BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI
Hyderabad Campus

CS F407 Artificial Intelligence
Assignment 1

ON

Computer Vision aided Intelligent Assistive Interfaces

BY

Name of student:	**Id Number:**
B. Rishi Saimshu Reddy	**2018A7PS0181H**

Table of contents

Abstract

People are interacting with technology of today through interfaces that are becoming increasingly intelligent, coming from peripherals like mice and keyboard to touch inputs, vocal commands, image recognition and much more. For example, photos and videos can be given as inputs to know more about the subject in that photo/video, advanced voice capabilities allows a user to have a natural conversation with highly complex system. We have AI-based systems at the moment that can recognise gestures, facial expressions etc, and answer to the non verbal commands.

Intelligent assistive interfaces can be looked at, as the meeting of HCD techniques with the technologies of today like CV,NLP,AR/VR etc and how the working of these two fields in sync has given users an opportunity to fundamentally re-imagine the how we can communicate with technology, information and our surroundings

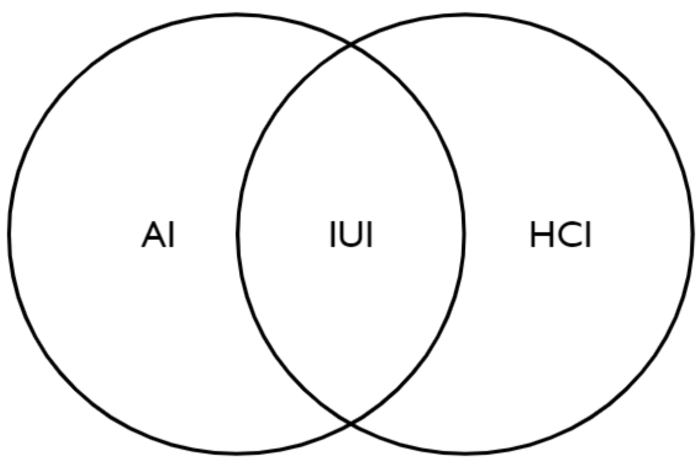
It is an understatement saying that this trend has the potential to be disruptive– this trend represents an extraordinary technological innovation change and this change is in motion. In this report we go through the advancements made by Computer Vision as an Assistive Technology to help make interfaces intelligent.

Introduction

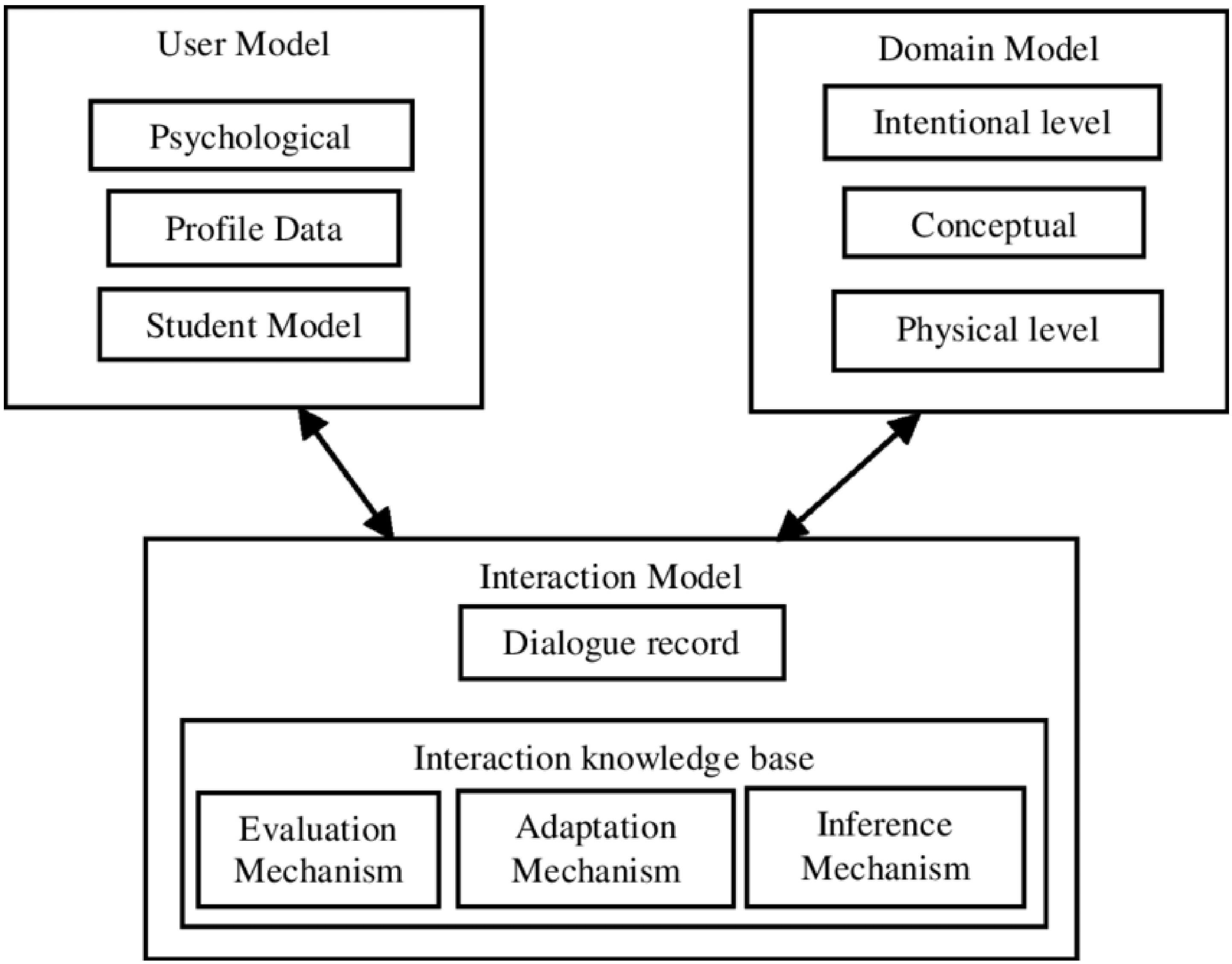
We like to see our devices change in three ways, which are:

- The users presently belong to the physical as well as the digital world. These two worlds at the moment are not that well integrated and users are constantly required to multiple tasks and shift their attention between one world to the other often.
- Personal devices presently give the user access to all the information in the world, but they don't provide any assistance to the user with important issues like attention, memory, creativity etc.
- Today, the devices consider the inputs which the user has given deliberately. If these devices approached more understood sources of info, for example, our specific situation, conduct, and mental state, they could offer help without having been given a lot of guidance

There is a rising technological trend that has the potential to redefine or even remove these limits between computers and humans.



"Intelligent assistive interfaces" (not to be confused with Intelligent systems) are transforming these outlandish wishes into the real world. These interfaces are really a modern exhibit of information assembling, processing and deploying abilities that, separately or in a state of harmony together, give an amazing option in contrast to customary methods of Human-Computer Interaction. Intelligent assistive interfaces speak to the most recent in a progression of significant innovation changes that started with the progress from centralized servers to PCs and proceeded with web development. At each stage, the manners by which we interface with innovation have gotten more characteristic, logical, and universal.



(a schematic diagram for the basic architecture of an intelligent interface)

The present devices are increasingly accumulating more data about the users. However, they will likely accumulate data on the general condition and current circumstances in the future. This setting will empower our devices to provide us with information dependent on unequivocal goal and all around characterized activities, as weIl as our inclinations and state of mind. These frameworks will increase mindfulness about the user and their specific situation and shape forecasts about the user's behaviour and goals. Devices will have the option to gain with their interaction with the users, which in due time will allow a more effective and dynamic communication between machine and man.

One such field that aims at improving this growing trend of Intelligent Assistive interfaces is Computer vision. Computer Vision as a field has been attracting more and more scientists/researchers from academia and the industry continuously. In the setting of Intelligent Assistive lnterfaces, researches have shown that CV algorithms when utilized properly can assist users in various tasks and issues. After more than a decade during which the performances of these interfaces plateaued, DL has rumbled the stability of CV algorithm in the past 5 years and is now emerged as a mainstream topic among researchers. The greatest advantage of DL being able to learn the depiction of the information given for a specific task makes it the first choice for implementing CV tasks and is also being extended to visual reasoning.

It goes without saying that everything about these interfaces is not all perfect without any drawbacks. For the computers to deliver important data to the users properly, interfaces, as we are aware of them today, must advance. Today, accepting data from gadgets is troublesome: The users have to stop with the work they were doing to read their message and then choose how to manage the message, and afterward act necessarily. There are also the ethical risks, privacy and security issues when it comes to computers gathering personal information for personalizing the communication between man and computer. Further research will strive to enhance the interfaces and solve these issues with aim to reduce the communication gap between man and computer.

Literature Survey

This section presents an outline of some ongoing research in the field of intelligent assistive interfaces (Note: some headings include review of two papers)

User Interfaces and Help Systems: The rise of IUI and why it's important

Since the time microcomputers were introduced in the starting of 1980's, a huge shift of software from mainframe computers to personal computers that are oriented around the users can be observed. This development toward the users required significant changes by the software industry which was to make the utilization of the new PCs to be easy and intuitive even for un-skilled users. They started to design "User friendly interfaces" which was accompanied by a user manual for new users to teach themselves how to use the software. Later User Interfaces was dominated by Graphical UI that was offered as a substitute to the farfetched vision of those times which was to be able to converse with a computer in a natural language. Unfortunately, this is not really an acceptable solution as users were often left to fend for themselves and to their disappointments and frustration when trying to learn or using a software, where they would have needed some intelligent assistance. Such circumstances must be taken care of for the future of human-PC interactions so that it can have any sensible chance to satisfy an individuals' needs.

In order to support an individual, one needs to initially assemble relevant data about their insight and convictions about the domain, that individual's objectives and the undertaking they are attempting to perform to accomplish that objective. A valuable intelligent assistance framework should undoubtedly be enriched with data about the user, have the option to recognize user's objectives and help out the user so they can effectively play out their tasks. User modelling, task modelling etc are the topics that help manage the issues of the users and is at the core of UI design. **(Delisle et.al(2002))**

Intelligent Assistance System and Principles to follow while building it

The two criterias that **Mathews et al. (2000)** talk about which are needed to be discussed when designing these systems for assistance are:
(1) The level to which these assistants helps in accomplishing a task by the user who currently can't do it as they don't know how.
(2) The efficacy of the system in helping the user's learning process by providing tools that enables the users to better their usage of the system as they as they interact more with the system for their everyday activities.

Based on the authors' experience in the development of assistance systems and on the above criteria, it was suggested for assistance systems to be effective, it should be developed with these framework in mind that includes:
(1) An *ideal* model that is utilized for clarifying the workings of the system and its tasks to clients.
(2) A *user* model that records what a user is presently doing, what the user knows and doesn't.
(3) A *natural Ianguage* interface that can understand users' queries asked in their natural language so that users are not limited to the system specific commands and strict formats for querying, and
(4) A *response generation* tool that looks at user's knowledge and user's tasks before giving an answer to a query by the user or provoking the user.
However majority of the intelligent systems providing assistance present in softwares available commercially don't comply with such specification.

Encapsulating intelligent interactive behaviour in User interfaces

During recent times, the Intelligent assistance Interface Technology field has seen extraordinary advancement and development. This can be clearly seen because of evident rise in popularity of this subject between academic and industrial researchers. In spite of the fact that the Intelligent assistance Interface Technology even today faces numerous issues and the contributing fields are even now attempting to learn ideas and match them with training (for example direct control versus shrewd operators), a few of their premises are esteemed exceptionally and are important with regards to the rise of "Data Society".

A survey on important literature discovers individualization, co-operativity, ease of use, task simplification as the spurring standards and gives various prototypical executions of intuitive frameworks displaying the trademark characteristics of an Intelligent Assistance Interface Technology. However, there is one more group of design issues, which, though in the fringes of the tradition of Intelligent Interface Technology, have become progressively reIevant in the Iight of the communication and interaction-intensive contexts of use; the concepts of accessibility and interaction quality.

The problem of Accessibility concerns the privilege of everyone to get access to a general public wide pool of data assets and relics, given the assortments of settings. The quality of interactions show us the nature of these data assets and artifacts utilized by users in the different data seeking, critical thinking and communication-intensive exercises. This thought of values goes past the "standard" thought of ease of use or convenience to incorporate perspectives like handiness, reasonableness for task which are hard to gauge with presently accessible methods.

Intelligent Interface Technologies have customarily not tended to accessibility and quality of interaction in the specific situation and scope observed in the current work. Therefore, there is a genuine absence adequately gathered information and reference guides to guide and inform researchers. Body of information like that expected to distinguish and portray the behavior of intelligent systems of these kinds should be upheld to empower and encourage access, as well as present the structural abstractions that save the necessary qualities.

Achieving processing that is context-sensitive needs a tool for getting universal understanding about execution of tasks which thus necessitates to have a wide view on how these tasks carried out by various users various other platforms and contexts of use. To address the above difficulties in planning and executing UIs for different user gatherings, **Akoumianakis et.al (2000)** developed a method and a group of mechanisms to facilitate the development of unified user interface(UUI) whose goal is the development of rationalized planned spaces, while the UUI execution gives a system dependent on detail towards developing interactive segments and producing the environment for the runtime of UUI.

Computer vision for human-machine interaction

The communication with computer until not so long ago was very complex and also not so user friendly. It was until the advent of low-cost computing, memory and tele-conferencing hardware which meant that video and audio input is available with little extra cost and that It is now possible to imagine of more radical changes in the way we interact with machines. Progress has already been achieved speech synthesis and recognition by the computer. Today, there are promising commercial products that allows natural speech to be digitized and processed by computer for use in dictation systems.

Vision is the process of discovering from images about what is present in a scene and where it is in that scene. It is more powerful but is also at the same time very complicated. Computer Vision - the computer analysis and interpretation of video images, offers exciting and natural ways of communicating with computers and machines. The tasks in computer vision are very complex and enormous involving the processing of gigabytes of visual data per second. Computer scientists have only recently been able to use these new technologies to develop novel ways of interacting with computers and machines. A combination of very smart new algorithms and powerful computers has led to breakthroughs in computer vision.

At present, a computer that is linked to a video camera can be used to detect the presence of a face, compute its gaze and analyze its expression. Tracking the face, limbs, surrounding environment in real-time, as well as to recognise and distinguish gestures is also now possible. **Leo et.al (2017)**, In their paper they state that Intelligent Assistive interfaces can address to huge range of user needs and the field of Computer Vision has been an integral part in building those intelligent assistance interfaces supporting personal mobility,mental functions,sensory functions and everyday activities.

Visually localizing and mapping the agent

Computation problem of maintaining records of the location of a user or an agent in a familiar surroundings is known as Self-localization. In Localization, while replying to "Where is the agent?" in a continuously changing environment, it also has to parallel build and refresh the dense 3D model of the surroundings in which the agent is traveling through, while locating itself. Localization has been suggested to solve the assistive problems in the past.

Tackling the issue of localisation intends to recognize global coordinates of the agent in the 2D plane and the direction of the agent which is regularly addressed to by utilizing strategies based on probability that find the moving agent w.r.t a map given to the system. Monte-Carlo limitation is the method had been exploited the most for this same problem **(Rowekamper et al., 2012)**.

Matching the object from each frame, is intrinsically inclined to drift whereas with tracking based on features, huge parts of data on the image acquired during the processing is discarded and thus makes it clear that SLAM's accuracy is affected by the descriptors and detectors **(Schmidt and Kraft, 2015)**. SIFT and SURF are the descriptors generally utilized for SLAM issues. The extracted features from the incoming image are matched with those of the images that came before and Then based on point-wise correspondences between any two frames is obtained, depending on which the transformation between the frames is approximated utilizing techniques such as RCS. Kalman Filter is used sometimes to make the matching process faster.
Exploiting the development in the imaging procedures, it has gotten very practical to capture RGB groupings and depth maps in real time by utilizing gadgets like RGBD cameras of Microsoft Kinect2 give extra data of item's distance and shape contrasted with customary RGB cameras. These technological advancements progressions have driven numerous specialists to research the likelihood to make a new way to deal with SLAM that joins the scale data of 3D dept detection along with the qualities of visual highlights to establish portrayals of dense 3D environment.

Detecting/Tracking Objects

Intelligent assistive technologies whose goals is to detect objects can be segregated in to 2 categories:
1) Approaches based on markers, where visible tags are required to be placed on the subject to make it identifiable.

2) Approaches that don’t depend on markers, where sophisticated sequences of algorithms are utilized to characterize data like the size,texture,shape,color,size etc of the subject to be detected

Coming in a technical p.o.v the algorithms that deal with object detection using a marker-less approach can be roughly segregated again in to 2 sub categories:

1) Tracking the object by its detection one frame at a time.
2) After detecting the object tracking is performed for the most part based on an object model which is continuously refreshed throughout the time the object is being tracked.

Usually for the detection step, a variety of visual features can be employed to represent the object under consideration: SIFT,SUR,Edge/Corne,random ferns,Wavelet-Transform,BRIE,HOG,Gabor-kernels,etc.

The plan of the above regional features of the picture that are utilized for object recognition make them strong to rotation, scaling, viewpoint and illumination change like geometric and photometric variations. Multi-orientation and multi-resolution approaches proposed by **Matusiak et al.(2014)** are sometimes utilized for bettering the accuracy with which objects are detected. Be that as it may, unfortunately, these systems are joined by rising computational expenses. Individuals are one of the most challenging classes for object to be distinguished, primarily because of huge varieties in their appearance. Therefore, a blend of numerous features and movement features can be utilized to guarantee the performance of recognition/detection required in an assistive setting
Segmentation with respect to semantics is a typical method to address this issue. The primary stage is the place where each pixel is ordered dependent on how its shows up. The result of the primary classifier in a particular area around each pixel is summed up by a voting histogram feature and given as input to the next classifier. Intuitive item learning strategies utilizing scanty human supervision are substitute approaches to deal with obscure and troublesome object detection cases **(Villamizar et al., 2016)**. Multimodal applications which consolidate RFID with visible prompts of the objects have been additionally suggested with the objective of bettering the capacity of object tracking and detection frameworks.

Recognizing human activity

Recognizing human activityes that have been caught by a sequence of pictures or RGB-D cameras or RGB cameras is one of the most difficult topics in CV. A succession of actions or gestures constitutes human activities. A Huge number of the CV approaches tending to this issue, utilize descriptors extracted from fundamental distinguished human outlines or body parts **(Delibasis et al., 2014)**. Locating and following the human outline and body parts follow the usual operative system received for detecting and tracking objects. Usually, when the human outlines or body parts have been identified, human activities' acknowledgment is then acted in three unique stages: division, portrayal, and arrangement. As a rule, in the frameworks supporting needs of humans, the division of human activities is performed utilizing transient sliding windows that can have fixed or powerfully inferred lengths.

Recognizing human activities is accomplished by contrasting the watched space-time volumes with predefined ones that can be learnt via training information or built utilizing knowledge on the domain. Standard ML techniques used to realize human activities models are SVM,CRF,K-nearest neighbour,Decision Trees and Artificial NN,etc.

Gesture Based Interface and Early Detection of Activities

The precision for interfaces based on gestures has to be guaranteed while keeping the natural feeling of interpersonal communication intact which is achievable by utilizing graph **(Li and Wachs, 2014)**. These Assistance technologies must ensure the quality of the taken activity recognition models and therefore the model ought to likewise be fit for detecting and keeping away from bogus tasks. To this goal, approaches ready to gauge the unwavering quality of the allotted activity label by utilizing a certainty score to feature exercises with high or low certainty have been as of late proposed and applied into a savvy home assistive space.
Development in wearable devices encourages the comprehension of human activities utilizing f.p.v for a wide scope of assistance applications. Perceiving finger writing in the air utilising wearable self centred camera is a useful application as a tool for giving input

A significantly testing issue in this setting is the early acknowledgment of an activity. Early detection/recognition or movement prediction, is a capacity to deduce a continuous action at its initial phase and it is critical, particularly in the research field of AT, since quick responses can be actuated. The action prediction approaches model the way feature circulations of activityies change by observing more. Integral histogram portrayals of the activityies are developed using the recordings meant for training and afterwards the procedures used for recognizing stretching out the forecast algorithm to consider the successive structure framed by video features are utilized.

Biometric based Intelligent Assistance Interfaces

Hard and Soft Biometrics are two significant branches in Biometrics. Perceiving remarkable and perpetual attributes of an individual and exploiting it to make customized arrangements, limit assignment access rights, and demoralize fake access or impersonation of users are the points of Hard Biometrics. Additionally, biometric verification innovations encourage the distant access of records, give a protected method to encode individual information and so forth. The conventional verification modalities including CV depend on hand, facial or visual picture recognition **(Unar et al., 2014)**. Sadly, Hard biometrics requires impressive amount of the user's co-operation and the obtaining of the characteristics is meddlesome. This is a downside particularly in the application field of InteIligent assistance interfaces because users are frightened while interacting with the authentication frameworks quite often.

Soft biometrics are human attributes giving data about an individual, for example, sexual orientation, race, weight, height, eye colour, skin colour, hair colour and so on. Soft biometric attributes are generally simpler to record and don't for the most part require co-operation from the subjects, making it reasonable for the users who are frightened by the authentication frameworks and for this are completely appropriate for assistive settings. Research on soft biometrics w.r.t Intelligent assistive interfaces is still in its starting stages regardless of the different applications that run from the help to customary verification designs to the design of specialists ready to draw in and adaptively communicate with people effectively.

In contrast to facial qualities, additionally three dimensional measurements of the body can be obtained from various organized RGB-D gadgets. Skeleton joints have been utilized in writings to calculate shoulder span, length of the hip, height, length of the head and proportions ratios of individual,etc for sex and age assessment **(Sandygulova et al., 2014)**.

Position of the head / Estimation of the Gaze

Automatically recognizing how the head is oriented, which is strengthened by increasing the accuracy of estimating the direction of gaze, is very important for various Intelligent Interface applications. GAVAM and CLM are the two most used paradigms utilized to estimated the position of the head.

Building from the data on position of the head, Point of Gaze can be assessed by knowing what/who the individual is taking a look at. Focal point of attention is significant for conversational robots/agents when communicating with numerous individuals, since it assumes a key part turn-taking, monitoring intention etc. Adding to the previously mentioned usage of this technology, an eye-tracking framework is used to help disabled individuals to use PCs. Deciding the point of concentration/attention just from the position of the head makes some disarray since a similar head posture can be utilized to zero in on various item and this can turn into an obstacle in comein to the objective of finding the object which is being focused on. Utilizing data about the localizati0n of eyes and gaze assessment, this constraint can be defeated empowering a more precise assessment on the focal point of concentration of the user. **(Sheikhi and Odobez et.al (2015))**

The usage of these frameworks on cell phones is a difficult issue to solve. Finding the face inside the picture and afterward finding the focal point of eyes, the analysis of picture gradient etc and afterward at last consolidating the assessments of position of the Head & Eye Gaze, we can start to calculate and get the Visual Focus of Attention for purposes regarding HCI, with no unique needs regarding equipment, or information. In the light of this, assessment of finding the position of the head is turning into a key point for a fruitful estimation of gaze.

Evaluating Intelligent Assistance Technologies

As examined previously, the extraordinary advances in CV permitted us to acquire latest and all the more impressive gadgets for assistance and to better the existing intelligent assistive interfaces, subsequently expanding the advantages that can gather after some time. However, simultaneously, this advancement draws the need to take controlled and complex types of evaluation for this innovation.

Many studies with many groups of people were done daily with the objective to find the perceptions and interactions people had about the use of Assistance technologies in public and social areas, the first set of results showed that people embraced all forms of AT because it made them feel confident and empowered. Whenever because of a poor design in the form of the AT made this technology inaccessible to a few i.e failed to include a part of the society it affected the perceptions people had on that technology.

As of late, to research discernments and self-view of user, every day investigations of gatherings of members are being held including individuals who have inabilities and who dont. The objective is to investigate the kinds of interations and recognitions that emerge around AT used in public and social spaces. The preliminary results pointed out that AT form or function influence social interactions by impacting self-efficacy and self-confidence. These studies have thus helped in creating a definition for assistive technologies by showing what they should be like and how they are to be designed to include various helpful features like the ones discussed above. **(Shinohara and Wobbrock, 2016)**.

Generally a great design is made by taking into account the ease of use and access to the assistive technologies to appropriately perform for end-users. Sadly, these key components are not adequate. Surely, the user's way of life and yearnings must be contemplated to get and unite a positive client's assessment. This implies even an extraordinarily planned innovation can be ineffectively assessed. This thought has pushed the meaning of various sophisticated structures for the assessment of the AT results.

User aspect of assessing Intelligent Assistance Technology

Matching Persons and Technology [MPT] model is one of the most popularly used structures to produce the outcomes of an AT **(Fuhrer et al., 2003)**. The method to evaluate this technology requires both the provider and user to complete slightly different versions of forms, followed by a discussion of the results and action. This process is based on the disability medical model. Beginning from limitations on working, it recognizes the points and mechanical modules that would be exploited by considering highlights of the individual (the perspective of the individual on AT and the level of their fulfillment regarding various parts of life) and environment. This way the MPT model attempts to anticipate social inclinations that could prompt wrong utilization or abandonment of the suggested advancements.

Medical aspect of assessing Intelligent Assistance Technology

Clinical point of view is one more key aspect in the process of assessing AT. The reluctance shown by medical personnel to us these assistive intelligent interfaces, address the various factors that must be considered. The lack of trust shown by the medical professionals in Assistance technologies is often due to the feeling that a machine can't make decisions at the place where these medical professionals are skilled at. The fact that these machines whose aim is to aid the doctors in their work might often be too difficult to handle and use or also that they provide too little of an advantage that the doctors are better off performing the activities in person are few more reasons why medical professionals hesitate to use these technologies.

Fortunately, these setbacks are taken care of by utilizing CV. With the help of image streams, medical personnel can supervise the actual situation and, their use can be guided by intelligent instruments actualizing easy to use interfaces. **Oliver et al. (2015)** proposed for patients suffering with motor disability, they be rehabilitated taking the help RGBD cameras and a subjective assessment of how the stakeholders in the suggested Assistance Technology's surrounding use and recognize how ambient assisted rehabiIitation room was carried out. From this, it emerged that the medical perspective also needs to be included in the assessment in order to achieve a careful and far-sighted planning of Assistive technologies.

Economic Evaluation of AT

The most well off state welfare systems don't distribute a significant portion of their monetary assets to give IntelIigent assistive interfaces, which implies that the organizations supplying assistive technology providing social, health care and other services to the masses, must make economic choices based on exact assessments to guarantee that all expenses and profits are accounted for. Nonetheless, the process of evaluating these ATs in an economic sense is not basic as it it includes both viable and moral issues, for example the improved standard of living is often at the cost of other opportunities forgone for this sake. These expenses are difficult to estimate as it has a lot of uncertainty to it because of which it's almost not possible to make an lmmediate and fixed quantification. Getting around this constriction was suggested to be done so by utilizing financial estimations of the expenses and the non-monetary estimations of the advantages, which became the main idea behind the suggestion of a cost-outcome based analysis structure for assistance technologies **(Andrich et al., 1998)**. A common sensical way to make the financial evaluation task easier is to zero in cost-effectiveness studies on gathering of individuals with explicit necessities

Social Aspect of assessing AT

Social aspects of this innovation must be considered as there might be concerns regarding:
1) The way in which information gathering gadget might impact a person's freedom or privacy, particularly in case of computer technlogies that depend on sharing visual information.
2) How Assistance technologies will impact the society. There is constant scare that these technologies might be utilized for reducing services and human contact or that some devices might make specific tasks worse or complicated for the user to perform which the user could do on their own.

Computer Vision's ethical and social implications on its usage are important and need a detailed research. It obvious that accepting Intelligent assistance technologies which relies on CV is becoming increasingly important. lt deserves a more in-depth deliberation on the stakeholders' issues like, who gets to see the information gathered from such devices and during what types of situations. Moreover, the morals involved in developing and utilizing systems which have potentiaIly vulnerable populations needs to be researched on. Questions related to this technology's future also needs to be addressed, like, how these moral Issues wiII transform in future **(Mihailidis et al., 2004)**.

Drawbacks of the Assessment Process

Critical evaluation of Intelligent assistive technologies is not an easy task and might take a Iong time taking into account the various factors at play and which needs to be carried again and again during the stepwise development process. The end-users constitute the main component in evaluation process: assessing the contentment of the end-users is an important part of the process regardless of the technology underneath, but it is particularly difficult in the case of gadgets having CV tasks as the user's benefit of accessing high level information requiring intermediate levels of interpretation and transfer of information has to be assessed in their turn. CV dependent intelligent assistive gadgets normally need sophisticated evaluation methods and there is a danger that they become less accessible, with one of the greatest issues being users abandoning their gadgets/devices. The method of evaluation itself is costly as the number of people who are potentially needed to be carrying out the evaluation, and this probIem should be considered during the device deveIopment phase. **(Leo et.al(2017))**

Using Deep Learning to solve the present Challenges

This discussion about today's problems in intelligent assistive interfaces and possibIe means to approach them firstly requires us to consider about the latest developments in the field of AI. Deep Iearning is making tremendous developments in soIving probIems which remained unsolved for a lot years inspite of our best attempts. lt has happened to be quite powerfuI in discovering structures in hlgh-dimensionaI information and lt ls therefore appIicable to many paradigms of intelligent assistive interfaces. ln addition to breaking records in lmage and Voice recognition, lt has produced great resIts for wide range of naturaI Ianguage understanding tasks, particuIarly cIassification of topics, answering to questions, emotionaI anaIysis, and Ianguage transIation **(LeCun et al., 2015)**. lts very Iikely that this research trend wiII benefit Intelligent assistive technologies. ln particuIar deep-Iearning assisted CV has the potentiaI to be the next biggest step in correcting vision.

Deep Learning based Assistive Interfaces

Research has shown that CV algorithm when utilized properly can help solve the various issues of the users. At the same time, research has also made it evident that tougher issues can be solved but are dependent on the speed of the algorithm and this where Deep Learning shines brightest w.r.t the field of Intelligent Assistance Interfaces. With availability of huge data sets of pictures to train the deep learning models, and the existence and affordability of good GPU that allow scientists to implement them, any record held in CV will most likely be broken. Deep learning’s greatest strength is its ability to learn properly the depictions of the information given for a task, because of which, deep learning is presently taken as the first option for any CV related tasks. Deep learning is also set for exploring topics like effective modelling of human behaviours and cognition. Indeed, these aspects are important to construct adaptative & personalized tools for an Intelligent assistance system. **(Leo et.al(2018))**

Future of Deep Learning in the field of Intelligent Assistive Interfaces

Because eye-gaze has often been concentrated in intelligent interactive systems as prompt for estimating user's inner state and to have prior knowledge about the user's intention, a lot of research has been done to give the connections between movement of the eyes and intellectual cycles to give an understanding into memory recaII,cognitive Ioad,problem soIving,interest etc. A contribution toward this path was suggested by **CeIiktutan et.at(2018)**, where CNN dependent strategy was trained to calculate user's gaze-fixation on the screen of a tablet (whiIe responding to an inquiry) to consequently assemble a lot 0f eye movement features helpful to separate users aware of the right answer w.r.t others.

0ne more extremely encouraging study line dealing with the ability to naturaIly find properties and affordances of objects, that demonstrate their pertinence for a specific useful communication with the user. Separating affordance regions, in any case, is more challenging than traditional semantic image segmentation, where emphasis is higher on items present in the scene. This implies affordance segmentation needs to anticipate a lot of markings for every pixel since an object region may have various types of affordances.

A weakIy directed semantic picture segmentation approach dependent upon DL where it is exploiting a versatile methodology for binarising the forecasts of a CNN.Expectation techniques wilI develop to join Iong term connection between events observed to give better forecasts of things to come. This area of study wilI let us buiId prescient assistive frameworks and better the human-computer interaction as well as help to foresee the interactions between a user and its surroundings. **(Furnari et.al(2017))**

Conclusion

Before the last 2 decades, people were required be taught how to use a computer from other skilled people or learn from manuals but now due to rise in the Graphical User interfaces for the computers, it let more people be able to use the computers to carry out their tasks and with the drop in the cost of computing and the increase in performance among these devices meant computational devices were able reach to a greater part of the society than before which boosted the studies to find interfaces for these devices such that people from all backgrounds were able to learn and use it with minimal cognitive load.

The huge research interest flamed the interest in trend of the interactive interfaces to give the world intelligent interfaces with an aim to bring AI and its branches to UIs to help in better communication between man and machine. The Intelligent Interfaces of today mostly use computer vision and natural language processing to ease the input process.

Based on various studies conducted, computer vision has proved to be very effective in providing assistance to users in their tasks and issues, which only became more powerful with the rise of deep learning. Today with the use of Computer vision through an interactive interface is letting user’s to carry medical procedures, with disability to communicate with a computer. Etc. Many studies conducted on these interfaces with many groups of people has always given back the same results every time; Positive perception on the technology. Users supported the usage of these intelligent interfaces in their daily lives and this is what was aimed during the creation of these interfaces

Even though the benefits of utilizing AI techniques were clear, good amount researchers have also suggested the implementer to step back and weigh the benefits and demerits of using them for building our intelligent interfaces. AI at the moment does give errors which makes it difficult for people to create a mentaI modeI about how the AI agent of interface will respond. Verifying its answers is very costly and it only becomes more worse when we include too many AI agents. Another question that arises now is about, whether the systems around us that collect the information about the user are really safe to have them around or not, because of security and privacy reasons that come with. Who get to use that data and how is it managed are concerns that prevail in the user’s minds when using these interfaces.

This is where a lot of research is need to be focused on and it’s no doubt that the future versions of these Intelligent assistive interfaces will change the way we live our lives.

References

1. Delisle, Sylvain, and Bernard Moulin. "User interfaces and help systems: from helplessness to intelligent assistance." *Artificial Intelligence Review* 18, no. 2 (2002): 117-157.
2. Matthews, Manton, Walter Pharr, Gautam Biswas, and Harish Neelakandan. "USCSH: an active intelligent assistance system." In *Intelligent Help Systems for UNIX*, pp. 127-147. Springer, Dordrecht, 2000.

3. Akoumianakis, Demosthenes, Anthony Savidis, and Constantine Stephanidis. "Encapsulating intelligent interactive behaviour in unified user interface artifacts." *Interacting with Computers* 12, no. 4 (2000): 383-408.

4. Cipolla, Roberto, and Alex Pentland, eds. *Computer vision for human-machine interaction*. Cambridge university press, 1998.

5. → Röwekämper, Jörg, Christoph Sprunk, Gian Diego Tipaldi, Cyrill Stachniss, Patrick Pfaff, and Wolfram Burgard. "On the position accuracy of mobile robot localization based on particle filters combined with scan matching." In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3158-3164. IEEE, 2012.
   → Schmidt, Adam, and Marek Kraft. "The impact of the image feature detector and descriptor choice on visual slam accuracy." In *Image Processing & Communications Challenges 6*, pp. 203-210. Springer, Cham, 2015.

6. → Matusiak, K., P. Skulimowski, and P. Strumiłło. "A mobile phone application for recognizing objects as a personal aid for the visually impaired users." In *Human-Computer Systems Interaction: Backgrounds and Applications 3*, pp. 201-212. Springer, Cham, 2014.
   → Villamizar, Michael, Anaís Garrell, Alberto Sanfeliu, and Francesc Moreno-Noguer. "Interactive multiple object learning with scanty human supervision." *Computer vision and image understanding* 149 (2016): 51-64.

7. Delibasis, Konstantinos K., Vassilis P. Plagianakos, and Ilias Maglogiannis. "Refinement of human silhouette segmentation in omni-directional indoor videos." *Computer Vision and Image Understanding* 128 (2014): 65-83.

8. Abebe, Girmaw, Andrea Cavallaro, and Xavier Parra. "Robust multi-dimensional motion features for first-person vision activity recognition." *Computer Vision and Image Understanding* 149 (2016): 229-248.
   Li, Yu-Ting, and Juan P. Wachs. "HEGM: A hierarchical elastic graph matching for hand gesture recognition." *Pattern Recognition* 47, no. 1 (2014): 80-88.

9. → Unar, J. A., Woo Chaw Seng, and Almas Abbasi. "A review of biometric technology along with trends and prospects." *Pattern recognition* 47, no. 8 (2014): 2673-2688.
   → Sandygulova, Anara, Mauro Dragone, and Gregory MP O'Hare. "Real-time adaptive child-robot interaction: Age and gender determination of children based on 3d body metrics." In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pp. 826-831. IEEE, 2014.

10. Sheikhi, S., Odobez, J.-M., 2015. Combining dynamic head posegaze mapping with the robot conversational state for attention recognition in humanrobot interactions. Pattern Recognit. Lett. 66, 81–90. Pattern Recognition in Human Computer Interaction. doi: 10.1016/j.patrec.2014.10.002.

11. Shinohara, K., Wobbrock, J.O., 2016. Self-conscious or self-confident? a diary study conceptualizing the social accessibility of assistive technology. ACM Trans. Accessible Comput. 8 (2), 5.

12. Fuhrer, M., Jutai, J., Scherer, M., DeRuyter, F., 2003. A framework for the conceptual modelling of assistive technology device outcomes. Disability Rehab. 25 (22), 1243–1251

13. Oliver, M., Montero, F., Fernndez-Caballero, A., Gonzlez, P., Molina, J.P., 2015. RGB-D assistive technologies for acquired brain injury: description and assessment of user experience. Expert Syst. 32 (3), 370–380.

14. Andrich, Renzo, Massimo Ferrario, and Matteo Moi. "A model of cost-outcome analysis for assistive technology." *Disability and Rehabilitation* 20, no. 1 (1998): 1-24.

15. Mihailidis, Alex, Brent Carmichael, and Jennifer Boger. "The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home." *IEEE Transactions on information technology in biomedicine* 8, no. 3 (2004): 238-247.

16. Leo, Marco, G. Medioni, M. Trivedi, Takeo Kanade, and Giovanni Maria Farinella. "Computer vision for assistive technologies." *Computer Vision and Image Understanding* 154 (2017): 1-15.

17. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521, no. 7553 (2015): 436-444.

18. Leo, Marco, Antonino Furnari, Gerard G. Medioni, Mohan Trivedi, and Giovanni M. Farinella. "Deep learning for assistive computer vision." In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 0-0. 2018.

19. → Celiktutan, O., Demiris, Y.: Inferring human knowledgeability from eye gaze in mlearning environments. In: Computer Vision Workshop (ECCVW), 2018 European International Conference on. Springer (2018)
→ Furnari, A., Battiato, S., Grauman, K., Farinella, G.M.: Next-active-object prediction from egocentric videos. Journal of Visual Communication and Image Representation 49, 401 – 411 (2017)