term models used.
- Combination rules similar to nonparametric model.
- Tradeoff between computational efficiency and foreground detection.

- Hyperspherical kernel computationally simple compared to Gaussian kernel.
- Gaussian kernel gives better estimate of the PDF and hence better detection results compared to hyperspherical kernel.
Model for colour videos

- Color channels assumed to be independent.
- Each channel processed separately like a gray-scale video.
- $R(x, y)$, $G(x, y)$, $B(x, y)$ – Processing results of the three colour channels at pixel $(x, y)$ (0 for background and 1 for foreground).
- Final result

$$O(x, y) = R(x, y) \mid G(x, y) \mid B(x, y)$$
Results: Lab video

- Detection results for 346\textsuperscript{th} frame of lab video under different noise levels.

Figure: (a) Original frames corrupted by noise; (b) Detection results of the proposed model; (c) Detection results of nonparametric model for low threshold; (d) Detection results of nonparametric model for high threshold.
Results: Lab video

- Detection results for $375^{th}$ frame of lab video under different noise levels.

Figure: (a) Original frames corrupted by noise; (b) Detection results of the proposed model; (c) Detection results of nonparametric model for low threshold; (d) Detection results of nonparametric model for high threshold.
Results: Lab video

Detection results for 408\textsuperscript{th} frame of lab video under different noise levels.

Figure: (a) Original frames corrupted by noise; (b) Detection results of the proposed model; (c) Detection results of nonparametric model for low threshold; (d) Detection results of nonparametric model for high threshold.
Results: Lab video

- Detection results for $434^{th}$ frame of lab video under different noise levels.

![Images of original frames and detection results](image)

Figure: (a) Original frames corrupted by noise; (b) Detection results of the proposed model; (c) Detection results of nonparametric model for low threshold; (d) Detection results of nonparametric model for high threshold.
Results: Crowd video

- Detection results for 253rd, 332nd and 340th frames of crowd video.

Figure: (a) Original frames corrupted by noise; (b) Detection results of the proposed model; (c) Detection results of nonparametric model for low threshold; (d) Detection results of nonparametric model for high threshold.
Results: Crowd video

- Detection results for 350\textsuperscript{th}, 378\textsuperscript{th} and 381\textsuperscript{st} frames of crowd video.

Figure: (a) Original frames corrupted by noise; (b) Detection results of the proposed model; (c) Detection results of nonparametric model for low threshold; (d) Detection results of nonparametric model for high threshold.
Results: Bottle video

- Detection results for 251\textsuperscript{st}, 291\textsuperscript{st} and 300\textsuperscript{th} frames of crowd video.

![Original frames and detection results](image)

**Figure:** (a) Original frames; (b) Detection results of the proposed model; (c) Detection results of nonparametric model for low threshold; (d) Detection results of nonparametric model for high threshold.
Results: Bottle video

Detection results for $302^{nd}$, $306^{th}$ and $328^{th}$ frames of bottle video.

Figure: (a) Original frames; (b) Detection results of the proposed model; (c) Detection results of nonparametric model for low threshold; (d) Detection results of nonparametric model for high threshold.
Results: Ducks video

- Detection results for 292\textsuperscript{nd}, 302\textsuperscript{nd} and 307\textsuperscript{th} frames of ducks video.

Figure: (a)Original frames; (b)Detection results of the proposed model; (c)Detection results of nonparametric model for low threshold; (d)Detection results of nonparametric model for high threshold.
Results: Ducks video

- Detection results for $340^{th}$, $363^{rd}$ and $372^{nd}$ frames of ducks video.

Figure: (a) Original frames; (b) Detection results of the proposed model; (c) Detection results of nonparametric model for low threshold; (d) Detection results of nonparametric model for high threshold.
Conclusion and future work

Advantages of the proposed approach:

- Computationally simple.
- Performs very well even in the cases of noisy videos and dynamic backgrounds.

Future work

- Considering all colour channels and using incremental PCA to avoid dealing with 27-dimensional data.
- Making the parameters (kernel radius and threshold) adaptive.
Applications

- Traffic monitoring
- Human motion estimation and tracking
- Human activity recognition
- Abandoned object detection
- Crowd density and behaviour estimation
- Silhouette extraction for object recognition
