iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** XII **Month of publication:** December 2019

DOI: http://doi.org/10.22214/ijraset.2019.12050

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

A Survey on Virtual Personal Assistant

Sudarshan Modhave[1], Suyash Soniminde[2], Aadesh Mogarge[3], Mr. Kapil Tajane[4], Ajay Sah[5]
[1, 2, 3, 4, 5]*Department of computer Engineering, Pimpri Chinchwad college of Engineering, Pune, India*

Abstract: ***Ongoing advancements in smart home automation and smart assistants are recently drawing in the intrigue and interest of customers and scientists. Discourse empowered remote helpers offer a wide assortment of system arranged administrations. The majority of those gadgets intensely depend on cloud- based administrations, in this manner transmitting conceivably delicate information to remote servers. To annihilation such issues probably the most exceptional procedures in computer vision, deep learning, speech generation and recognition, and artificial intelligence, into a virtual assistant architecture for smart home automation systems are utilized. One of the objectives of Artificial intelligence is the acknowledgment of normal exchange among people and machines. as of late, the exchange frameworks, otherwise called intuitive conversational frameworks are the quickest developing region in AI. Numerous organizations have utilized the discourse frameworks innovation to build up different sorts of Virtual Personal Assistants dependent on their applications and territories. VPAs are utilized to build the association among people and the machines by utilizing various innovations, for example, motion acknowledgment, picture/video recognition, speech recognition, the huge exchange furthermore, conversational information base, and the general learning base.***

Keywords: Virtual Personal Assistant, Speech Recognition, Face Detection, Natural Language Processing.

I. INTRODUCTION

Computer systems are becoming perpetually intricate, both vertically and evenly, while the capacity and readiness of clients to endure multifaceted nature is relentlessly diminishing. This problem influences all items in all business sectors. A rich arrangement is to engage items with worked in counselors that address client issues at the interface: Virtual Personal Assistant. Nowadays, virtual personal assistants are winding up well known. They are being favored by individuals and for what reason would it be a good idea for them to not be? They can do almost every task that human assistant can do with less cost and more reliability. Unlike, humans they can serve you 24X7, 365 days in a year. By performing most of your tasks, they can contribute to raise your productivity. They provide large number of applications in various domains like business, education, healthcare, entertainment, etc. Also, intelligent personal assistants can be used in automobiles to enhance the user experience and ease of doing business of the customer. People prefer giving voice commands rather than typical typing or operating device manually because it is quicker and makes the task more interactive. Virtual personal assistant accepts multimodal inputs like voice, text, image, video and they process it to give better output. They have proven to be very useful in day to day life for everyone. You can do lots of things like booking flight tickets or movie tickets, ordering food, fixing appointment with doctor and many more things.

The feature that makes it more useful is voice commands. User can give voice commands even if he is busy in any other task, thus providing multitasking.

II. RELATED WORK

Two principle challenges in the territory of human- Computer interaction are: first, how to know the client, and second, how to help the client. The vast majority of the past work has added to the first test by improving the cooperation among clients and PA specialists, learning client's inclination and objectives, giving assistance at the ideal time, etc. Issues can be illuminated in a smart path by thinking, making inferences and learning.

Moreover, SARA is a large, complex system that could be hardly squeezed into a cheap, small-sized and resource-constrained hardware [7]. In the literature [8][9], several other architectures and implementations have been suggested. However, in our knowledge, so far, no embodied, vision and speech enabled virtual agents have been presented and released to the public, VPAs can acknowledge customers who discriminate against their faces and operate on low-cost and small-scale consumer appliances.

The absence of face recognizable proof capacities of most virtual specialist programming is most likely because of the deficiency, before, of appropriate lightweight, yet powerful and precise, face distinguishing proof approaches. By and large, until a couple of years back, face acknowledgment was off base or required amazing computational assets for the acknowledgment procedure or for the disconnected enlistment of the client pictures. The ongoing utilization of profound

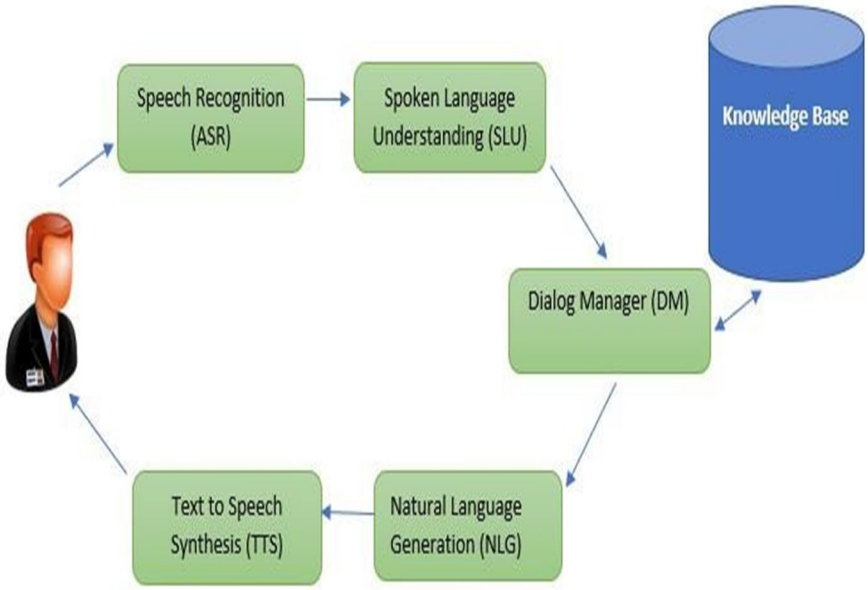
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

neural systems is delivering a problematic change in this pattern, allowing the improvement of powerful, exact and lightweight systems for face acknowledgment, that are right now moreover abused for client recognizable proof in cell phones. It is about time to coordinate those advancements into a virtual assistant.

III. SYSTEM ARCHITECTURE

The dialogue system is one of a functioning territory that numerous organizations use to structure and improve their new frameworks. There are various methodologies used to design the talk structures, in perspective on the application and its multifaceted nature. In view of method used to control trade, a talk structure can be assembled in three classes: Finite State (or graph) based systems, Frame based system and Agent based systems. The multi-modular dialogue frameworks process at least two joined client input modes, for example, discourse, picture, video, contact, manual gestures, look, and head and body development so as to plan the Up and coming Generation of VPAs model.

The proposed design is completely developed with modular approach. It is made of a lot of administrations, a graphical frontend and an organizer that influences on the administrations to offer to the client a multimodal and including collaboration with the associated home computerization and smart assistant systems [3][8].



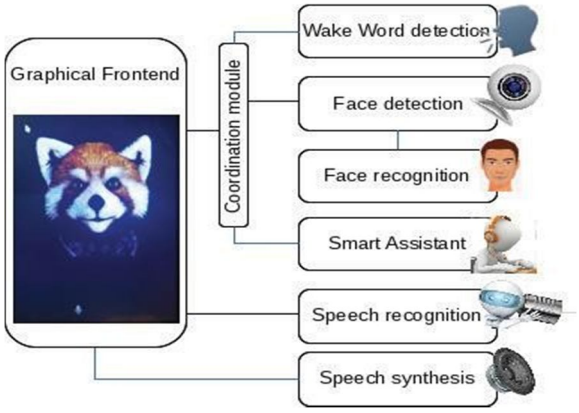
The modules pass on through connections and RESTFul relationship, thusly, if vital, different modules can be apportioned on different planning center points. The GoogleTTS module relies upon the notable Google Cloud Speech API. Despite relying upon cloud benefits and requiring a paid participation, Google Cloud Speech API offers the upsides of supporting various languages and including both male and female, amazingly common, voices.

Current virtual assistants can talk and tune in to their clients, yet, can't "see" them. In addition, by and large they don't highlight any sort of visual passionate input. They are visually impaired what's more, unremarkable to the client, consequently their communication is frequently hindered what's more, fragmented and, consequently, less viable and productive. To defeat the issue recorded over, this system gives a design for structure vision-empowered brilliant aides, given expressive and vivified graphical characters what's more, discourse acknowledgment and combination.

The absence of face ID capacities of most virtual assistant programming is likely because of the deficiency, previously, of reasonable lightweight, yet powerful and exact, face recognizable proof approaches. Much of the time, until a couple of years prior, face acknowledgment was wrong or required amazing computational assets for the acknowledgment procedure or for the disconnected enlistment of the client pictures. The ongoing use of profound neural systems is creating a troublesome change in this pattern, permitting the advancement of successful, exact and lightweight strategies for face acknowledgment, that are as of now moreover misused for client distinguishing proof in cell phones. It is about time to incorporate those innovations into a menial helper.

A. *Virtual Personal Assistant*

A specific virtual assistant is introduced, that exploits: Six service modules, a graphical front-end and a module for coordination.



B. *The Graphical Frontend and the Speech recognition module*

The Graphical Frontend relies upon a HTML5 report, containing the Javascript code expected to grant through a Websocket relationship with the Coordination module and through a RESTful API with the Speech association and the Talk affirmation modules. The character picked for the interface, was stimulated in order to make advancement hoc, expressive and noteworthy chronicles, that are joined and synchronized with the talk at run time. This results in a counting and reasonable collaboration with the user.

The HTML5 record is conveyed by a HTTPS server, privately introduced, to a Chromium internet browser, that was explicitly picked as it is open source and completely underpins the Google Speech to Text (GSTT) administration. The GSTT administration was chosen as it gives a successful and free answer for dependable what's more, multilingual speech recognition. As an outcome, the speech recognition module was acknowledged by methods for an outside administration. An elective module dependent on the Mozilla open source execution of DeepSpeech [9], that can run locally, is as of now being researched. The Speech acknowledgment module is typically inert, not recording nor detecting sound contribution, so as to save the client's security. It is unequivocally enacted by the Graphical Frontend in foreordained periods of the association (e.g., in the wake of representing a question to the client) and when the Coordinator module signals that the Wake Word has been distinguished by the Wake Word module.

C. *Speech synthesis module*

The Speech amalgamation administration gives the discourse capacity to the Graphical Frontend. Two Speech amalgamation modules were used: MaryTTS, in light of the MaryTTS programming, and the Flite2 module [4].

Because of the primary memory impediments of the embraced equipment, highlighting just 2GB of Ram, the lighter Flite2 module was picked. In spite of the fact that during the tests Red communicated uniquely in English language, an Italian voice created for a previous rendition of Flite is at present being ported to this module. An adaptation of the Flite2 module was additionally discharged as an open source node.js bundle.

D. *The Wake-up Word detection module*

The Speech combination administration gives the discourse capacity to the Graphical Frontend. Two Speech amalgamation modules were used: MaryTTS, in light of the MaryTTS programming, and the Flite2 module.

There are at present two modules in the Wake Word class, to be specific, the Snowboy module [10], coordinating the SnowBoy profound neural system-based programming, and the PocketSphinx module, in view of the homonymous open source programming created via Carnegie Mellon University [11]. The first module depends on an exclusive library containing the profound 1npm bundle named ianniTTS. neural system that is the establishment of the Snowboy programming, furthermore, accordingly isn't completely convenient. On the other hand, the subsequent module is open source and completely convenient, yet the methodology for encoding the wake word is unwieldy.

E. Face Detection module

A Face recognition module persistently filters the video contribution from an associated webcam and, at whatever point it distinguishes a human face, it alarms the Coordination module. At the point when a face is identified, assistant directs its concentration toward the client confronting the camera and welcomes her/him. In the event that the personality of the client is accessible (as the client has been perceived by the Face acknowledgment module), the client is called by name and each adjustment in the personality of the client in front of the camera is motioned by a comparing articulation.

The Face identification module picked for the assistant is the Haar module, in view of the quick and lightweight face identifier by Viola and Jones [12] and advanced in the latest rendition of the OpenCV library [13], however different modules were added to the Face Detection class to be tried with colleague.

F. Face Detection and Recognition module

The Face acknowledgment administration permits the remote helper to perceive the client before the camera by coordinating her face against a lot of photos that dwell in a nearby database. The quantity of photos isn't fixed and new ones can be included whenever, as the administration convention has a particular direction for adding another photo to the accumulation, together with the related character. Despite the fact that the administration isn't expected as a security instrument, it must be extremely exact so as to keep the important unwavering quality and the trust of the client. Also, a right distinguishing proof of the client is required for keeping up the cooperation setting (the bit of the past connection expected to effectively bargain with the present exchange). As an outcome, just the most precise and ongoing ways to deal with face acknowledgment were taken into record for the usage of the Face acknowledgment modules.

This method embraced for the module utilized depends on a picture metric acknowledged through a ResNet- 34 profound neural system [14].

G. Face Detection and Recognition module

A virtual assistant is a product specialist that can perform assignments or then again benefits for human clients. Current brilliant assistants depend on regular language handling (NLP) for parsing the contribution from the client and showing directions and demands, and, in a few cases, misuse man-made consciousness to expound explained and meaningful answers when required. The client information is generally printed (for example chatbots), vocal (as on account of present-day shrewd speakers). At times, it can likewise utilize pictures, that can be submitted to the specialist by taking pictures with a cell phone. The Smart assistant module in Red goes about as an interface to a smart associate stage, accordingly coordinating its administrations into the Virtual Assistant. As of now the class just contains the Mycroft module, that incorporates the Mycroft open source smart assistant platform [15] into VPA. Mycroft gives essential administrations and a negligible conversational capacity however it is measured and new administrations can be included by effectively putting in new "aptitudes". On account of the Mycroft module, Assistant can address to inquiries concerning the climate or the time around the world, recover data from a web search tool or from Wikipedia, straightforwardly turn on a light or interface to a home mechanization stage, leave a message for another client, set an alert, and then some.

IV. CONCLUSION

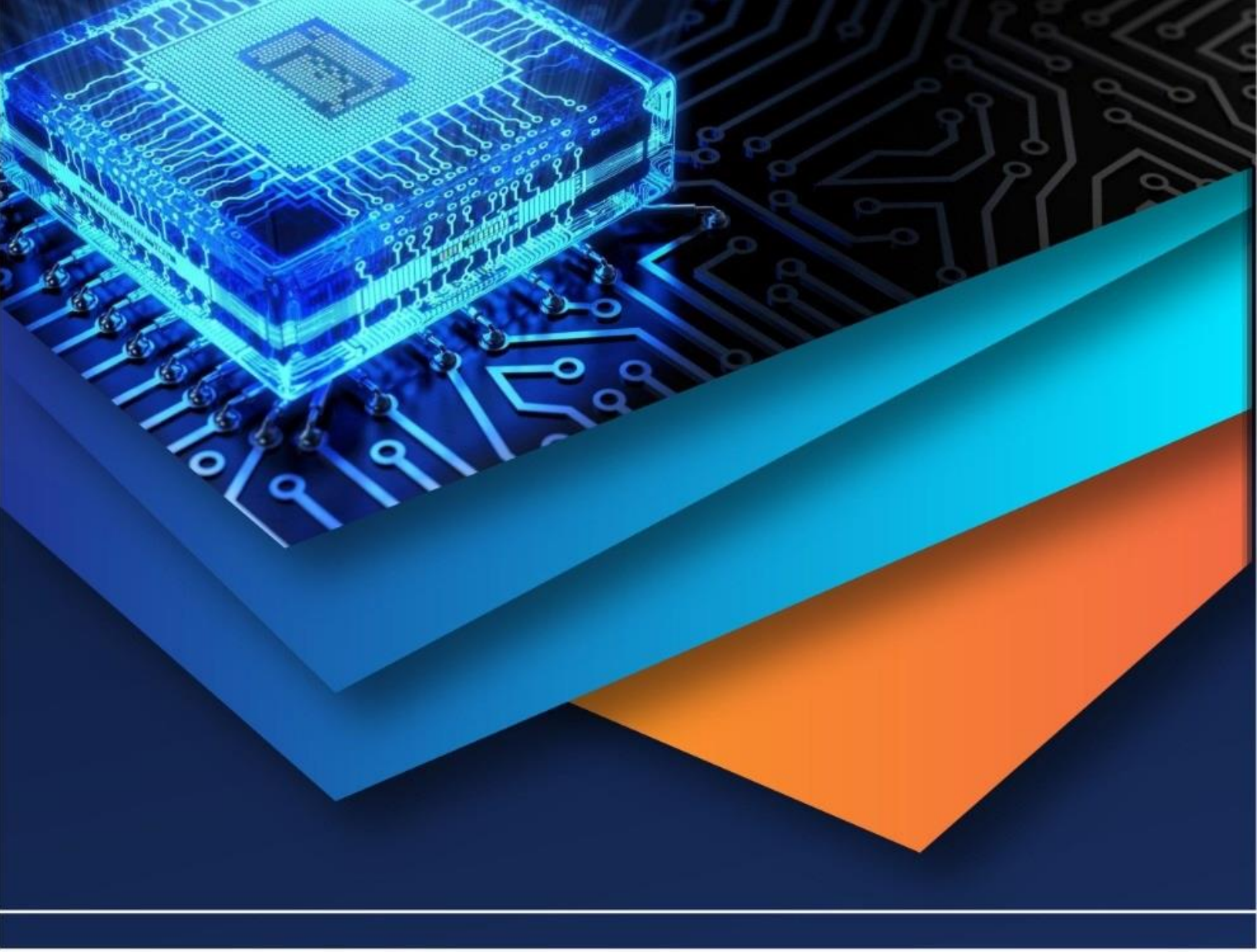
The ways of this examination in regards to VPAs is expected to uncover a review on how and to what degree these gadgets may be utilized in human-PC cooperation and learning. In this association, the working frameworks of the VPAs specifically Apple's Siri, Google Now and Microsoft Cortana are modified inside the setting of AI. This proposition presents the structure of Next-Generation of Virtual Personal Assistants that is another VPAs framework intended to speak with a human, with an intelligent structure. Likewise, the VPAs framework will be utilized to build the connection among clients and the PCs by utilizing a few innovations, for example, motion acknowledgment, picture/video acknowledgment, discourse acknowledgment, and the Knowledge Base. In such manner, it might be proposed that the two gadgets (PDAs) and applications (IPAs) may be utilized as plausible devices for language adapting; so increasingly subjective and quantitative investigations might be directed likewise. In addition, this framework can be utilized in various errands, for example, instruction help, therapeutic help, mechanical autonomy and vehicles, inabilities frameworks, home mechanization, and security access control. A total model application was likewise created, highlighting a sensible realistic collaborator ready to demonstrate outward appearances and empowered with discourse combination and acknowledgment, face location and face acknowledgment for client ID. The colleague was additionally associated with a brilliant home collaborator stage,

consequently constructing a total "encapsulated" virtual home partner that, uniquely in contrast to most normal brilliant speakers, can "see" and "be seen" by the client and draw in her in a multimodal association.

A straightforward counter-analyze was set up with a free form of the virtual operator, demonstrating that the capacity to distinguish and perceive the client, just as the graphical interface, to a great extent improve the client experience. The displayed design is still being worked on and a few upgrades are being included.

REFERENCES

[1] Nil Goksel-Canbek and Mehmet Emin Mutlu, "On the track of Artificial Intelligence: Learning with Intelligent Personal Assistants."
[2] Aditya K, Biswadeep G, Kedar S and S Sundar, "Virtual Personal Assistance."
[3] Veton Kepuska and Gamal Bohouta, "Next- Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)."
[4] Giancarlo Iannizzotto, Lucia Lo Bello, Andrea Nucita and Giorgio Mario Grasso, "A vision and speech enabled, customizable, virtual assistant for smart environments."
[5] Y. Matsuyama, A. Bhardwaj, R. Zhao, O. Romeo,
S. Akoju, and J. Cassell, "Socially-aware animated intelligent personal assistant agent," in Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 224–227, Association for Computational Linguistics, 2016.
[6] M. Schroeder, E. Bevacqua, R. Cowie, F. Eyben, H. Gunes, D. Heylen, M. ter Maat, G. McKeown, S. Pammi, M. Pantic, C. Pelachaud, B. Schuller, E. de Sevin, M. Valstar, and M. Wllmer, "Building autonomous sensitive artificial listeners," IEEE transactions on affective computing, vol. 3, pp. 165– 183, 4 2012. eemcs-eprint-22932.
[7] B. Martinez and M. F. Valstar, Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition, pp. 63–100. Cham:Springer International Publishing, 2016.
[8] S. Arora, K. Batra,and S. Singh. Dialogue System: A Brief Review.Punjab Technical University.
[9] A. Y. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, and A. Y. Ng, "Deep speech: Scaling up end-to-end speech recognition," CoRR, vol. abs/1412.5567, 2014.
[10] KITT.AI, "Snowboyhotworddetection.", 2018.
[11] D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar, and A. I. Rudnicky, "Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices," in 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 1, pp. I–I, May 2006.
[12] P. Viola and M. J. Jones, "Robust real-time face detection," Int. J. Comput. Vision, vol. 57, pp. 137– 154, May 2004.
[13] Itseez, "Open source computer vision library." https://github.com/itseez/opencv, 2015.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, June 2016.
[15] MycroftAI, "Mycroft, an open source artificial intelligence for everyone." https://github.com/MycroftAI/mycroft-core, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY