IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 30, 2018, accepted March 12, 2018, date of publication April 5, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2823590

Towards a Generalized Approach for Deep Neural Network Based Event Processing for the Internet of Multimedia Things

ASRA ASLAM AND EDWARD CURRY
Insight Centre for Data Analytics, National University of Ireland Galway, H91AEX4 Galway, Ireland
Corresponding author: Asra Aslam (asra.aslam@insight-centre.org)

This work was supported by the Science Foundation Ireland under Grant SFI/12/RC/2289 INSIGHT.

ABSTRACT Event processing systems serve as a middleware between the Internet of Things (IoT) and the application layer by allowing users to subscribe to events of interest. Due to the increase of multimedia IoT devices (i.e. traffic camera), the types of events created are shifting more toward unstructured (multimedia) data. Therefore, there is a growing demand for efficient utilization of effective processing of streams of both structured events (i.e. sensors) and unstructured multimedia events (i.e. images, video, and audio). However, current event processing engines have limited or no support for unstructured event types. In this paper, we described a generalized approach that can handle Internet of Multimedia Things (IoMT) events as a native event type in event processing engines with high efficiency. The proposed system extends event processing languages with the introduction of operators for multimedia analysis of unstructured events and leverages a deep convolutional neural network based event matcher for processing image events to extract features. Furthermore, we show that neural network based object detection models can be further optimized by leveraging subscription constraints to reduce time complexity while maintaining competitive accuracy. Our initial results demonstrate the feasibility of a generalized approach toward IoMT-based event processing. Application areas for generalized event processing include traffic management, security, parking, and supervision activities to enhance the quality of life within smart cities.

INDEX TERMS Internet of Multimedia Things, Smart Cities, Event-Based Systems, Internet of Things, Multimedia Stream Processing, Distributed Systems, Smart Environments.

I. INTRODUCTION

The Internet of Things (IoTs) is designed to support intelligent systems. Event based systems are introduced to serve as a middleware between the Internet of Things and its applications. They are based on tracking and analyzing (processing) streams of information (structured data) about things that happen (events) and deriving a conclusion from them [1]. Publish/subscribe systems form the basis for event processing systems to facilitate smooth interactions between publishers and subscribers. Event processing systems are designed to process the subscription of a user in the form of a language for rules and query the structured events [2]. In addition to structured data, diverse sensors within smart cities are producing huge amounts of unstructured (multimedia) data. For instance, processing readings from temperature or energy sensors are examples of structured events and events produced by traffic cameras relate to traffic awareness, like traffic light switching, car parking, pedestrian detection, etc., belongs to unstructured events. Due to the lack of support for processing multimedia events in existing event-based systems, there is the need for an Internet-of-Multimedia-Things (IoMT) based event processing system which can also process images/videos. IoMT can be defined as an IoT-based paradigm which allows objects to connect and exchange structured as well as unstructured data with one another to facilitate multimedia-based services and applications within smart cities [3].

Image processing is a relevant domain in the context of smart cities and object detection is a common challenge in this area [4]. Most of the existing real-time image-based systems are efficient in recognizing objects while having domain specific characteristics. Deep convolutional neural networks [5] are proven to be suitable for the application of image classification in achieving high performance results.

As deep learning has made significant progress in the field of image recognition, it is desirable to bring its capabilities to identify objects, in smart cities infrastructure.

Existing event processing systems [6] are not designed to consume unstructured event types produced by the IoMT-generated events and there is no provision of multimedia query languages for events. At the same time, image processing systems [7], [8] are focused on handling real-time data and are highly efficient in recognizing objects in image events, but most are domain specific, and are not constructed by considering the event-based paradigm. Since there is limited provision of any query language in image processing systems, user expressiveness is also very low. Currently, developers have to build an application every time to merge the results of event processing and image processing systems for the efficient consumption of multimedia event data, which ultimately requires high setup cost with challenges in merging the results between both systems. Thus, there is a need for an IoMT aware event processing engine to support adaptive multimedia services between heterogeneous event environment and applications within smart cities.

To realize the goal of processing and analyzing IoMT based events as native event type with high performance, an event based Multimedia Stream Processing Engine (MSPE) is proposed along with an optimization technique, consisting of neural network based feature extraction operators, which extends event query language to support multimedia analysis within event-based systems. Within MSPE the user can define subscriptions using the proposed *detect* operator based on object detection. The subscription is used by the object detection model to choose the relevant classifier which is needed to identify the prescribed attribute. We proposed an optimization model which uses subscriptions at two different stages: (1) analyzing the query, and (2) optimize the processing of a neural network based matcher based on subscriptions constraints. The resulting model is proficient in processing multimedia event streams belongs to multiple applications while achieving high throughput and comparable accuracy which also confirms its generalizability within smart cities infrastructure.

The contributions of this paper can be summarized as follows:

- An architecture for a multimedia stream processing engine with neural network based event matcher using feature extraction.
- The "detect" operator in event query languages to support object detection in multimedia events.
- A neural network based matcher optimized based on subscription constraints.

The remainder of the paper is organized as follows. An introduction of IoMT with brief description of IoT is presented in Section–II. An outline to motivate the problem formulation is presented in Section–III. A brief overview of existing techniques with background are presented in Section–IV. Details of the proposed approach with implementation are discussed in Section–V. Simulation results and comparisons are given in Section–VI. Lastly, concluding remarks and avenues for future work appear in Section–VII.

II. INTERNET OF MULTIMEDIA THINGS (IOMT)

Traditionally, Internet of Things (IoT) can be defined as a "network of interconnected objects embedded with sensors to acquire information and to make this information available on the Internet" [9]. The basic idea of IoT is to empower the things so that they can gather information, understand each other and are able to interact with one another to reach the common goal of an intelligent system [10], [11]. Technological advancement in the field of Radio-Frequency IDentification (RFID), sensing and actuating capabilities is an important step towards the direction of IoT which enables the understanding of the world without the intervention of humans. The evolution of IoT from RFID tags to sensors and actuators and then towards virtualization of real-world objects, confirms the potential of IoT to change the physical objects into smart objects [9]. Much of the research on IoT is focused on improving the efficiency of handling large scale real-time data with limited consideration for the challenges of multimedia processing. The need to enable smart objects with the power of observing, sensing and understanding of the world through multimedia data in an efficient way, shifts the focus of research from conventional IoT to multimedia-based IoT [12]–[16].

The Internet of Multimedia Things (IoMT) is the term coined recently to represent multimedia communications using IoT. IoMT can be defined as an IoT-based paradigm which allows objects to connect and exchange structured as well as unstructured data with one another to facilitate multimedia-based services and applications [3]. As the concept of integrating multimedia with IoT is very recent, it has been referred in the literature with IoMT and MIoT synonymously, and needs standardization. In literature [3], [9], [17], IoMT has also been realized as additional challenges over IoT which may include Security, Routing, Quality of Service (QoS) and Quality of Experience (QoE) concerns, Heterogeneity of multimedia sensors etc. Layered based protocols have been presented for handling the challenges of QoE associated with concept of Multimedia IoT (MIoT) and analyzed using IoT based vehicle application [18], [19]. Similarly, challenges related to low cost data acquisition [20] and efficient cloud based transmissions [21], including robust multicast routing [22], have also been realized due to high bandwidth requirements of multimedia applications in the Internet of Things [23].

Table 1 represents a comparison of IoT and IoMT based technologies on the basis of a layered architecture [10]. It includes distinguished functionalities across sensing, network, service and application layer that should be taken into consideration for analyzing Internet of Things with/without Multimedia data. The high power consumption of multimedia sensor nodes as compared to scalar sensor nodes is one of the most important factors that affects efficient multimedia

TABLE 1. Comparison of IoT and IoMT based systems.

Characteristics	Internet of Things (IoT)	Internet of Multimedia Things (IoMT)
Resources	Low Energy Consumption	High Energy Consumption
Deployment	RFID tags (one-time or application dependent)	Video and Audio Sensors
Data Heterogeneity	Limited Heterogeneity (Scalar data)	Heterogeneous (Multimedia data)
Quality of Service (QoS)	Low Bandwidth	High Bandwidth
Data and Signal Processing	Scalar data to knowledge (Structured Data Analytics)	Multimedia data to knowledge (Multimedia Analytics)
Communication Protocols	Standardized	Non-Standardized
Service Composition	SOA-based and event-based middleware	No available specialized middleware

communications over IoT. In existing research, IoT focuses on the sensing, actuating and networking techniques relying on deployed RFID tags which is moving/shifting towards video and audio sensors for the processing of multimedia traffic. Scalar Sensors data coming from various sources of temperature, light, humidity, etc., are relatively structured with limited heterogeneity as compared to unstructured (multimedia) data. The fixed bandwidth of IoT based systems, which is usually low due to the expectation of small size packets, is not sufficient to analyze high bandwidth multimedia data. Moreover, communication protocols are also not standardized for multimedia data which associate challenges in designing of multimedia based applications over existing Wireless Sensor Networks (WSNs). Multimedia streaming (especially videos) is one of the most common type of events within applications of smart cities. However, it is clear from the analysis that existing approaches lack support for specialized middleware [12] to provide services required by IoMT-based applications.

III. MULTIMEDIA STREAM PROCESSING

A. PRELIMINARIES

- Information Flow Processing (IFP) refers to an application domain in which information is collected from multiple producers (distributed sources *i.e.* sensors) and processed to extract relevant information within timeliness [1]. Examples of IFP applications may include environmental monitoring, financial applications, fraud detection, inventory management, manufacturing control systems etc.
- Sensor Networks consisting of sensing nodes communicating in a dynamic environment to exchange information which can be in the form of scalar (structured) or multimedia (unstructured) data. Conventionally Radio Frequency IDentification (RFID) tags form the basis of wireless sensor networks.
- Middleware provides general-purpose services between distributed (multiple) platforms and domain-specific applications.
- Publish/Subscribe is a message-oriented interaction paradigm in which publishers send messages to the middleware, and the consumers express their particular interest for receiving messages [1], [2].
- Stream Processing refers to the processing of data streams with provision of querying active databases arriving in arbitrary order and perform actions on appearance of specific conditions, by considering time and size constraints. Stream processing evolved into event processing systems for association of notifications to subscriptions on matching of events, which can be further explained with the following entities:
  - Events: An event is anything that happens and is the significant observable occurrence.
  - Subscriptions: Subscription is a registration and association of an event action to indicate that a particular event is of interest to the user. Usually, it is an expression of subscriptions to match an event, if all primitive constraints are satisfied.
  - Matcher: The matcher is responsible for analyzing events and communicating them according to relevant subscriptions. In an event paradigm, the matcher will detect single events or pattern of events depending on the complexity of the matcher.

B. MOTIVATION

Recently, leveraging Internet of Things (IoT) for the processing of information related to various large-scale real-time data processing applications is becoming a popular trend to realize the goal of smart cities. However, the multimedia communication is gradually becoming an important source of information in multiple scenarios including traffic management, security, supervision activities, terrorist attacks and natural hazards. In the current scenario, existing multimedia-based communication technologies are domain specific and research on IoT mainly focuses on handling the challenges of big data excluding multimedia, leaving a gap between the advancement of IoT and multimedia-based technologies. The enormous generation of multimedia data within smart cities belonging to increasing number of applications, imposes a requirement of efficient handling of multimedia based events.

1) MOTIVATIONAL SCENARIOS

Since image recognition is the most common challenge in the context of smart cities, consider scenarios of object detection (shown in Fig. 1) for analyzing real-time multimedia events. Information associated with public transport and parking

(a) (i) Bus present; and (ii) Bus absent in image event

(b) (i) Car present; and (ii) Car absent in image event

FIGURE 1. Multimedia Processing Events Scenario.

related events can be further utilized with the following subscribed query statements:

Example 1 (Public Transport Management): Suppose a user subscribes for the status of "Bus" as "*Is a Bus on the Bus Stand?*". This type of query can be answered from a camera observing the bus stand (see Fig. 1a) and producing multimedia events consisting of bus status related information. In the first image, the bus has already arrived at the bus stand and in the second there is no bus. If we have a system to answer such types of queries then it will be easy for users to monitor such a situation.

Example 2 (Car Parking): If a user subscribes for car parking events such as "*Is Parking spot empty?*", this type of query can be answered using multimedia events (shown in Fig. 1b) related to the parking status of one parking lot with the presence and absence of the car.

2) CURRENT APPROACHES

In the current framework, real-time image events within smart cities are analyzed using image processing systems by adding the capabilities of event based systems. Based on real time video analysis, a real time event detector [24] is constructed for each action of interest by learning a cascade off filters based on volumetric features that efficiently scan video sequences in space and time. The presented system follows the model based approach for event detection, for constructing a framework to analyze videos efficiently. Another technique [25] targets video stream for the analysis of moving objects in the scene. It also provides a configuration for event detection and behavior analysis of video-surveillance streams. Due to the low expressiveness for user query and lack of provision of publish-subscribe paradigm, the merging of event based systems with image processing systems is an essential requirement. However, the need of integration of event based model for the processing of videos, limits the performance of the system and also requires high setup cost. Similarly video event monitoring systems [7], [8] [26] have been developed for real-time traffic management in smart cities. These systems are focused on handling the multimedia (unstructured) events with the aim of having high efficiency, but most of them are domain specific and cannot be generalized for multiple applications.

Fig. 2 illustrates the current implementation of processing multimedia events within smart cities. It consists of two key components: image processing system and event processing system to handle IoMT based events. Image based systems are responsible for analyzing images/video using domain specific feature extraction methods. On the other hand, event based systems provide the publish-subscribe paradigm to facilitate distributed interaction in large-scale applications. Thus, these two components need to be merged using another application to support user request across both systems.

3) TECHNICAL LIMITATIONS

Due to the lack of middleware support between heterogeneous environment and its applications, merging of image and event processing systems is the best possible solution for efficient handling of adaptive multimedia events. In order to accomplish the task of combining both types of systems, developers have to build dedicated applications for the respective domain which suffer from the following limitations:

- Need for an application to process IoMT generated data with change of application domain.
- High setup cost of combining image and event processing systems, which hinders the growth of multimedia based applications and thus resulting the main cause of restricted consumption of multimedia data.
- Variance in performance due to repetitive implementation of merging procedures with respect to application domains having different requirements/criteria.

Practically, it is not possible to design an application each time with change of application domain for the efficient consumption of multimedia data, thus it is necessary to develop a generalized approach that can handle multimedia events using IoT based technologies. In order to process the Internet of Multimedia Things based generated data as native event type, we must need to implement the multimedia aware event based system which can serve as a middleware to support adaptive multimedia services between heterogeneous environment and its applications within smart cities.

C. PROBLEM DESCRIPTION

1) REQUIREMENTS

The following requirements are tackled in this paper to process multimedia events with high performance, within an event based paradigm:

- Support of unstructured multimedia events which may contain information in the form of images or video.

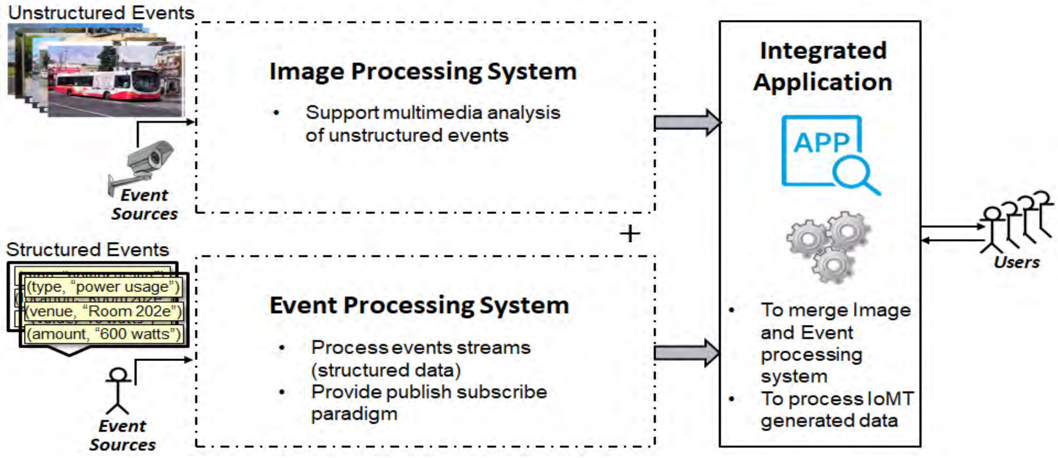


FIGURE 2. Current Approaches.

- High accuracy of feature extraction method used with matcher.
- Low execution time of feature extraction for the processing of multimedia events.

2) RESEARCH QUESTIONS

The above requirements can be further elaborated into the following research questions:

1) How can we perform feature extraction for IoMT events with high performance?
2) How can we optimize feature extraction based on subscription constraints?

3) PROBLEM STATEMENT

Based on the above requirements and research questions, this paper is focused on the designing of a generalized event processing system that can consume Internet of Multimedia Things (IoMT) generated data as a native event type, support multimedia analytics, and react to users situations of interest with high performance.

IV. RELATED WORK

A. EVENT-BASED SYSTEMS

Traditionally, event based systems follow this model for the processing of structured event streams such as SIENA [27]. This approach realizes an event notification service using matching on events with high scalability and user expressiveness that supports structured (scalar) events. Similarly, a semantic based publish/subscribe system [28] targets the heterogeneity of structured events at the middleware level to operate over existing publish/subscribe paradigm on the basis of ontologies. Approximate semantic matching [29] is one of the recent methods designed for handling heterogeneous event in the field of event based systems. Existing event-based systems are only focused on structured events for the processing of subscriptions of a user and with no provision of handling the feature extraction requirements of multimedia events.

B. IMAGE-BASED SYSTEMS

1) REAL-TIME BASED IMAGE PROCESSING

Image processing is a method to convert an image into digital form and perform operations on it, in order to extract useful information. The goal of manipulation of images at different levels of complexity, can be broadly classified into three categories: preprocessing, analyzing, and understanding of an image. Higher-level processing of images involves "making sense" which may include representation, description and recognition of objects. There has been substantial research in the area of image processing, where systems are focused on handling real-time data and are highly efficient in recognizing objects in image events. In the paper [7], an efficient real time traffic sign recognition scheme is proposed using intelligent vehicles. A traffic recognition system is used to automatically inform the driver about the traffic sign information by a head up display (HUD) monitor or audio device. In terms of requirements, unstructured events are covered by the traffic recognition system but only for traffic images and especially for detecting circular and triangular shapes with the requirement of specific mathematical equations for particular shapes. However, the accuracy of the system is quite high with acceptable execution time, given that all events are related to traffic signs and cannot be generalizable. A software architecture for Video based Traffic Congestion Prediction is presented in [8]. The paper includes two contributions; firstly it presents an embedded computer vision system which collects traffic data. Secondly it reports an innovative method for predicting traffic congestions. From the perspective of requirements, the presented system achieves the real-time performance in terms of analyzing traffic in advance while struggling for achieving high accuracy. The data collected for the enhancement of performance is only specific for the two-lane motorway, thus it is far for its commercialism and cannot be generalized for handling all types of unstructured events. Similarly most of the existing image based systems [26], [30] are domain specific, and they cannot be applied in multiple situations. Moreover, because of a limited user interface, expressiveness for the user is low in these types of systems.

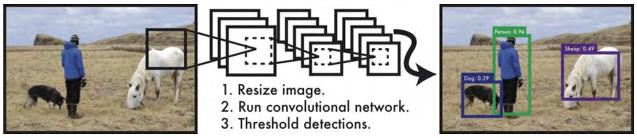


FIGURE 3. You only look once (YOLO) [31].

2) DEEP NEURAL NETWORK BASED IMAGE PROCESSING

In the domain of image recognition, deep convolutional neural networks are capable of achieving record-breaking results using purely supervised learning [4], [5]. You only look once (YOLO) is a machine learning based approach for object detection [31], [35] (shown in Fig. 3). It predicts both bounding boxes and class probabilities in an image to detect multiple objects of multiple classes. It is pre-trained on the Image Net [36], Pascal VOC [37], and Microsoft COCO [38] datasets. Training is based on a neural network based model having 20 convolutional layers including average pooling layer and fully connected layer. YOLO can detect more than 200 classes; it predicts detections for more than 9000 different object categories while running in real-time. In recent trends, it is fast and accurate which meet the expected requirements of execution methods for processing unstructured events. Since YOLO is designed only for the purpose of object detection based applications, it is only applicable for the domain of image recognition while processing unstructured events within smart cities.

In order to satisfy the requirements of execution method and processing of unstructured events on a broader perspective, it is essential to examine optimization techniques related to training configuration of DNN based methods. Several researchers investigate the impact of class distribution over classifier performance and concluded with general guidelines, that for a particular dataset a different class distribution may lead to substantial improvements in classifier's performance [39]–[41].

C. MULTIMEDIA-QUERY LANGUAGES

There are few languages like CVQL, SVQL, SPARQL-MM which can process multimedia data [42]. Content-based video query language (CVQL) is an extended version of existing structured query languages, that allows querying in video databases [32]. It is based on the spatial and temporal relationships of the content objects. Videos are divided into different classes, and then into their respective hierarchical categories. For example videos can be divided into *sports*, *politics*, *economics*, etc., and *sports* can be further classified into *basketball* and *tennis*. CVQL requires the knowledge of classes of fixed domain, thus cannot fulfill the requirement of processing any type of unstructured events. Moreover, performance requirements are not addressed in this work. SVQL is based on the structure of sequential query language SQL with videos [33] by modifying the "where" clause to process them. It is more expressive as it includes a variable declaration, structure specification, feature specification and spatial-temporal specification in addition to the existing conditional expressions. With respect to requirements, the SVQL expressions are generalizable for handling unstructured events, however, execution methods have not been evaluated for efficiency and accuracy. Similarly, SPARQL-MM [42] is also an extended version of SPARQL, by introducing spatio-temporal filter and aggregation functions for multimedia data. Although processing unstructured data is the primary goal of SPARQL-MM, it is not focused on its performance requirements. The goal of the MPEG Query Format (MPQF) is to facilitate and unify access to distributed multimedia repositories. MPQF can be used as a standard interface for multimedia retrieval engines [34]. It has the ability to access distributed multimedia repositories and uses an interactive feedback approach to improve efficiency which is not a recommended solution for real-time applications.

As a result of the analysis of related work, Table 2 shows a comparison of existing approaches with mapping of requirements (suggested in Section–III-C.1) for the proposed system.

It is clear from the research reviewed that there is no event processing system that can process unstructured events, in addition to structured events. Also, no image based systems are able to support the event-based paradigm with high user expressibility. Deep neural network based methods open new paths in the direction of image recognition techniques that can be beneficial in the design of multimedia event processing based systems. Moreover, the potential improvement in performance of the system with changes in configuration of classifiers can be useful for the further optimization of existing neural network based systems.

V. PROPOSED APPROACH

In this paper we propose a generalized approach for handling IoMT-based events which can be broadly classified into two categories according to their abstraction levels: (i) providing a generalized multimedia aware middleware; and (ii) facilitating multimedia events services using deep convolutional neural networks.

A. GENERALIZED IOMT-BASED MIDDLEWARE

In order to process IoMT based data, a multimedia stream-processing engine has been proposed within event-based services which can serve as middleware between multimedia heterogeneous sensor networks and their application portal within smart cities. Fig. 4 demonstrates the interaction of the proposed system with wireless sensor networks for different domain-specific applications using middleware. The incorporated IoMT-aware middleware provides event based services by consuming events (scalar/multimedia) generated from sensors (like camera, RFID tags, temperature sensors etc.), process them using multimedia event processing engine

TABLE 2. Comparison of related-work with requirements dimensions.

Category	Approach	Requirements		
		High Accuracy	Low Execution Time	Events Type (Unstructured)
Event-Based Systems	Carzaniga et al. (SIENA) [27]	N.A	N.A	Structured events only
	Zeng & Lei (Semantic Publish Subscribe) [28]	N.A	N.A	Structured events only
	Hasan et al. (Approx. Semantic Matching) [29]	N.A	N.A	Structured events only
Image-Based Systems	Glas et al. [8]	Limited Accuracy	Efficient	Image/Video events
	Shu et al. [30]	Accurate	N.A	Image/Video events
	Lai & Yu [7]	Average Accuracy	Efficient	Image/Video events
	Redmon et al. (YOLO) [31]	Accurate	Fast	Image/Video events
Multimedia Query Languages	Kuo & Chen (CVQL) [32]	N.A	N.A	Video events
	Lu et al. (SVQL) [33]	N.A	N.A	Video events
	Doller et al. (MPQF) [34]	N.A	Efficient	Multimedia events

and react to users accessing multiple types of applications. The enhanced event processing engine for IoMT events at the middleware, enables the analysis of multimedia (unstructured) data along with scalar (structured) data. Moreover, it also permits easy migrations from one domain to another within multimedia processing, using an adaptable feature extraction model that may encompasses multiple types of operators. Examples of events to support the proposed generalized approach can be analyzed along two dimensions:

1) Nature of Events:
   - Scalar
   - Multimedia
2) Application of Events:
   - Transportation
   - Entertainment
   - Security
   - Energy Consumption
   - Temperature readings etc.

The proposed framework is generalizable to support different kinds of operations on multiple application events, using existing event processing query languages and proposed multimedia feature extraction model, irrespective of their domain and nature of events.

B. DNN-BASED MULTIMEDIA EVENT PROCESSING

The proposed Multimedia Stream-Processing Engine (MSPE) is based on event processing, multimedia analysis, and deep convolutional neural networks, to meet the requirements of IoMT based systems in real-time. The incorporation of event-based systems with multimedia analysis provides the support for processing of multimedia events streams within the publish-subscribe paradigm. Deep convolutional networks based techniques are included with multimedia analysis to facilitate the processing of IoMT generated data with high performance. Deep Neural Network (DNN) based feature extraction is proposed in the form of operators, which extends the event query language to support multimedia analysis within event-based systems. A new "detect" operator has been developed to provide the requisite feature extraction, for the purpose of detection of objects inside image events.

In order to accomplish the goal of processing IoMT based events, the proposed approach can be summarized as the contribution of three systems shown in Fig. 5 with their respective endeavor. The system must design and implement an event processing engine with multimedia analysis using deep convolutional neural network based techniques to process multimedia event streams in publish-subscribe paradigm with high performance.

With respect to the research questions in Section–III-C, we propose an IoMT based model using "DNN based Feature Extraction Operator" and its optimization with the use of a subscription based "Classifier Division and Selection" approach.

C. SYSTEM ARCHITECTURE

On the basis of an architecture of existing Information Flow Processing (IFP) systems [1], a functional model (shown in Fig. 6) has been designed for a Multimedia Stream Processing Engine (MSPE) that consists of a matcher, Multimedia Event Processing Language (MEPL) statement, Subscription Covering based Optimization, feature extraction and a collection of classifiers.

1) RECEIVER

Event Sources ("sensors" in the present case) create events, which gets received by the *Receiver*, that sends events to MSPE for the processing. Receiver implements the transport protocol to communicate information over the network [1]. It is also responsible for receiving information in the form of events from multiple sources and act as an intermediator to send them one by one to the information flow processing (IFP) system which is multimedia stream processing engine in the proposed method.

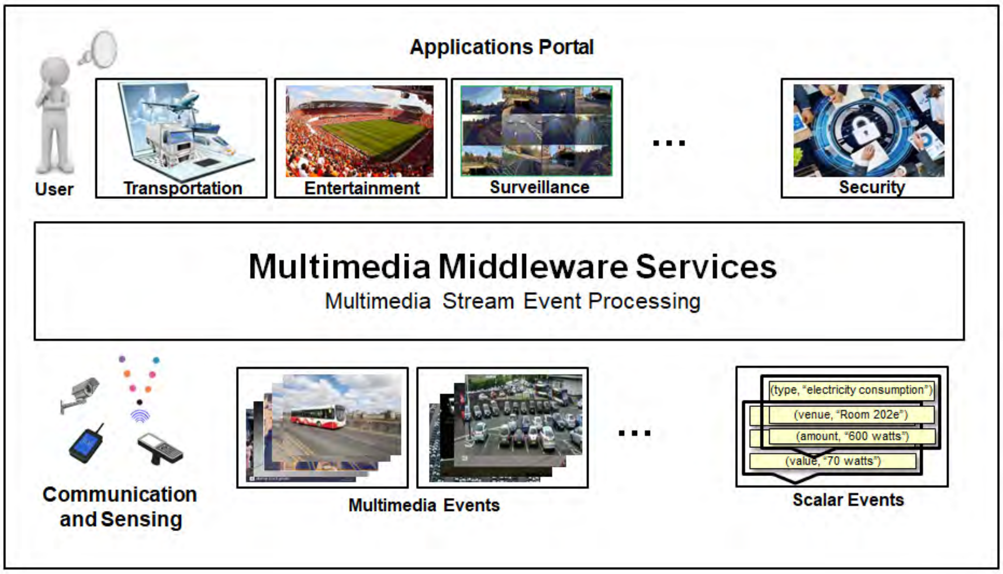


FIGURE 4. IoMT Aware Middleware.

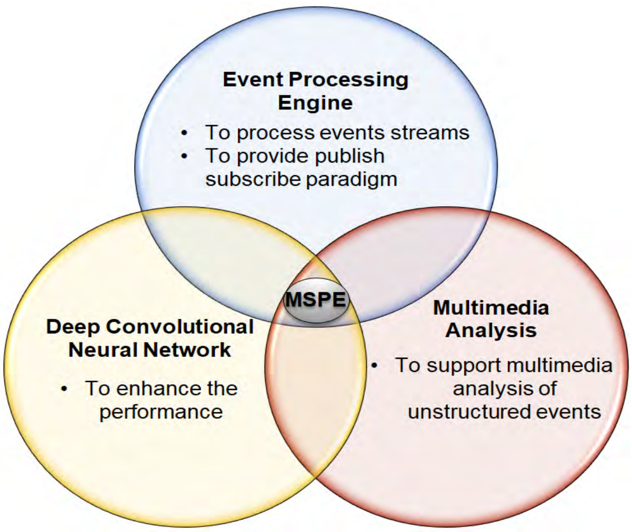


FIGURE 5. Systems required for IoMT.

2) MULTIMEDIA STREAM PROCESSING ENGINE (MSPE)

a: MATCHER

The *matcher* is responsible for the detection of conditions which hold in image events according to the user query (which has been evaluated using MEPL statements) and the propagation of notifications to the forwarder according to the condition detected in multimedia events.

b: MULTIMEDIA EVENT PROCESSING LANGUAGE (MEPL)

Subscriptions are received by *MEPL Statement*, with "Image Event" from *matcher* which analyzes the structure of query and instantiates a feature extraction model while using *Subscription Covering based Optimization* for filtering commonalities. MEPL will be responsible for resolving the signatures of operators associated with multimedia event based query languages such as the "Detect" operator in the present scenario described in Section–V-E.

c: SUBSCRIPTION COVERING BASED OPTIMIZATION

Subscription Covering based Optimization receives subscribed keywords with identifiers of subscribers from *MEPL*. It removes common keywords to consider them only once for further processing of multimedia events and sends the aggregated subscriptions to the *Feature Extraction* model. For instance, if multiple subscribers are looking for the same object (say "person"), then the keyword "person" should be analyzed once associated with multiple subscribers.

d: FEATURE EXTRACTION

The feature extraction model performs operations on image events according to the subscriptions using image processing operators ("detect" operator in the present case) and a collection of *classifiers*. The DNN based feature extraction model is presently using "You Only Look Once" [31], [35] for the purpose of object detection. Object detection is used for extraction of image features as it is the most common problem in the context of smart cities. Moreover, subscriptions from subscribers in the form of "keywords" will also direct the feature extraction model to use suitable classifiers, for the enhancement of performance.

The feature extraction model also facilitates the proposed system to include multiple types of operators for processing different features of the multimedia events which also makes it easily transportable to multiple domains and hence generalizable.

e: CLASSIFIERS

DNN based feature extraction model interacts with classifiers using keywords which is a key requirement of the proposed optimization methodology. Classifiers are trained on classes belonging to real-world objects to perform detection.

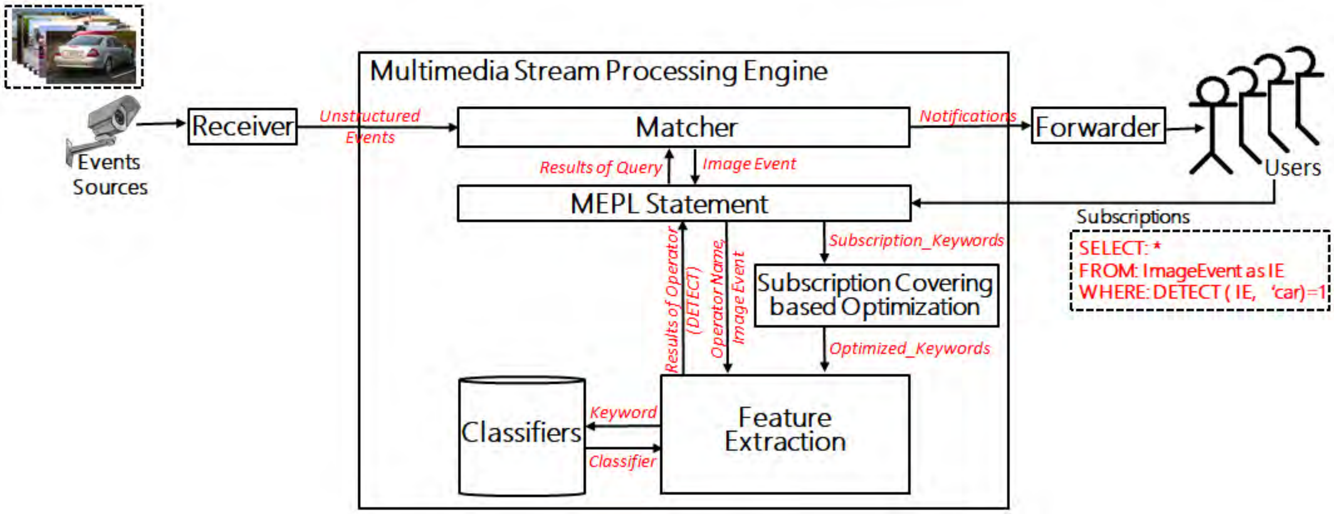


FIGURE 6. IoMT based model.

The number of classes per classifier configuration may vary with a change of domains and will be responsible for the robustness of the system. The input "keywords" will direct a feature extraction model to choose the suitable classifiers for the processing of image events.

3) FORWARDER

Finally events propagate to the forwarder which will notify users according to their registered subscriptions. The "forwarder" is also necessary for the implementation of transport protocols for the purpose of communication.

D. SYSTEM OPTIMIZATION

In order to achieve the goal of performance requirements, a subscription-based optimization technique has been proposed along with MSPE. The feature extraction model within MSPE uses a DNN based technique to identify objects efficiently. Irrespective of providing high-performance capabilities in image recognition, DNN based techniques are also dependent on the trained classifiers for the purpose of object detection. Conventionally, these classifiers are trained on general-purpose datasets consisting of a large number of classes, which may reduce the performance. The division of classifiers on the basis of domain and selection of classifiers on the basis of subscriptions could be a possible solution resulting in improvements in classifier performance. Thus, the proposed optimization is based on the inclusion of "classifier division and selection approach" which enables the proposed feature extraction model to choose a suitable classifier on the basis of subscription constraints. The approach is presented with implications of using only n-class classifiers belonging to a particular domain related to subscriptions with optimal values of "n" meanwhile neglecting classes of irrelevant domains.

Fig. 7 demonstrates the flow of subscriptions for the optimization of MSPE. Suppose registered subscriptions of users are $Sub_1 : \{detect(IE, person) \vee detect(IE, tennis\_racket) \vee detect(IE, dog)\}$ and $Sub_2 : \{detect(IE, car) \vee detect(IE, person) \vee detect(IE, dog)\} \vee detect(IE, motorbike)\}$. Subscriptions will be analyzed using MEPL, disintegrated into keywords and communicated to the "feature extraction" model via "Subscription Covering based Optimization". The objective of prior subscription covering based optimization is to remove common *keywords* resulting in the detection of only *person*, *tennis_racket*, *motorbike*, *car* and *dog* as keywords for the feature extraction in the example. The feature extraction model will extract objects using an object detection model (presently "YOLO" [31]) using only specific classifiers related to the prescribed attributes (keywords). For instance "car" classifier (single class classifier) will be selected by "YOLO" model for the detection of a *car*. The basic idea is to use only the available classifier which is closely related to the attribute, which can vary from single to n-class classifiers having n ranges 1 to ∞.

Practically, it is not possible to design a classifier consisting of all probable classes (∞), and also it is very unlikely that such a classifier would perform better than domain-specific classifiers. The utilization of n-class classifiers related to the particular domain may lead to improvements in performance as compared to the use of a general-purpose classifiers having a large number of classes. For further illustration, we are also considering the traffic, sports, animal and home subscriptions related to classifiers having 8, 9, 10 and 17 classes respectively derived from Pascal VOC [37] and Microsoft COCO [38] datasets. Since the constructed n-class classifiers belong to different domains, they also comply with the goal of generalizability. Lastly, MSPE classifiers also include two general-purpose classifiers with 20 and 80 classes constructed using Pascal VOC and Microsoft COCO for the proof of optimization described in Section–VI-C. The inclusion of classifiers with other larger values of "n" is not viable in the present case due to the limitations of YOLO model

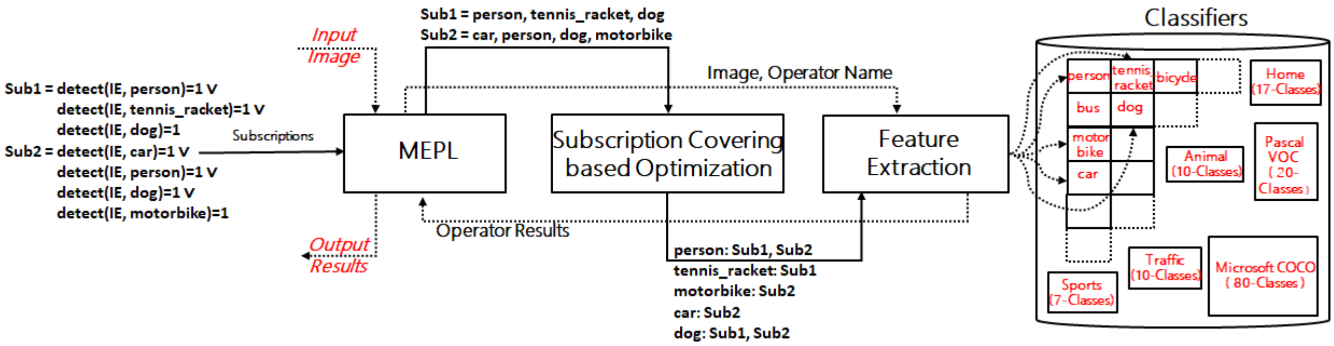


FIGURE 7. System Optimization using Subscriptions.

performing worse on ImageNet dataset [35] having more than 1000 classes and outperforms other object detection models [35] for real-time event detection.

E. DETECT OPERATOR

The *detect* operator is a general purpose operator that has been designed for detection of objects in image events. Detect is a binary operator, which consists of two inputs: *image event* and *keyword*. Events contain the details of image event, while keyword denotes the name of object that user intends to detect in an image. The return type of the operator is Boolean, either true or false depending on the detection of object in an image. The *detect* operator is:

boolean DETECT (Image_Event, Keyword)

Consider the situation in which a subscriber wants to know that a particular object (like car, bus etc.) is present or absent in the current image event. Examples of such kind of queries by using the proposed *detect* operator, are shown below:

Exmple 1: Query statement "*Is a Bus on the Bus Stand?*" for public transport management can be expressed as:
SELECT $*$
FROM Image_Event AS IE
WHERE $DETECT(IE, \text{'}bus\text{'})$

Exmple 2: Query statement "*Is Parking spot empty?*" related to car parking can be expressed as:
SELECT $*$
FROM $Image\_Event.win : time(1\ min)$ *AS IE*
WHERE $DETECT(IE, \text{'}car\text{'})$

The "DETECT" operator is implemented using "You Only Look Once" [31], [35], for the purpose of classifier based object detection. The performance of the *detect* operator on multiple types of events for handling multiple domain related subscriptions is shown in Fig. 8.

F. MULTIMEDIA STREAM PROCESSING ALGORITHMS

The utilization of DNN based systems for the purpose of high performance requires a classifier selection technique for the analysis of multimedia events. Thus, we present a method that can analyze multimedia events only once to process multiple subscriptions. In addition, an optimization technique "subscription-based classifier selection" for finding classifiers is proposed in this section. The implementation procedures for optimized IoMT based model and handling of commonalities among subscriptions are shown in Algorithm 1 and 2, with description of symbols in Table 3.

TABLE 3. Description of symbols.

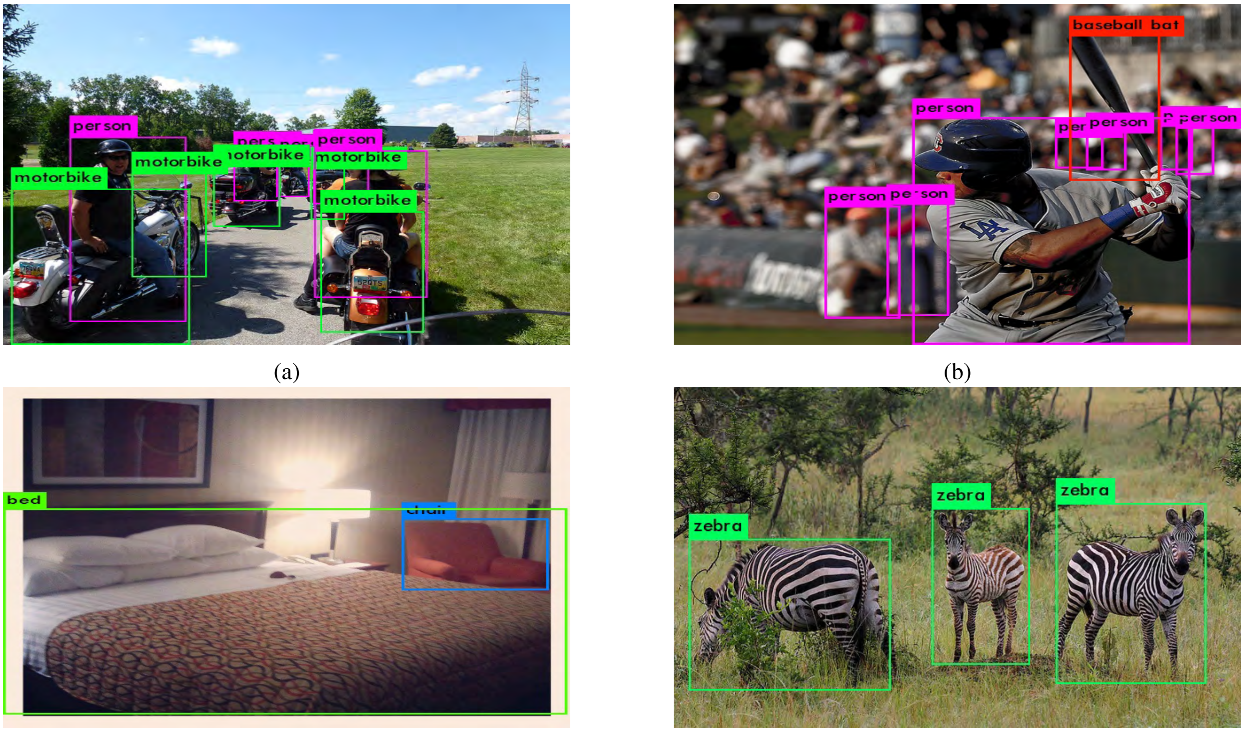
Symbol	Description
$List(L)$	List (L) of sets of Subscriptions
$T(K, S)$	Two-dimensional matrix for indexing Keywords and Subscribers
K	Keywords
S	Subscribers
S_{IE}	Stream of Image Events

1) MULTIMEDIA STREAM PROCESSING ENGINE

The proposed multimedia stream processing algorithm uses the following four steps and is based on keyword based optimization for handling commonalities:

1) Finding commonalities among subscriptions for subscription covering based optimization using keywords.
2) Identification of classifier according to the subscribed keyword.
3) Application of object detection for the processing of image events.
4) Notification of user on the matching of image event with subscription.

Let *List(L)* consist of a sets of subscriptions (Sub_i) with each set consisting of the identity of subscribers (S_i) with a set of keywords ranging from 0 to K_{j_i}. Consider a stream of image events as S_{IE}. The algorithm starts with distribution of subscriptions into a two-dimensional adjacency matrix $T(K, S)$ with K as keywords and S as indexes of subscribers corresponding to keywords, using the procedure of handling commonalities. Initialize m and n with the total number of keywords and subscribers respectively. The algorithm keeps on receiving the image events as IE and process them for each keyword k_i belonging to $T(K, S)$ with i ranging from 1 to m. Each iteration begins with the identification of *class name* related to keyword k_i and recognition of specific *classifier* using class name. Thereafter, objects are extracted from



(a) (b) (c) (d)

FIGURE 8. **Detect Operator on Traffic, Sports, Home and Animal related events. (a) Traffic Event. (b) Sports Event. (c) Home Event. (d) Animal Event.**

image events using the *matcher* and users notified on matching of class name with objects detected.

2) SUBSCRIPTION COVERING BASED OPTIMIZATION

Algorithm 2 shows the steps of keywords based optimization used for handling commonalities among subscriptions. Keywords based optimization is the procedure of identification of common keywords among all subscriptions and considering them only once for the processing of image events. The procedure of handling commonalities consist of following major steps:

1) Instantiate keyword-subscriber two dimensional matrix using input subscriptions.
2) Identification of keywords for each subscriber.
3) Association of subscriber identities with index of requisite keywords.

Let the *List(L)* be the input and $T(K, S)$ be the output two-dimensional adjacency matrix, consisting of keywords and subscribers identifiers. Next, initializes the total number of subscribers as i, counting the total number of keywords as j and assign $T(K_{j_i}, S_i)$ as 1 with each keyword K_{j_i} belonging to subscriber (S_i) using subscription set (Sub_i).

VI. EVALUATION

On the basis of traditional evaluation metrics, we are using throughput and accuracy to evaluate the efficiency and effectiveness of the multimedia event based system. In addition to these performance metrics, we are also comparing the

---

Algorithm 1 Multimedia Stream Processing Engine

Input: List (L) of sets of Subscriptions:
$(Sub_1 : \{S_1, K_{1_1}, K_{2_1}, ..., K_{j_1}\},$
$Sub_2 : \{S_2, K_{1_2}, K_{2_2}, ..., K_{j_2}\}, ...$
$Sub_i : \{S_i, K_{1_i}, K_{2_i}, ..., K_{j_i}\})$ and
S_{IE}: Stream of Image Events.
Output: Notifications

T(K,S)← $SCBO(L)$
{// - - - Subscription_Covering_based_Optimization}
$m \leftarrow count\_keywords(T(K, S))$
$n \leftarrow count\_subscribers(T(K, S))$
while true **do**
  $IE \leftarrow Image\_Event(S_{IE})$
  for $i = 1$ to m **do**
    $classname \leftarrow \{k_i | k_i \epsilon T(K, S)\}$
    $classfier \leftarrow find\_classifier(classname)$
    $objects \leftarrow matcher(IE, classfier)$
    if $(classname \ \epsilon \ objects)$ **then**
      for $j = 1$ to n **do**
        if $(T(k_i, s_j) = 1)$ **then**
          $notify(s_j)$
        end if
      end for
    end if
  end for
end while

TABLE 4. Performance of proposed system on different classifiers in terms of accuracy and throughput (FPS).

Exp. No.	Events	Example of Subscriptions	Accuracy				Throughput (fps)			
			Proposed Approach			Overall Change	Proposed Approach			Overall Change
			80-Class Classifiers	1-Class Classifier	N-Class Classifier		80-Class Classifiers	1-Class Classifier	N-Class Classifier	
1.	Traffic Events ≈ 25818	person, bicycle, car, motorbike, bus, truck, traffic light, stop sign, etc.	16.34%	94.58%	16.35% (N = 8)	+ve (0.06%)	108.08	114.67	111.18 (N = 8)	+ve (2.87%)
2.	Home Events ≈ 17218	chair, sofa, potted plant, bed, dining table, toilet, tv, refrigerator, etc.	83.15%	74.93%	83.19% (N = 17)	+ve (0.05%)	107.83	114.04	110.20 (N = 17)	+ve (2.20%)
3.	Sports Events ≈ 7776	frisbee, snowboard, sports ball, baseball, bat, skateboard, tennis racket, etc.	90.26%	68.04%	90.26% (N = 9)	No Change (0.00%)	107.72	114.94	110.65 (N = 9)	+ve (2.72%)
4.	Animal Events ≈ 8265	bird, cat, dog, horse, sheep, cow, elephant, bear, zebra, giraffe, etc.	86.51%	68.60%	86.45% (N = 10)	-ve (0.07)%	108.01	114.56	110.81 (N = 10)	+ve (2.59%)
5.	Mixed Events ≈ 40137	person, aeroplane, boat, bottle, wine, glass, book, glove, surfboard, cat, etc.	49.45%	81.72%	55.45% (N = 44)	+ve (12.13%)	106.23	114.55	107.16 (N = 44)	+ve (0.88%)

Algorithm 2 Subscription Covering based Optimization

Input: List (L) of sets of Subscriptions S:
$(Sub_1 : \{S_1, K_{1_1}, K_{2_1}, ..., K_{j_1}\},$
$Sub_2 : \{S_2, K_{1_2}, K_{2_2}, ..., K_{j_2}\}, ...$
$Sub_i : \{S_i, K_{1_i}, K_{2_i}, ..., K_{j_i}\})$
Output: Decision Tree
$T(K, S) \leftarrow 0$
$i \leftarrow Count\_Subscribers(S)$
for 1 to i **do**
  $j \leftarrow Count\_Keywords(Sub_i)$
  for 1 to j **do**
    $T(K_{j_i}, S_i) \leftarrow 1$
  end for
end for
return $T(K, S)$

single-class classifiers with existing n-class classifiers using precision and recall for the proof of optimization. All experiments have been conducted on Ubuntu 16.04.3 LTS (GNU/Linux 4.13.0-26-generic x86_64), with NVIDIA TITAN Xp GPU at 33MHz.

A. EVALUATION METRICS

- Throughput: Number of events matched in a unit time, measured in terms of frames/sec (fps).
- Accuracy: Ratio of correctly predicted observation to the total observations.

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \quad (1)$$

- Precision: Ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = TP/(TP + FP) \quad (2)$$

- Recall: Ratio of correctly predicted positive observations to the all positive observations of actual class.

$$Recall = TP/(TP + FN) \quad (3)$$

B. EVALUATION OF FEATURE EXTRACTION

In order to answer "*How can we perform feature extraction for IoMT events with high performance?*", the proposed approach of using "DNN based feature extraction operator" has been evaluated on multiple domain image event streams for its performance and generalizability.

1) EXPERIMENTAL SETUP

- Event Sets
  Four event sets consisting of events related to traffic, sports, home and animals, have been prepared manually using classes of Microsoft COCO dataset [38].
- Subscription Sets
  Random subscriptions have been generated manually for each of the applications, by varying the number of users and number of subscriptions per user. Subscriptions consist of attributes having the name of objects for detection like $\{cat\}$, $\{dog\}$, $\{cat, dog, horse\}$, $\{car, motorbike\}$ etc.
- Matcher Constraints
  Given a set of n subscriptions $S = \{s_1, s_2, ..., s_n\}$ with collective attributes $A = \{a_1, a_2, ..., a_t\}$ and a set of classifiers $C = \{c_1, c_2, ..., c_m\}$, the event matcher has to match events with subscriptions S using classifiers C. The matcher constraint is a condition specified on set of attributes A that belongs to subscriptions S. There exists a subset C' of available classifiers C, which covers all attributes of A, *i.e.*, $\exists\, C'\ \forall (a_i \in A)(C' \subseteq C \wedge a_i \in C')$, to achieve high performance.

2) EXPERIMENTS AND RESULTS

Experiments have been conducted on randomly generated subscriptions by considering multiple users ($m \epsilon [1, 10]$) and a large number of subscriptions ($n \epsilon [1, 100]$) per user. Table-4 summarizes the average accuracy and throughput of the proposed system and also demonstrate the overall improvement in performance for analyzing traffic, sports, home, animal and mixed image event streams. Accuracy of the proposed system is measured with the help of correctly predicted observations (true positives and negatives) with respect to the total number of observation/image events, whereas throughput indicates the number of frames processed per second (fps). Performance has been evaluated for each type of event stream by using three types of classifiers: single class, N-class and 80-class classifiers, to demonstrate the impact of an increased number of classes per classifier on performance. Here the value of "N" depends on the application domain which is taken as 8, 9, 17 and 10 for traffic, sports, home and animal classifier respectively. The 80-class classifier serves the purpose of a general classifier having Microsoft COCO classes consisting of multiple domain categories [38], and it will remain the same throughout all experiments. Additionally, performance has been evaluated by replacing single 80-class classifier with category related single-class classifiers. Results shows that the throughput of the system is increased while having competitive accuracy. The only drawback of using single-class classifiers is the requirement of loading all classifiers at the same time. However, the decrement in throughput as well as accuracy on using 80-class classifier as compared to single class classifiers, shows that the performance of the system will decrease with an increase in the number of classes per classifier. Thus, it will be beneficial to choose the optimal value of "N" and consider only the related classes for the construction of a classifier for the purpose of optimization.

Consequently, we also evaluated our system on domain-specific (N-Class) classifiers that consistently outperform the other classifiers with an average throughput and accuracy of 110 fps and 66.34% respectively. The high average throughput of the proposed system consisting of N-Class classifiers, for different event streams while achieving high accuracy, signifies the generalizability, efficiency, and effectiveness of the proposed system for real-time applications.

C. PROOF OF OPTIMIZATION

The research question "*How can we optimize feature extraction based on subscription constraints?*", has been analyzed using proposed "Classifier Division and Selection" approach with the help of throughput and precision-recall for the proof of optimization.

1) EXPERIMENTAL DATASETS

- Training Datasets
  All 80 classifiers (classifier with 1 class, classifier with 2 classes, classifier with 3 classes and so on) are trained on Microsoft Common Objects in Context (MCOCO) training dataset [38], to demonstrate the dependence of throughput on the number of classes. Also two other specialized classifiers are trained on Pascal VOC [37] and Microsoft COCO [38] having 20 and 80 classes respectively, for the purpose of comparison of precision and recall.
- Testing Datasets
  N-Class classifiers are tested on testing datasets of Pascal VOC [37] and MCOCO [38].

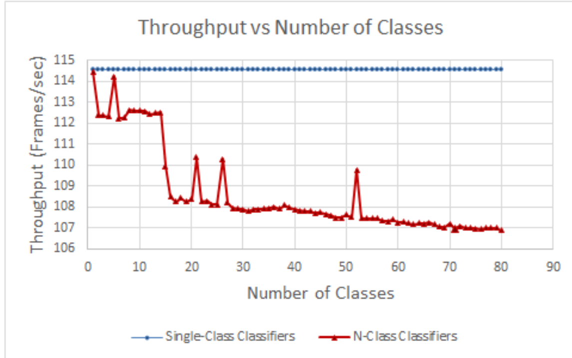


FIGURE 9. **Average Throughput of Classifiers.**

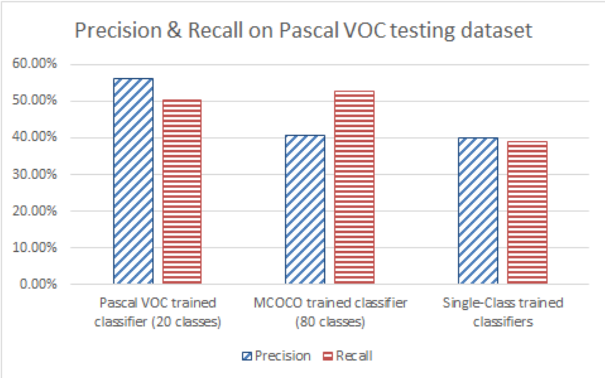
2) RESULTS & COMPARISON

a: THROUGHPUT vs NUMBER OF CLASSES

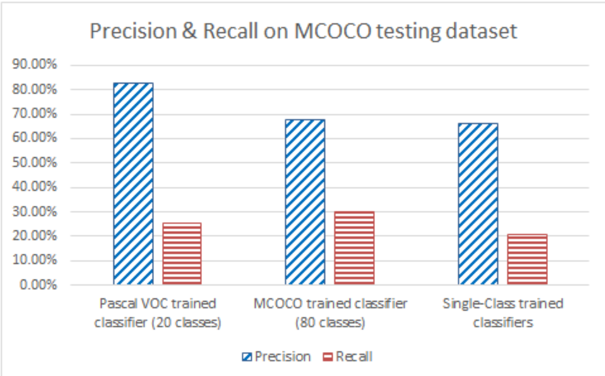
Fig. 9 illustrates the relationship of throughput with the number of classes, for n-class classifiers where n ranging from 1 to 80. Here, n-class classifiers refers to a classifier which is trained on n classes. The performance of the system on using single-class classifiers (classifier trained on only one class) will remain constant, with the condition of loading of all classifiers at the same time. Also it can be seen from the graph that the throughput of n-class classifiers is continuously decreasing with an increase in number of classes. This shows that the number of classes is among one of the configuration parameters of classifiers which affect their time complexity with loose upper bound $o(n)$. Thus choosing an optimal value of n according to the required application, is recommended for the optimized solution of constant time complexity $\mathcal{O}(1)$.

b: PRECISION-RECALL vs N-CLASS CLASSIFIERS

Fig. 10 represents the comparison of average precision and recall for three types of classifiers: Pascal VOC, Microsoft COCO, and single-class trained classifiers. Here a Microsoft COCO trained classifier with 80 classes serves the purpose of testing the best case analysis as compared to a single class classifier trained on 1 class for the worst case analysis. However, Pascal VOC with 20 classes is considered to analyze the performance on average cases. Fig. 10a compares the performance of all types of classifiers on Pascal VOC testing dataset. It shows that the Pascal VOC trained classifier is performing best on Pascal VOC testing dataset, while the precision-recall of Microsoft COCO with 80 classes is the almost equivalent to the performance of Pascal VOC having 20 classes. However, the values of precision and recall



(a) Average Precision and Recall on Pascal VOC testing dataset



(b) Average Precision and Recall on MCOCO testing dataset

FIGURE 10. Average Precision-Recall of Classifiers.

achieved by single-class trained classifiers, which could be the worst possible case of N-class trained classifiers, are also quite promising as compared to other classifiers. Similarly Fig. 10b shows the performance of all three types of classifiers on Microsoft COCO testing dataset. The competitive precision-recall of the approach having N-class classifiers with N=1 is practically indistinguishable as compared to N=20 and N=80 classes, which also indicates the suitability of keeping the value of N as minimal as possible.

It implies that choosing lower values of "N" related to application domain can improve the throughput of the system and also does not influence its accuracy. Hence the "classifier division and selection" approach has proven to be useful for the purpose of optimization of feature extraction.

VII. CONCLUSION

To process multimedia events within the event based paradigm a generalizable IoMT based system has been proposed in this paper. It is based on a deep neural network based feature extraction model expressed as an object detection operator. The proposed model has been optimized by using a classifier selection approach based on subscription constraints. Experiments show that the proposed system achieves an average throughput of 110 frames/sec with an approximate accuracy of 66.34% on real-world events in various applications of smart cities. We show the decrease in performance of classifiers with an increase in the number of classes per classifier, which indicates the effect of number of classes on the performance of proposed system by time complexity $o(n)$. Precision and recall have been evaluated to show the reasonable performance of n-class classifiers even in a worst case scenario. The reduction in throughput with increase in number of classes and promising precision and recall even on small values of n for n-class classifiers, supports choosing the optimal number of classes per classifier to achieve constant high performance. Thus, it is evident that the proposed approach is capable of providing the desired optimization on the basis of classifier configuration using subscription constraints.

In future, it can be useful to analyze the system with other image processing operators. The proposed model can be extended for online training with the introduction of auto/semi-auto generation of classifiers. Such a direction could lead to the inclusion of adaptability in the sense that the system can be generalized or specialized according to the practicality of the model in various scenarios. Optimization of subscriptions and event streams is also another area for future research, which requires the knowledge modeling of IoMT generated data. Lastly, only a single event matcher is developed in the proposed model, support for complex pattern matching is an essential area of interest, which may include detection of patterns in multimedia event processing.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

REFERENCES

[1] G. Cugola and A. Margara, "Processing flows of information: From data stream to complex event processing," *ACM Comput. Surveys*, vol. 44, no. 3, 2012, Art. no. 15.

[2] P. T. Eugster, P. A. Felber, R. Guerraoui, and A.-M. Kermarrec, "The many faces of publish/subscribe," *ACM Comput. Surveys*, vol. 35, no. 2, pp. 114–131, 2003.

[3] S. Alvi, B. Afzal, G. Shah, L. Atzori, and W. Mahmood, "Internet of multimedia things: Vision and challenges," *Ad Hoc Netw.*, vol. 33, pp. 87–111, Oct. 2015.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[6] S. Hasan and E. Curry, "Thematic event processing," in *Proc. 15th Int. Middleware Conf.*, 2014, pp. 109–120.

[7] C.-H. Lai and C.-C. Yu, "An efficient real-time traffic sign recognition system for intelligent vehicles with smart phones," in *Proc. Int. Conf. Technol. Appl. Artif. Intell. (TAAI)*, Nov. 2010, pp. 195–202.

[8] H. Glasl, D. Schreiber, N. Viertl, S. Veigl, and G. Fernandez, "Video based traffic congestion prediction on an embedded system," in *Proc. 11th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2008, pp. 950–955.

[9] A. Floris and L. Atzori, "Quality of experience in the multimedia Internet of Things: Definition and practical use-cases," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 1747–1752.

[10] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[11] F. Wortmann and K. Flüchter, "Internet of things," *Bus. Inf. Syst. Eng.*, vol. 57, no. 3, pp. 221–224, 2015.

[12] Q. Wang *et al.*, "Multimedia IoT systems and applications," in *Proc. Global Internet Things Summit (GIoTS)*, Jun. 2017, pp. 1–6.

[13] T. Balan, D. Robu, and F. Sandu, "Multihoming for mobile Internet of multimedia things," *Mobile Inf. Syst.*, vol. 26, pp. 1–16, Sep. 2017, doi: 10.1155/2017/6965028.

[14] J. Yang, J. Li, and S. Liu, "A new algorithm of stock data mining in Internet of multimedia things," *J. Supercomput.*, pp. 1–16, Nov. 2017.

[15] H. Noura, A. Chehab, L. Sleem, M. Noura, R. Couturier, and M. M. Mansour, "One round cipher algorithm for multimedia IoT devices," *Multimedia Tools Appl.*, pp. 1–31, 2018.

[16] M. Jridi, T. Chapel, V. Dorez, G. Le Bougeant, and A. Le Botlan, "SoC-based edge computing gateway in the context of the Internet of multimedia things: Experimental platform," *J. Low Power Electron. Appl.*, vol. 8, no. 1, p. 1, 2018.

[17] L. Zhou and H.-C. Chao, "Multimedia traffic security architecture for the Internet of Things," *IEEE Netw.*, vol. 25, no. 3, pp. 35–40, Jun. 2011.

[18] A. Floris and L. Atzori, "Managing the quality of experience in the multimedia Internet of Things: A layered-based approach," *Sensors*, vol. 16, no. 12, p. 2057, 2016.

[19] A. Aliyu *et al.*, "Towards video streaming in IoT environments: Vehicular communication perspective," *Comput. Commun.*, vol. 118, pp. 93–119, 2017.

[20] Y. Zhang *et al.*, "Low-cost and confidentiality-preserving data acquisition for Internet of multimedia things," *IEEE Internet Things J.*, 2017.

[21] J. Yang, S. He, Y. Lin, and Z. Lv, "Multimedia cloud transmission and storage system based on Internet of Things," *Multimedia Tools Appl.*, vol. 76, no. 17, pp. 17735–17750, 2017.

[22] J. Huang, Q. Duan, Y. Zhao, Z. Zheng, and W. Wang, "Multicast routing for multimedia communications in the Internet of Things," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 215–224, 2017.

[23] A. Elshafeey, N. S. A. Elkader, and M. Zorkany, "Compressed sensing video streaming for Internet of multimedia things," *Int. J. Cyber-Security Digit. Forensics*, vol. 6, no. 1, pp. 44–54, 2017.

[24] Y. Ke, R. Sukthankar, and M. Hebert, "Efficient visual event detection using volumetric features," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1. Oct. 2005, pp. 166–173.

[25] G. Medioni *et al.*, "Event detection and analysis from video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 873–889, Aug. 2001.

[26] A. Kanungo, A. Sharma, and C. Singla, "Smart traffic lights switching and traffic density calculation using video processing," in *Proc. Recent Adv. Eng. Comput. Sci. (RAECS)*, 2014, pp. 1–6.

[27] A. Carzaniga, D. S. Rosenblum, and A. L. Wolf, "Achieving scalability and expressiveness in an Internet-scale event notification service," in *Proc. 19th Annu. ACM Symp. Principles Distrib. Comput.*, 2000, pp. 219–227.

[28] L. Zeng and H. Lei, "A semantic publish/subscribe system," in *Proc. IEEE Int. Conf. E-Commerce Technol. Dyn. E-Business*, Jul. 2004, pp. 32–39.

[29] S. Hasan, S. O'Riain, and E. Curry, "Approximate semantic matching of heterogeneous events," in *Proc. 6th ACM Int. Conf. Distrib. Event-Based Syst.*, 2012, pp. 252–263.

[30] C.-F. Shu *et al.*, "IBM smart surveillance system (s3): A open and extensible framework for event based surveillance," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill. (AVSS)*, Sep. 2005, pp. 318–323.

[31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

[32] T. C. T. Kuo and A. L. P. Chen, "Content-based query processing for video databases," *IEEE Trans. Multimedia*, vol. 2, no. 1, pp. 1–13, Mar. 2000.

[33] C. Lu, M. Liu, and Z. Wu, "SVQL: A SQL extended query language for video databases," *Int. J. Database Theory Appl.*, vol. 8, no. 3, pp. 235–248, 2015.

[34] J. R. Smith *et al.*, "The MPEG query format: Unifying access to multimedia retrieval systems," *IEEE Multimedia*, vol. 15, no. 4, pp. 82–95, Dec. 2008.

[35] J. Redmon and A. Farhadi. (2016). "YOLO9000: Better, faster, stronger." [Online]. Available: https://arxiv.org/abs/1612.08242

[36] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[37] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

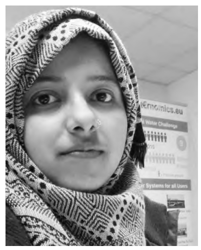
[38] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[39] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural Netw.*, vol. 21, nos. 2–3, pp. 427–436, 2008.

[40] T. Kavzoglu and I. Colkesen, "The effects of training set size for performance of support vector machines and decision trees," in *Proc. 10th Int. Symp. Spatial Accuracy Assessment Natural Resour. Environ. Sci.*, 2012, pp. 1–2.

[41] G. M. Weiss and F. Provost, "Learning when training data are costly: The effect of class distribution on tree induction," *J. Artif. Intell. Res.*, vol. 19, no. 1, pp. 315–354, Jul. 2003.

[42] C. Angsuchotmetee and R. Chbeir, "A survey on complex event definition languages in multimedia sensor networks," in *Proc. 8th Int. Conf. Manage. Digit. EcoSyst.*, 2016, pp. 99–108.



ASRA ASLAM received the B.Tech. and M.Tech. degrees in computer engineering from Aligarh Muslim University, India, in 2013, and 2015, respectively. She is currently pursuing the Ph.D. degree with the Insight Centre for Data Analytics, National University of Ireland Galway, Galway. She was a Guest Teacher with the Department of Computer Engineering, Aligarh Muslim University, from 2015 to 2016. Her research interests include smart cities, image processing, artificial intelligence, and machine learning.

EDWARD CURRY is currently a Research Leader with the Insight Centre for Data Analytics, and a funded Investigator at LERO The Irish Software Research Centre. He has involved extensively with industry and government advising on the adoption patterns, practicalities, and benefits of new technologies. He has authored or co-authored over 130 scientific articles in journals, books, and international conferences. He has presented at numerous events and has given invited talks at Berkeley, Stanford, and MIT. His research projects include studies of smart cities, energy intelligence, semantic information management, event-based systems, stream processing, and human-in-the-loop. He is a Vice President of the Big Data Value Association—a non-profit industry-led organization with the objective of increasing the competitiveness of European Companies with data-driven innovation. He is a member of the scientific leadership committee of Insight, and a Lecturer in informatics with the National University of Ireland Galway.