

# A Survey on Real Time Object Detection using Voice Activated Smart IoT

Tejas Jnanesh Ghalsasi

Graduate Student, Dept. of Computer Science, California State University, California, United States of America

\*\*\*

**Abstract** - One of the goals of Artificial intelligence (AI) is the realization of natural dialogue between humans and machines. In recent years, the dialogue systems, also known as interactive conversational systems are the fastest growing area in AI. Many companies have used the dialogue systems technology to establish various kinds of Virtual Personal Assistants (VPAs) based on their applications and areas, such as Microsoft’s Cortana, Apple’s Siri, Amazon Alexa, Google Assistant, and Facebook’s M. At the research in the Areas of Object Detection using Advanced neural Network Techniques such as Deep Learning and Reinforcement learning have paved path to the boom of Machine learning. This proposal is an idea put forward to combine the workings of both these areas and create a system which will effortlessly help Disabled and Normal humans in various fields like Home automation, networking, data monitoring,

frameworks resulting in a global competition. As we know competition enhances the overall value of the subject, the same happened with the Field of AI.

	Languages	Tutorials and training materials	CNN modeling capability	RNN modeling capability	Architecture: easy-to-use and modular front end	Speed	Multiple GPU support	Keras compatible
Theano	Python, C++	++	++	++	+	++	+	+
Tensor-Flow	Python	+++	+++	++	+++	++	++	+
Torch	Lua, Python (new)	+	+++	++	++	+++	++	
Caffe	C++	+	++		+	+	+	
MXNet	R, Python, Julia, Scala	++	++	+	++	++	+++	
Neon	Python	+	++	+	+	++	+	
CNTK	C++	+	+	+++	+	++	+	

**Keywords—** Virtual Personal Assistants; Internet-of-Things, Home Automation, Machine learning, Real Time Object Detection, Template Matching.

## Introduction

To design and implement natural and intuitive interaction modalities is a primary research field in the Human-Computer Interaction Domain. Systems that can interact with user in their natural language are being researched vigorously at present. [3] Voice enabled chatbots are becoming more and more popular with the advent of devices and technologies like Google Home, Amazon Echo, NLP, ML, AI etc. Chatbot is an artificial service that can start, continue and handle complex interactions with human partners in their natural language. Voice enabled chatbots, today, are considered as classical yet innovative interfaces for natural language interaction with machines. [1]

Here out of all the options mentioned I would like to highlight Amazon Alexa as I would be using it as a component in the project. That is because Alexa is backed by a strong developer community from amazon and has been the earliest player in this area of the market. Amazon Alexa devices launched in 2014 which gave them a strong lead over the market. [2]

Speaking of Machine learning, the growing boom in the field of AI is attributed to the various developments over the last 5 years resulting in competition between AI frameworks. Popular companies such as Airbnb, Uber, Facebook, Microsoft, Google open sourced their AI

Due to the wide availability of frameworks and repo packages on GitHub we have seen a community style rise in the development of AI techniques and applications. [4] This is also a result of a large-scale data boom and the availability of public data sets. Google, UC Irvine and many other academic and non-academic institutions have already open sourced their datasets which cover a wide variety of subjects. Government organizations such as public water systems, NASA, Police departments have also made their statistics public resulting in better efficiency of data.

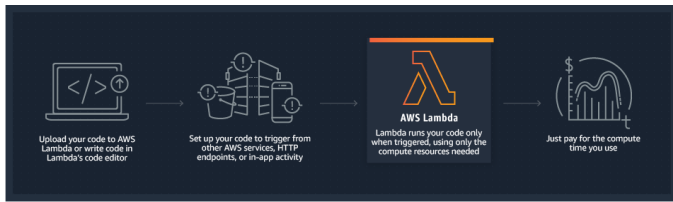
## Technological Background

To give a bit of a technological background we are going to go through the different Techniques of Machine learning and the working of AWS in 2 sections within this section

### A. AWS Architecture

AWS Lambda lets you run code without provisioning or managing servers. You pay only for the compute time you consume - there is no charge when your code is not running. [11]

With Lambda, you can run code for virtually any type of application or backend service - all with zero administration. Just upload your code and Lambda takes care of everything required to run and scale your code with high availability. You can set up your code to automatically trigger from other AWS services or call it directly from any web or mobile app.



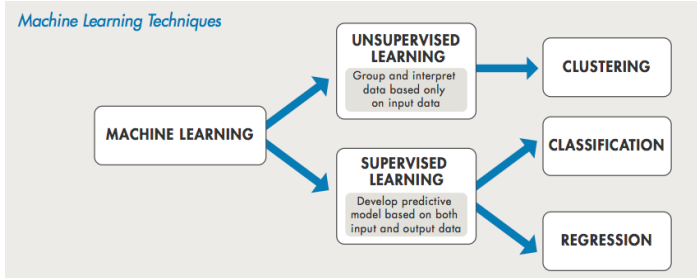
You can use AWS Lambda to execute code in response to triggers such as changes in data, shifts in system state, or actions by users. Lambda can be directly triggered by AWS services such as S3, DynamoDB, Kinesis, SNS, and CloudWatch, or it can be orchestrated into workflows by AWS Step Functions. This allows you to build a variety of real-time serverless data processing systems.

You can use AWS Lambda to perform data validation, filtering, sorting, or other transformations for every data change in a DynamoDB table and load the transformed data to another data store. [11]

In our case we are dealing with an IoT case and using AWS Lambda we can build serverless backends to handle web, mobile, Internet of Things (IoT), and 3rd party API requests.

These are very few of the applications in which AWS has proven its mastery which is why a lot of companies are using it in the industry today.

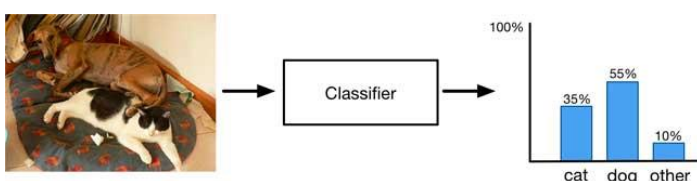
**B. Machine Learning in Object Detection**



Object detection is one of the classical problems in computer vision:

Recognize what the objects are inside a given image and also where they are in the image.

Detection is a more complex problem than classification, which can also recognize objects but doesn't tell you exactly where the object is located in the image — and it won't work for images that contain more than one object. [10]

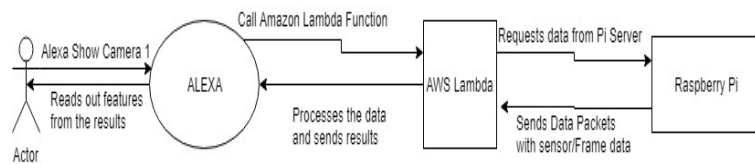


An object detection model predicts bounding boxes, one for each object it finds, as well as classification

probabilities for each object. It's common for object detection to predict too many bounding boxes. Each box also has a confidence score that says how likely the model thinks this box really contains an object. As a post-processing step we filter out the boxes whose score falls below a certain threshold (also called non-maximum suppression). [10]

**Discussion**

The proposed system aims to connect the power of Artificial intelligence to the ease of access of the smart appliances and Serverless AWS Architecture. The process flow of information would be displayed as follows:



Imagine a situation where a blind person wants to keep track of the activity going on in his front yard. The system consisting of a motherboard and a camera would sit outside connected to Wi-Fi. When the person would invoke Alexa to open the camera, the AWS Lambda function would be triggered.

The AWS Function would request for data packets from the Assembled system and would return results if any to the Alexa. Imagine the event of 3 men coming inside the court yard. The system would immediately detect and send processed results to Alexa which would then read out the message to the user.

This system can be used for plenty other business and industrial scenarios such as, patients waiting outside the clinic or some person carrying a harmful object such as a gun or knife, a baby is sleeping or awake, if the baby is inside the cradle or trying to jump outside. The future scope to this system is that we can enhance the camera used and also use other detection methods such as audio. In the case of audio recording we can use techniques like collaborative filtering to differentiate environmental high pitch voice and the voice made by the target i.e. baby, patient, etc.

The **SWOT** Analysis of the proposed system is as follows:

**Strengths:**

Innovative use of Smart devices along with Cloud computing and machine learning. Will help establish solutions to drive towards an automation driven economy Total Cost of the setup is around \$60-100. An efficient home security system at this cost is a strength.

**Weakness:**

In situations such as moisture, rain, snow or equipment malfunction, our system may malfunction or give false alarms.

### Opportunity:

Marketing and selling such a deal would bring us a good chance to make money. Such a system can be patented based on its unique approach.

### Threats:

The back bone of this system is the AWS cloud infrastructure. The rates are decided by AWS which is why any hike from the side of Amazon would result in surging of prices. Apart from that the data required to train the system need to be clean and coming from diverse sources. Mistakes caused due to machine learning due to bad ingestion data may result in false alarms

### Conclusion

As time moves on the drive for automation and mechanization is going to increase. We are seeing it in the form of chatbots and self-driving cars. In this proposal we went through different ideas about how to implement the proposed system which serves the purpose of smart devices interfacing with Raspberry Pi through Amazon AWS Cloud Infrastructure. We went through the SWOT Analysis and found out that this project needs to be implemented to study the impact of Automation and Smart Devices. Further the idea can be expanded into a number of other domains which is why I believe we need to implement a basic stage of the project.

### References

- [1] Reshmi, S., & Balakrishnan, K. (2016). Implementation of an inquisitive chatbot for database supported knowledge bases. *Sadhana*, 41(10), 11731178
- [2] S. Sikand, "Deep learning: Comparison of Available Frameworks – Zeg.ai – Medium," *Medium*, 17-Feb-2017. [Online]. Available: <https://medium.com/zeg-ai/deep-learning-comparison-of-available-frameworks-ab89610f8d2b>. [Accessed: 04-Oct-2018].
- [3] W. Wayt Gibbs, Build your own Amazon Echo-Turn a PI into a voice controlled gadget [Resources-Hands on]. *IEEE Spectrum*. 54(2017), pp. 20–21.
- [4] Q. Xu, F. Gao, K. Shi, G. Xu, Real-time rear of vehicle detection from a moving camera. *Control and Decision Conference ( ...)*, 4575–4578 (2014).
- [5] C. Z. Yue, S. Ping, in *Proceedings of 2017 2nd International Conference on Frontiers of Sensors Technologies, ICFST 2017*(Institute of Electrical and Electronics Engineers Inc., 2017), vol. 2017–January, pp. 489–492.
- [6] G. Bohouta, V. Z. Kępuska, "Real Time Speech Translation (Google TV and Chat) Live Speech Translation (Google TV) View project Comparing Speech Recognition Systems (Microsoft API, Google API And CMU Sphinx) using Java and C# language View project," (available) at <https://www.researchgate.net/publication/322686710>.
- [7] Y. Wang, J. F. Doherty, R. E. V. Dyck, Moving Object Tracking in Video. *Camera*.
- [8] V. Kepuska, G. Bohouta, in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018*(Institute of Electrical and Electronics Engineers Inc., 2018), vol. 2018–January, pp. 99–103.
- [10] M. Hollemans, "One-shot object detection," *Google's MobileNets on the iPhone*. [Online]. Available: <http://machinethink.net/blog/object-detection/>. [Accessed: 04-Oct-2018].
- [11] "AWS Lambda – Serverless Compute - Amazon Web Services," *Amazon*. [Online]. Available: <https://aws.amazon.com/lambda/>. [Accessed: 04-Oct-2018].

### BIOGRAPHY



Tejas Ghalsasi is pursuing Masters in Computer Science from California State University, Fullerton, California, United States of America. He completed his bachelors in 2015 from Fr. Conceicao Rodrigues College of Engineering which is affiliated with Mumbai University. Tejas has Software development work experience at Accenture which a Fortune500 company. Tejas will graduate from his Master's program in June 2019. You can find more about Tejas' projects on <https://www.linkedin.com/in/tejasghalsasi/>.