

Systems and Methods for Detecting Anomalies from Data

Cross Reference

[0001] The present application claims priority from Australian Provisional Patent Application No. 2009901406 entitled "Detecting Anomalous Behaviour" filed on April 1, 2009; and Australian Provisional Patent Application No. 2009905937 entitled "Detecting Anomalies from Data" filed on December 4, 2009. The entire content of each of these applications is hereby incorporated by reference. In addition the entire content of each of the following references is incorporated by reference: <http://www.abilene.iu.edu>; ~~A. Lakhina et al.~~ A. Lakhina, M. Crovella, and C. Diot. Diagnosing Network-Wide Traffic Anomalies. In Proceedings of ACM SIGCOMM, pp. 219--230 2004; M. Elad. Optimized Projections for Compressed Sensing. IEEE Transactions on Signal Processing, 55(12): 5695-5702, 2007; and S. Budhaditya, D-S. Pham, M. Lazarescu, and S. Venkatesh. Effective Anomaly Detection in Sensor Networks Data Streams. In Proceedings of the Ninth International Conference on Data Mining, Miami, Florida, pp. 722--727, 2009. ~~Diagonising network wide traffic anomalies. In Proc. ACM SIGCOMM, 2004;~~ M. Elad. Optimized projections for compressed sensing. IEEE Trans. Sig. Process., 55:5695-5702, 2007 and Effective Anomaly Detection in Sensor Networks Data Streams presents at the December 2009 Ninth IEEE International Conference on Data Mining, Miami Florida, S. Budhaditya et. al.

Technical Field

[0002] The present disclosure concerns the detection of anomalies or infrequent events from data. In certain embodiments this detection may be carried out in an automatic or substantial automatic manner and if desired on large-scale data. Certain applications are directed to analysing video surveillance records to identify aberrant behaviour among small numbers of individuals in a crowd. Certain applications are directed to analysing data streams from large-scale sensor networks and/or large on-line databases. In certain embodiments, the disclosure is to methods and/or systems of detecting anomalies from data; including large-scale data. Certain applications may be carried out in real time or substantially real time.

Background

[0003] Anomalous event detection from video data is a complex problem that has been the subject of investigation over the past decade. The number of surveillance cameras in the public domain is large, and as a result it is practically often impossible to manually monitor the video streams from all the cameras. As a result there is a pressing need for automatic, or substantially automatic, and/or scalable detection of abnormal events from video streams. There is also a pressing need to for being able to do this in substantially real time, in realtime, or in an acceptable time period.

Claims

1. A method for processing, detecting and/or notifying for the presence of at least one infrequent event from at least one large scale data set comprising:
 - receiving time series data;
 - representing either the time series data, or one or more features of the time series data, as sets of vectors, matrices and/or tensors;
 - performing compressive sensing on ~~the~~ at least one vector, matrix and/or tensor set;
 - decomposing the at least one ~~selected sets of~~ compressive sensed vectors, ~~matrices~~ matrix and/or tensors set to extract ~~the~~ a residual subspace; and
 - identifying, using a computing device, potential infrequent events by analysing compressive sensed data projected into the residual sub-space.
2. A method according to claim 1, wherein the one or more features of the time series data are one or more variables associated with a subset of the time series data.
3. A method according to claim 1, wherein the one or more features of the time series data are the difference in any variable between any two subsets of the data of the time series data.
4. A method according to claim 3, wherein the time series data is a stream of at least one of video and audio surveillance data, each subset is a frame and the feature is a motion or frequency information between each successive pair of frames.
5. A method according to claim ~~4~~3, wherein the time series data is a stream of video data, each subset is a frame and the feature is motion information between each successive pair of frames, the motion information being determined by the analysis (of optical flow) ~~involves~~ dividing each frame of video by a grid into cells, and then counting ~~the~~ a number of optic flow vectors in each cell to provide optical flow information.
6. A method according to claim 5, wherein the optical flow information is represented using bag-of-visual-words.
7. A method according to claim 6, wherein from the bag-of-visual-words a feature-frame matrix is constructed by amalgamating the feature vectors for all frames in the sequence.

8. A method according to claim 7, wherein the feature-frame matrix is then structurally decomposed into the observed into principal and residual components.
9. A method according to claim 8, further comprising detecting abnormal events in the residual subspaces using ~~the~~ a Q-statistic based test statistic.
10. A method according to claim 1, wherein a pre-cursor step is employed to transform the data to the Compressed Domain, to reduce the data to a manageable size.
11. A method according to claim 10, wherein the pre-cursor step of transformation transforming the data to the compressed domain is deployed for reducing ~~the~~ a feature dimension in the compressed domain and employs a sensing matrix having entries with values of either 0 with probability 2/3 or +/-1 with probability 1/6.
12. A method according to claim 11, wherein a random gossip algorithm is applied.
13. A method according to claim 10, wherein the pre-cursor step of transformation transforming the data to the compressed domain is deployed for reducing ~~the~~ a time instances in the compressed domain and employs frame sub-sampling.
14. A method according to claim 10, wherein the features are represented as sets of vectors before being transformed to the Compressed Domain.
15. A method according to claims 1-13 or 14, wherein said identifying potential infrequent events involves thresholding the data projected into the residual sub-space.
16. A system comprising:
 - at least one sensor to receive data comprising a time series of data subsets for analysis;
 - a computer memory for storing the data;
 - a computer processor for:
 - representing either the time series data, or one or more features of the time series data, as sets of vectors, matrices and/or tensors;
 - performing compressive sensing on the at least one vector, matrix and/or tensor set;
 - decomposing the selected sets of compressive sensed vectors, matrices and/or tensors to extract ~~the~~ a residual subspace; and

identifying potential infrequent events by analysing compressive sensed data projected into the residual sub-space.

17. The system of claim 16, wherein the computer processor is co-located with the at least one sensor.

18. The system of claim 16, wherein the computer processor is remotely located from the at least one sensor.

19. The system of claim 16, wherein a portion of the computer processor is remotely located from the least one sensor and performs the step of identifying potential infrequent events by analysing compressive sensed data projected into the residual sub-space.

20. The system of claim 16, wherein the functionality of the computer processor ~~can be~~is divided between a sub-processor co-located with the at least one sensor and a sub-processor remotely located from the at least one sensor.