Poster: Challenges with Image Event Processing

Asra Aslam
Insight Centre for Data Analytics, NUI Galway
asra.aslam@insight-centre.org

Souleiman Hasan
Lero- The Irish Software Research Centre, NUI Galway
souleiman.hasan@lero.ie

Edward Curry
Insight Centre for Data Analytics, NUI Galway
edward.curry@insight-centre.org

ABSTRACT

There has been substantial research in the area of event processing where systems are focused on event processing of structured data. However, in the context of smart cities, significant number of real-time applications for event-driven systems consist of image data, rather than structured events. Therefore, there is a need for a system that can process multimedia events such as images. This paper discusses challenges with processing images within event-based systems.

CCS CONCEPTS

•Information systems → Multimedia streaming; •Software and its engineering → Publish-subscribe / event-based architectures;

KEYWORDS

Image Event Processing, Smart Cities, Event-Based Systems, Internet of Things

1 INTRODUCTION

With the evolution of concept of smart cities, event based systems are introduced to serve as a middle-ware between the Internet of Things and the applications layer [1]. Existing event-based systems process the subscription of a user based on standard languages for rules or queries which conform to nature of structured events. However, much of the data in real world can be non-structured and may be in the form of images or videos. There is a need for a system which can take input in the form of images/videos and process it according to the event-based paradigm. The contribution of this paper is to highlight the challenges to build a system for image event processing.

2 PROBLEM STATEMENT

How to design an intelligent event driven system, which can process multimedia events, and react to users situations of interest with a low latency. Ideally, the system will be responsible for facilitating smooth interactions between three main entities: *subscriber*, *image producer* and *image analysis*. To achieve this goal consider an example of object detection in real-time images (Fig. 1). Objects such as 3 cars and 4 persons are detected in the image by using the *yolo* model [8, 9] from a frame of MIT traffic dataset [12]. Object detection can be used in applications like adaptive traffic light switching, car parking, pedestrian detection, etc. The system should abstract image analysis tasks, including object detections and bring them to the core of the event-based engine.

3 RELATED WORK

The main entities of event-driven systems are producers, consumers and a network for communication [3]. Existing event-driven systems allow the interaction via structured queries in response to events. The Content-based video query language (CVQL) is an extended version of existing query languages, by allowing the query of video databases [6]. It is designed for specifying spatial and temporal relationships. Structure-based video query language (SVQL) is more expressive as it includes variable declaration, structure specification, feature specification and spatial-temporal specification [7]. Both CVQL and SVQL are applicable only for videos, but they need to be improved with more operators to handle real-time complex multimedia events. In existing multimedia query languages, the MPEG Query Format (MPQF) can be used as a standard interface for multimedia retrieval engines [2].

The work in [10] presents an event-based surveillance system which involves high-level image processing, but it is only applicable for the detection of unusual events for an airport environment. An event-driven smart city architecture is presented in [4] for the management and interaction of different sensors mounted at multiple public places. A neural image caption generator which processes both text and images has also been proposed [11] to describe the images and convert them into text. Thus, there is a need for a multi-purpose system that can handle generic multimedia events.

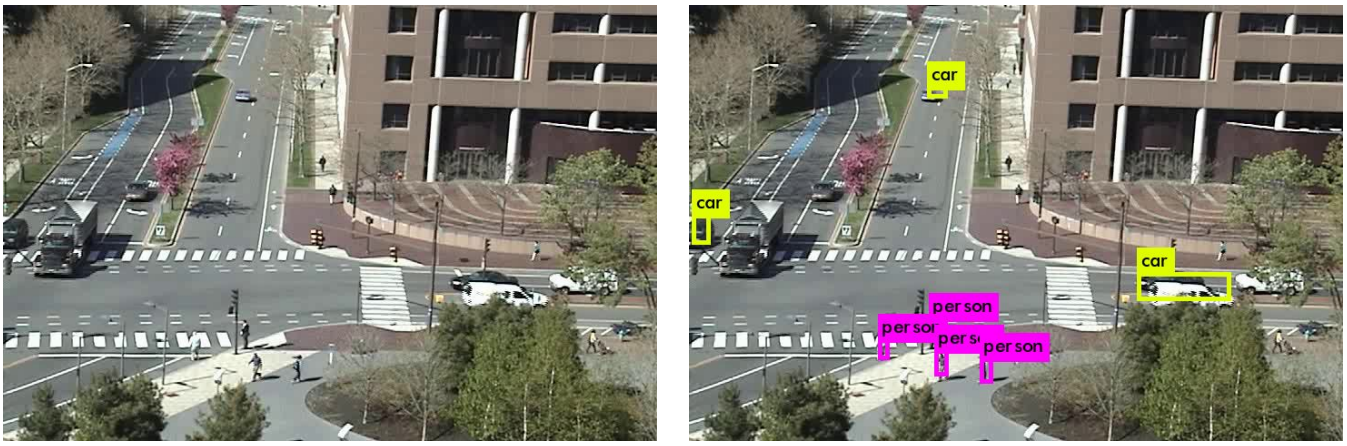
4 CHALLENGES

Challenges for image event processing includes:

Domain and Range of Textual Descriptions: Making the range of possible image classification outputs equal to the domain of the user is a challenge. For example in the current scenario shown in Fig. 1, if we want to match the word "car" then the system will make a match, but if the subscriber subscribes to "automobile", then it will not be able to identify the object, because it does not have the classifier for automobiles. Besides, it does not have any knowledge-base to recognize that an automobile is a superset of a car. The objective of our system should be to increase the *range* to such an extent that it can be equal to *domain*, and it can respond to any word that the subscriber might use. Using some knowledge bases can allow the system to find synonyms and sub or super classes. Semantic matching is also a possible solution [5].

DEBS '17, Barcelona, Spain

ACM ISBN 978-1-4503-5065-5/17/06.
DOI: http://dx.doi.org/10.1145/3093742.3095095



(a) (b)

Figure 1: An example image from the MIT Traffic Dataset (a) Original frame, and (b) Objects Detected by Designed System

Multimedia Data: Handling and availability of usable multimedia data is also a challenging task of the system at various levels. For example in the case of object detection, classifiers are required for labeling objects. However, image databases with annotations prerequisite to build a classifier are needed.

For evaluations, constructing results in terms of comparative accuracy and loss, ground truth datasets are needed. Content, size and quality of images received from the sensors will also directly impact the latency and accuracy of the results. By using existing databases like ImageNet, Pascal VOC, and COCO can be a good start.

Quality of Image Analysis to Detect Events: The analysis results will depend on many factors like training databases and categories. The efficiency and accuracy of existing image processing algorithms will ultimately decide the limitations of the system. High performance specialized models, like *darkflow* for object detection, can be used to address this. These models will require pre-trained classifiers with annotated databases.

Classifiers: Practically, constructing an infinite number of classifiers for detecting all types of images, is not possible. On-line training of classifiers at runtime may largely affect the processing time. On the other hand, having information available about all the categories prior to processing is likely unfeasible to achieve. So, at the application level, we are compelled to use pre-trained classifiers for common categories and automate the process of training for newly subscribed categories. Construction of classifiers in an efficient way is itself a challenging task. Continuous stream machine learning and crowd tagging can be used to address this challenge.

Optimization for Multiple Subscribers: At the subscriber level, in the case when multiple subscribers who are looking for the same event or object (e.g. both want to detect a "car"), the event should only be analyzed once. Also, if the subscriptions have synonyms, then knowledge-base concept hierarchies could also be used for increasing the performance. Providing suggestions to subscribers while querying can also help in optimizing the process.

ACKNOWLEDGMENTS

This work was supported, by *Science Foundation Ireland* under grant SFI/12/RC/2289 INSIGHT and grant 13/RC/2094.

REFERENCES

[1] Edward Curry, Schahram Dustdar, Quan Z Sheng, and Amit Sheth. 2016. Smart cities–enabling services and applications. *Journal of Internet Services and Applications* 7, 1 (2016), 6.
[2] Mario Döller, Ruben Tous, Matthias Gruhne, Kyoungro Yoon, Masanori Sano, and Ian S Burnett. 2008. The MPEG Query Format: Unifying Access to Multimedia Retrieval Systems. *IEEE MultiMedia* 15, 4 (2008), 82–95.
[3] Patrick Th Eugster, Pascal A Felber, Rachid Guerraoui, and Anne-Marie Kermarrec. 2003. The many faces of publish/subscribe. *ACM Computing Surveys (CSUR)* 35, 2 (2003), 114–131.
[4] Luca Filipponi, Andrea Vitaletti, Giada Landi, Vincenzo Memeo, Giorgio Laura, and Paolo Pucci. 2010. Smart city: An event driven architecture for monitoring public spaces with heterogeneous sensors. In *Fourth Int. Conf. on Sensor Technologies and Applications (SENSORCOMM)*. IEEE, 281–286.
[5] Souleiman Hasan and Edward Curry. 2015. Thingsonomy: Tackling variety in internet of things events. *IEEE Internet Computing* 19, 2 (2015), 10–18.
[6] Tony CT Kuo and Arbee LP Chen. 2000. Content-based query processing for video databases. *IEEE Transactions on Multimedia* 2, 1 (2000), 1–13.
[7] Chenglang Lu, Mingyong Liu, and Zongda Wu. 2015. Svql: A sql extended query language for video databases. *International Journal of Database Theory and Application* 8, 3 (2015), 235–248.
[8] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 779–788.
[9] Joseph Redmon and Ali Farhadi. 2016. YOLO9000: Better, Faster, Stronger. *arXiv preprint arXiv:1612.08242* (2016).
[10] Chiao-Fe Shu, Arun Hampapur, Max Lu, Lisa Brown, Jonathan Connell, Andrew Senior, and Yingli Tian. 2005. *IBM* smart surveillance system (*S3*): a open and extensible framework for event based surveillance. In *IEEE Conf. on Advanced Video and Signal Based Surveillance*. IEEE, 318–323.
[11] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2015. Show and tell: A neural image caption generator. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*. 3156–3164.
[12] Xiaogang Wang, Xiaoxu Ma, and W Eric L Grimson. 2009. Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. *IEEE Transactions on pattern analysis and machine intelligence* 31, 3 (2009), 539–555.