Abstract

We propose a real-time anomaly detection system for video streams. Spatio-temporal features are exploited to capture scene dynamic statistics together with appearance. Anomaly detection is performed in a non-parametric fashion, evaluating directly local descriptor statistics. A method to update scene statistics, to cope with scene changes that typically happen in real world settings, is also provided. The proposed method is tested on publicly available datasets.

1. Motivation

- To render scalable the evergrowing amount of human operated CCTV surveillance systems automatic real-time video analysis has to be performed.
- Detect and localize anomalous events to warn operators or drive the system to a more specific high-level task such as active PTZ tracking, face detection and logging or action recognition.
- Existing approaches rely on trajectory[5] or optical flow[4, 2] features which can detect only a subset of the anomalies present in an image.

2. Non-parametric Anomaly Detection

![Image of system overview]

Figure 1: System overview. Each cell features are stored on efficient k-means tree based indexes. Planes underneath represent a simplified view of the high dimensional feature space; dashed circles are plotted at the optimal radius value.

Learning the scene

Our approach is purely non parametric and exploits robust space-time descriptors.

- We extract features with a regular grid over the image and store them in fast nearest-neighbor search structures.
- At run-time, given an optimal radius \( r \), we perform a range query for each feature extracted at each image location. In absence of neighbors we consider it an anomaly.
- The parameter \( r \) is estimated using a simple data-driven technique. All nearest neighbor distances \( d\_i \) are stored and a cumulative distribution function \( CDF(d) \) is estimated for each cell. Given a probability \( p\_i \) below which we consider an event anomalous (e.g. \( 10^{-3}, 10^{-4} \ldots \) ), we compute

\[
  r = CDF^{-1}(1 - p\_i).
\]

Model update

Since this kind of anomaly detection applications are thought to be run for a long time, it is very likely that a scene will change its appearance over time. Again a very straightforward data-driven technique is used.

- Anomalous patterns detected at each location are stored in an abnormality list.
- Model update is performed learning an optimal radius \( r\_a \) for the abnormality list.

![Diagram of K-means tree]

3. Space-time features

Space-time volumes extracted on the regular grid are represented as in the following. To compute the representation of each volume we define a descriptor \( d \) based on three-dimensional gradients. The 3D gradient magnitude and orientations are:

\[
  M_{3D} = \sqrt{G_x^2 + G_y^2 + G_z^2},
\]

\[
  \phi = \tan^{-1}(G_y / \sqrt{G_x^2 + G_z^2}),
\]

\[
  \theta = \tan^{-1}(G_z / G_x).
\]

Each volume is divided in 18 sub-regions (three along each spatial direction and two along the temporal). \( \phi \) and \( \theta \) values are accumulated in local histograms of 4 and 8 bins respectively.

4. Results

![Image of results]

Figure 2: Anomaly detection results on UCSD dataset. Detected anomalies are skaters, bikers and trolleys or vehicles (see Fig. 4). Our system also detects a wheelchair and people walking off the walkway as anomalous patterns.

Figure 3: Spatial anomaly localization varying the spatial cuboid extension (20, 30, 40, 60 pixels).

Figure 4: Qualitative comparison with methods presented in [3, 4]: our method, mixture of dynamic textures, social force and mixture of principal components analyzers, social force only.

![ROC curve]

Figure 5: ROC curve to compare of our method with state-of-the-art approaches. Our system outperforms other real-time approaches and runs at 20 fps on a standard machine. State-of-the-art method [3] requires 25 seconds to process a frame.

References