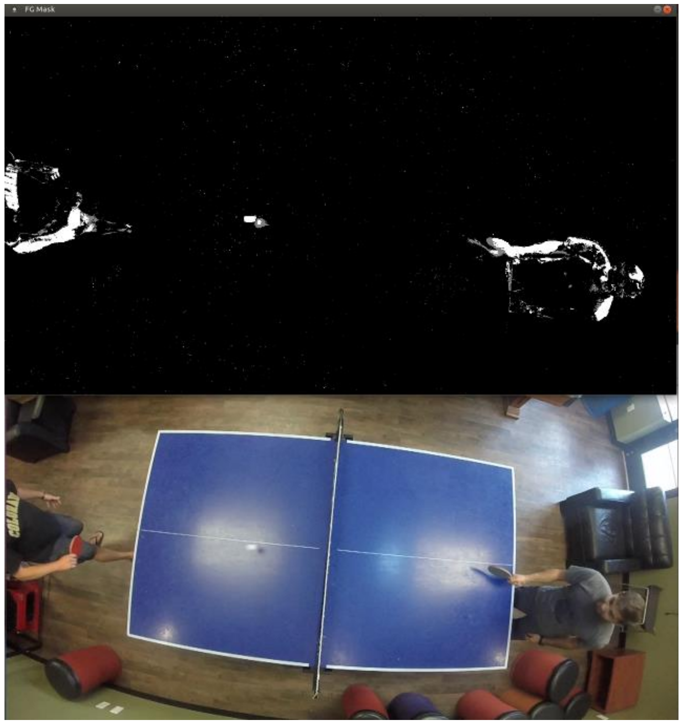
Human Computer Interaction Exercise #3

Andrea Gray
CS 415 – Dr. Siewert
DUE: October 4, 2019

1.a. After reading the paper, *Cloud scaling, Part 3: Explore video analytics in the cloud*, by Dr. Siewert I have compiled three main points that I gathered from the article. These points may not capture the main three points the author was attempting to convey, but during my reading these following points stood out to me the most. First, and mentioned multiple times in class, is that CV, computer vision, is not a trivial task at all and even though there has been much progress in the field in the past half-century, "most applications still cannot match the scene segmentation and recognition performance of a two-year-old child…" (page 2). This fact is very interesting because it is obvious that the CV development is still in it's beginning stages, and yet, as humans in an exponentially advancing and fast-paced world, it could be expected that the implementations of CV would be more advanced. Contrastingly, when reading the article and how it describes some of the implementations of CV today and upon inspection of the code examples and their outputs it is amazing and humbling to realize how complex even a simple implementation is. With that in mind, it makes one realize how the developments in CV thus far are notable. Subsequently, the article supports the fact conjectured above that photorealistic rendering, although impressive today, is nowhere near close to real time or interactive applications. The article goes on to say that "with sufficient data-centric computing capability leveraging the cloud and smart cameras, the dream of inverse rendering can perhaps be realized where, in the ultimate 'Turing-like' test that can be demonstrated for CV, scene parsing and re-rendering display and direct video would be indistinguishable…" (page 4). This gives the reader hope that while the technology today may not be able to support the idealistic computer functionalities, there is expectation that with time and greater progress in storage, networking, database, and computing the CV potential can be reached. Finally, this article gives a well-known example of computer vision in today's world: Augmented Reality (AR). Made popular by the recent public excitement of the app Pokémon Go, AR is an exciting development using real-time video analytics that is wedging itself into the common industry and developing progressively with popular demand. This article elaborates on this example by making a statement that proves how meticulous and precise real-time video analytics and CV must be. Dr. Siewert states, "… the ideal solution [of AR] provides [a] seamless transition from scenes captured with digital video to scenes generated by rendering for a user in real time, mixing both digital video and graphics… Poorly designed AR system distract a user from normal visual cues, but a well-designed AR system can increase overall situational awareness…" (page 5). This statement shows the reader that on top of the painstaking task of developing AR, the functionality of the system's human computer interaction must not degrade the usability of the system. While the general goal of creating an AR system is a common goal of many developers and researchers, the true test of success is the public's acceptance of the technology and its ability to aid the user substantially more than it could have the possibility of hindering the user. This was seen in the creation of WIMPs; where some companies excelled in their interfaces, some companies fell into ruin from a design that they might have found useful, but the user did not. Although, of course, the real-time expectations of AR and CV in general make the task of interfacing much more complex and weighted. In conclusion, this article has proven how interesting CV can be and how that while machine vision has helped automate the industry for years, CV now shoes promise for providing vision-based automation in the everyday world where the environment cannot be well controlled, like in machine vision, and where hard-time real-time systems are mandatory.

1.b. The main points of the OpenCV Background Subtraction Method Tutorial was that the method is used for creating a foreground image from a video usually for moving images within a capture of the video, the method is named such because it subtracts the moving object from the calculated background to present the object as a foreground outline, and that the method can be used on both movies and still images as long as enough data is provided for the still image option. The Background Subtraction code can be shown working below on a screenshot of the ping pong video:



The top portion is the background subtraction of the video frame below with the white portion of the background subtracted image being the moving elements from the video.

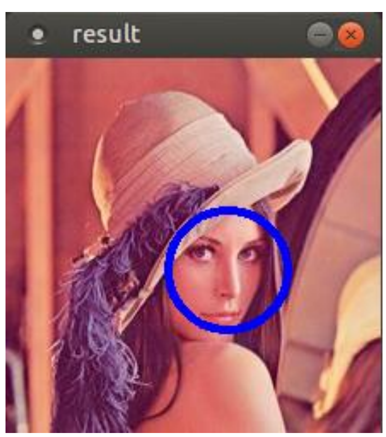
1.c. The Jetson OpenCV installation was tested by running the code. The Sobel Transformation is shown on Trees.jpg in the images below. From the OpenCV Tutorial on the Sobel Transformation, the way that the Sobel Transformation works is by using the derivatives of pixels within an image to detect an edge in the image. The gradient of a pixel in the image that is higher than its neighbors shows that that pixel represents an edge in the image. The Sobel operation uses this mathematical point to perform discrete differentiations, or to use the math explained above to detect all the edges in an image. The Sobel Transformation combines different mathematical methods to approximately compute the gradient of an image intensity function, or where the edges are, using the data of a certain pixel in an image and comparing that pixel to its neighboring pixels to see if the gradient of a pixel is a local maximum which would indicate an edge.

Trees.jpg before Sobel Transformation

Trees.jpg after Sobel Transformation

1.d. The OpenCV Haar Cascades for Face Detection Tutorial explains that Haar Cascades work by gathering input from a video and selecting a scene within that video for the code to run on. The code then translates the image to greyscale and detects a face within the frame using a multiscale detection cascade identifier included within the code as "objdetect.hpp". The code then determines where the eyes are by using the multiscale detection cascade identifier again and drawing ellipses around the eyes with the center point at the center of the eyes. This code was then tested on the famous lena.jpg and Dr. Siewert's CU faculty photo, both shown below.

Lena.jpg before the Haar Cascade Face Detection

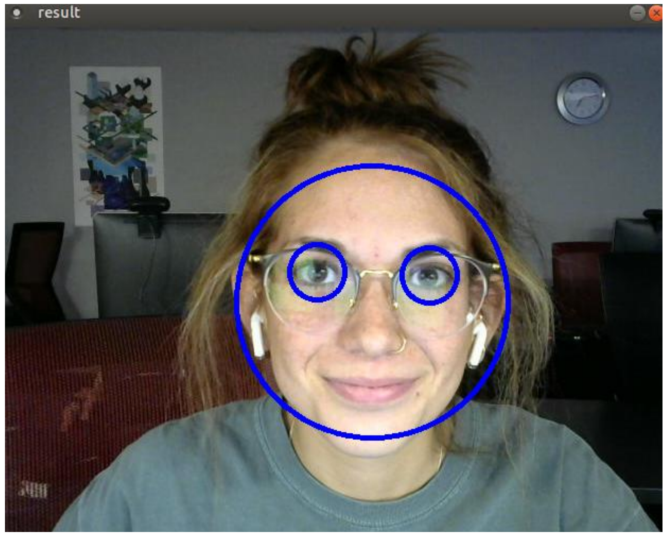


Lena.jpg after the Haar Cascade Face Detection

Dr. Siewert's CU Faculty Photo before Haar Cascade Face Detection

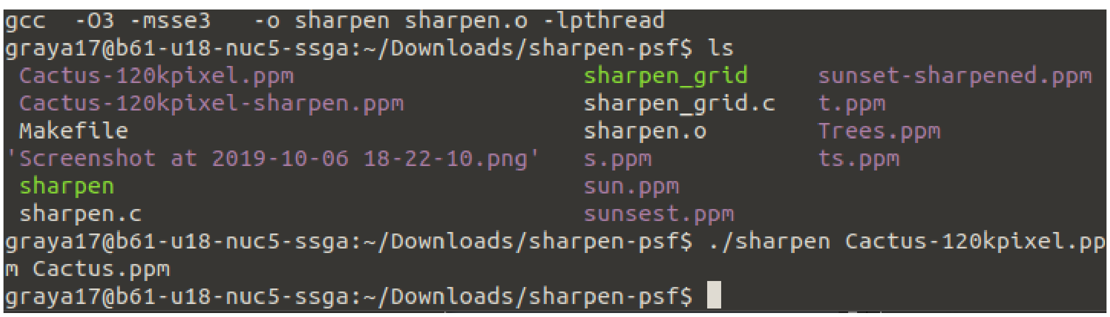
Dr. Siewert's CU Faculty Photo after the Haar Cascade Face Detection

Unfortunately, the face detection did not detect the eyes on either of the pictures. I am not sure what happened with the face detection code to make the eye detection not work. When ran with input from the camera on the computer though, the face detection worked fully; shown below.

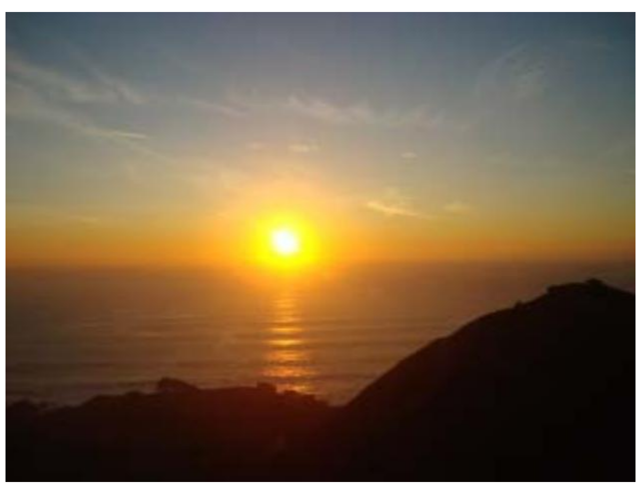


Face Detect working from Computer Camera Video Input

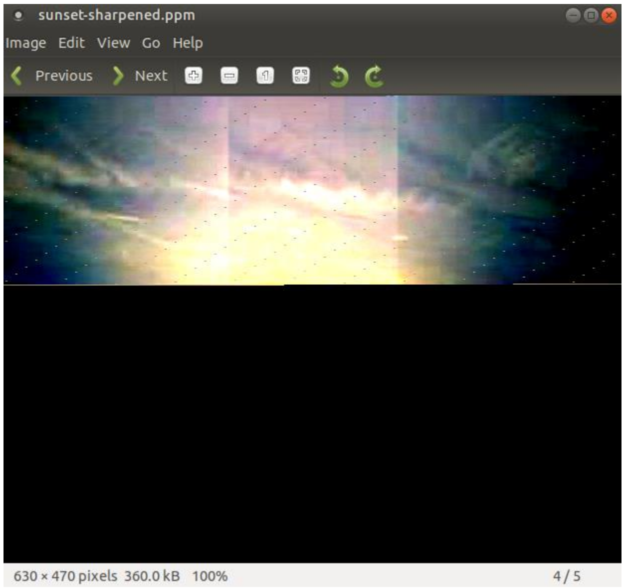
2.a. The code for image sharpening can be shown with a command line screen shot with the initial test ran on the 120K Cactus image included within the code file. The code was then ran on the sunset picture and below shows the before and after picture of the sunset in Dr. Siewert's publication, *Using Intel Streaming SIMD Extensions and Intel Integrated Performance Primitives to Accelerate Algorithms*.



Command Line from Sharpening Test Run



Sunset Photo before Sharpening

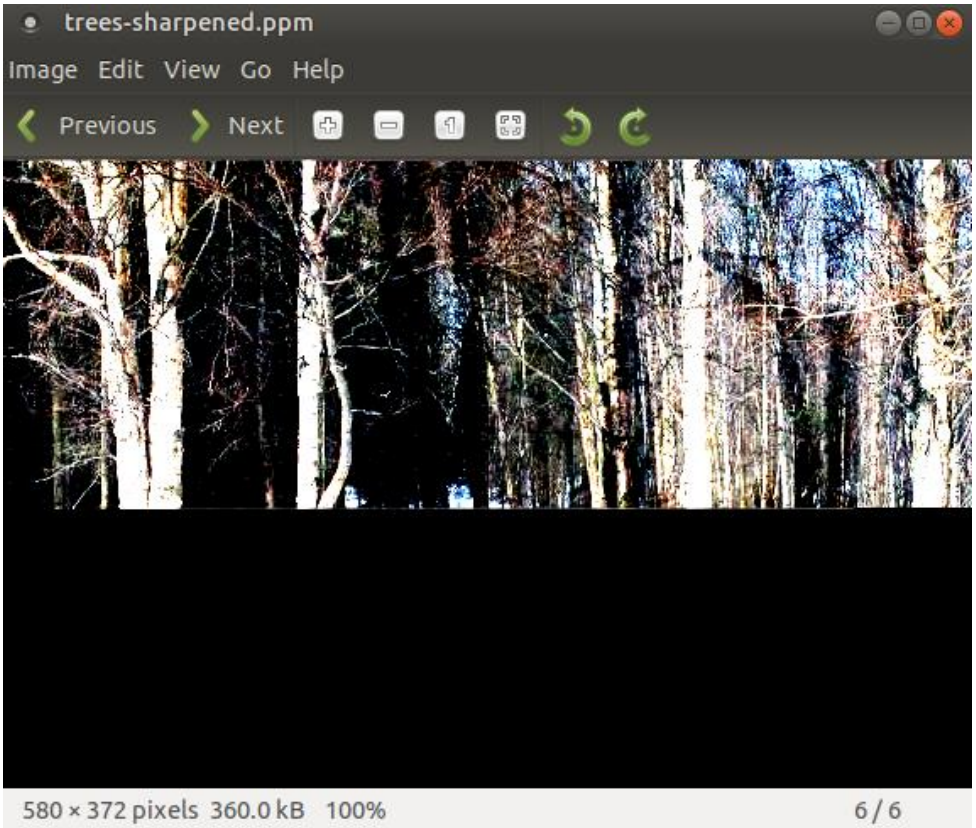


Sunset Photo after Sharpening

Unfortunately, the sharpening on the sunset image did not allow the photo to be loaded completely. I am unsure of the reason for this.

2.b. Modifications of the code, shown below, were made so that trees.jpg could be ran using the sharpening code in the previous question. To do this Trees.jpg was saved as Trees.ppm using GIMP so the code could run it. The output is below. The PSF (Point Spread Function) provides edge sharpening by taking a point of interest, a pixel corresponding to an edge of an image in this case, and decreases the brightness of the surrounding pixels, the pixel's region of influence, in a process called lateral inhibition in humans to make the point appear more prominent and therefore making the edges more prominent, or sharper. The PSF exaggerates the brightness of a point to amplify its presence in an image.

Trees.jpg before Sharpening using PSF



Trees Image after Sharpening

Just like the sunset image, the trees image after sharpening only loaded a portion of the total image.

2.c. The sharpen_grid code, after being read and ran, is different in comparison to the sharpen code as it uses threading for the image sharpening whereas the sharpen.c code does not. Threading is useful for real-time applications because it allows multiple things to be ran at once and it returns an error if the code does not perform properly whereas without threading the code could just return an inaccurate result without notifying the user of it. The sharpen_grid code may be better for real-time frame transformations because just from the test running the sharpen_grid code was more successful in creating a full sharpened image of the original picture than the sharpen code. The sunset and trees images after ran on the sharpen_grid code is shown below and the results show that the image may be grainier than the partial image created from the sharpen code, but the main success is the sharpen_grid produced the full image whereas the sharpen code couldn't. This would be more useful for real-time system because real-time systems require success and the sharpen-grid proved that whereas the sharpen code couldn't.


Trees Image ran through Sharpen_Grid.c


Sunset Image ran through Sharpen_Grid.c

3.a. I believe that we have still not progressed beyond WIMP in general purpose computing to 3D interaction because, as mentioned in answer to question 1.a., the technology for seamless interaction in 3D systems is not advanced enough yet, the platforms on which these technologies need to run on are insufficient for the computational load necessary, and the general public is more satisfied with a WIMP that has been “perfected” than a 3D Interaction that cannot run on their embedded device alone nor can have the desired interaction capabilities (not meeting the -ilities standards that the public will put 3D interfaces up to from their experience with WIMP).

3.b. I do not believe that the web or mobile computing is a distraction to the integration of 3D interaction, in fact, I believe that it has helped the public become excited about it. Since the recent advancements in 3D interaction and computer vision, that can be read on the web, I think people are expecting the interaction and excited for it. However, I still believe they are not satisfied with the progression that has been made with it up until now and will demand more before the public will accept it and begin to integrate it into their daily lives. Flat panels are a common occurrence in daily computing and moving on from them will be a major adjustment period for humanity; just as the movement from paper to the beginning of computing was a transition. Despite the expected acceptance and adjustment period, the real issue for 3D interaction, as supported through the articles given for this assignment, is the rendering and the real-time applications. Rendering 3D interaction platforms currently exceed the computational and storage load of computers and devices available today. While there are 3D interactions that are used today, in VR, they do not meet the standards required for daily computing or use and cannot be used on a mobile device. This is the real limitation of 3D interactions that is being tackled by developers and researchers today.

3.c. The notion that the WIMP is the peak of human computer interaction is an argument that could only be supported by a closed-minded individual. Before WIMPs were used, the common standard was equal to a command line interface. When the next human-computer interaction development is the standard, the WIMP will look just as the command line does now: useful but not desired for daily use. In a paper published by Andy van Dam from Brown University [1], “the longevity and ubiquitousness of the now [four]-decades old WIMP… should not mislead us into thinking that it’s an ideal interface.” Van Dam suggests that the replacement for wimp will be a form of virtual or augmented reality that allows the user to use more natural sensory channels than the implementations of today’s 2D WIMP. While the paper [1] agrees that WIMPs will not disappear, just as the command line hasn’t disappeared, it is noted that the ‘one-size-fits-all design philosophy of WIMP becomes increasingly irrelevant. Van Dam then brings in the persuasive argument of Bill Buxton saying, “specialized devices and interfaces are often much more satisfactory than general purpose ones.”

The replacement for 2D WIMP, in my opinion, will absolutely be Augmented Reality interfacing. While Virtual Reality is a cool concept with many noteworthy applications, the commonly used interface will be Augmented Reality, which allows for people to have one foot in the computing world and one foot in reality – whereas Virtual Reality submerges the user entirely into a computer-rendered world. My position is supported by the excitement for AR by Tim Cook, the CEO of the tech-giant Apple. He’s stated, “I am excited about Augmented Reality because unlike Virtual Reality which closes out the world, Augmented Reality allows individuals to be present in the world but hopefully allows an improvement on what’s happening presently… That has resonance.” Tim Cook has spoken out about his excitement for Augmented Reality over various platforms including Twitter, which leads the consumer to expect the application of AR-technologies into Apple’s product in the future. In fact, at the beginning of this year, Apple filed a patent for a synchronized, interactive augmented reality display for multifunction devices [4]. This display, mainly demonstrated through the example of an interactive augmented reality map, synchronizes live video and an information layer as the perspective of a video camera that, with the use of multiple on-board sensors, can alter the information layer with the live feed video layer with movement so that the there is no latency in user interaction. Apple notes that while many AR systems are not yet realizable, their patent and efforts are making the computationally complex functionalities a widely-available feature.

Innovative developments in the 3D human-computer interaction field have been sweeping the market far and wide as technologies advance towards the ideal 3D interface. Most notably, technologies such as virtual keyboards, hand gesture interaction devices, Microsoft's HoloLens, Google Glass, and even the incredible, futuristic creation of Bionic contact lenses by Google [6]. These devices allow the augmented interaction display to become a reality that allows the user to use a post-WIMP interface without being completely detached from, and unable to interact with, the world around them. The ability to interface the upcoming interaction technologies somewhat seamlessly with the user's everyday life is why I believe that AR will be the next post-WIMP rather than the impressive, but constricting, VR technologies.

The famously-popular app, Pokémon Go, was the fuel that ignited the excitement from the public for the AR industry. In the dissection of why users are so drawn to the award-wining game [3], the authors state that the use of AR allows users to alter their physical realities, which seems to have more of an excitement and interest from the general public compared to the immersion into a virtual reality. In the study, users stated that a highly valued aspect of AR is the ability to be outdoors within their own world. This study supports my position in the post-WIMP debate with the reference of the popularity of Pokémon Go compared to VR games. While the AR interaction implementation is not complete, users from the study also mentioned that the privacy concerns that are out of the hands of IT professionals do not bother them or hinder their willingness to use the technology. The paper states, "much of the design is focused on the game space and everyday space blending into each other in a hybrid reality that takes cues from both, but is restricted by neither." This aspect allows the user to experience the computer-generated reality of 'the future' without losing their grounding to the physical world. The paper supports that the AR industry is a key point-of-interest for consumers. The example of the successful app, Pokémon Go, proves that users are ready and excited to use AR technologies ("with no usability problems", which, of course, is one of the main concerns in the field).

My final point to this argument for AR as the post-WIMP solution goes back to my first in-class Minute Paper on what I predicted to be the WIMP replacement: AR-Holography. The research paper published by Andrew Maimone et al. [2] explains how holography is not what most people think of it to be. Holography is generally thought of by the general public as being noisy, low-contrast, bulky, expensive, and "futuristic". However, as supported by the examples and developments made by the paper, the application of a digital holography device has been realized for years now and is currently being perfected. Microsoft's HoloLens was the first mass-market instance of a holographic AR device. While it was revolutionary, there were issues with resolution, tracking, and latency. The research and developments done in [2] show that computing and processing today has reached the state that allows the wearable AR device to be efficient, smooth, and have no resolution trade-off.

The AR industry as a whole may still be in its infancy and the WIMP is still the infamous human-computer interface, but the developments and technologies in AR are rapidly becoming more functional, user-friendly, and popular. It is realized that computational displays, such as AR and Holography, are an attractive solution that push complexity to software where it is easier to fulfill the demands of the public and the expectations of the future [2]. Soon enough using WIMP will seem like a step back in productivity and everyone will be interacting, not on a 2D flat screen, but on a 3D AR display; one foot on the ground and one foot in the cloud.

References

[1] Van Dam, A., "Beyond WIMP," *IEEE Computer Graphics and Applications*, January/February 2000. [Online]. Available at: https://ieeexplore.ieee.org/document/814559. [Accessed: 2 Oct. 2019].

[2] Maimone, A., Georgiou, A., and Kollin, J., "Holographic Near-Eye Displays for Virtual and Augmented Reality," *ACM Transactions on Graphics*, Vol. 36, No. 4, Article 85, July 2017. [Online]. Available at: https://www.microsoft.com/en-us/research/wp-content/uploads/2017/05/holo_author.pdf. [Accessed: 2 Oct. 2019].

[3] Hamari, J., Malik, A., Koski, J., and Johri, A., "Uses and Gratifications of Pokémon Go: Why do people play mobile location-based augmented reality games?" *International Journal of Human-Computer Interaction*, Vol. 35, No. 9, July 2018. [Online]. Available at: https://www.tandfonline.com/doi/full/10.1080/10447318.2018.1497115#aHR0cHM6Ly93d3cudGFuZGZvbmxpbmUuY29tL2RvaS9wZGYvMTAuMTA4MC8xMDQ0NzMxOC4yMDE4LjE0OTcxMTU/bmVlZEFjY2Vzcz10cnVlQEBAMA==. [Accessed: 2 Oct. 2019].

[4] Bilbrey et al., "Synchronized, Interactive Augmented Reality Displays for Multifunction Devices," *US Patent: Apple Inc.*, Patent: US 10 , 176 , 637 B2. January 2019. Available at: https://patentimages.storage.googleapis.com/df/bb/aa/9b499465743402/US10176637.pdf. [Accessed: 2 Oct. 2019].

[5] Siewert, S., "Cloud Scaling, Part 3: Explore video analytics in the cloud," *IBM developerWorks*, June 2013. [Online]. [Accessed: 30 Sep. 2019].

[6] Hasan, M. and Yu, H., "Innovative Developments in HCI and Future Trends," *International Journal of Automation and Computing*, Vol. 14, No. 1, December 2016. [Online]. Available at: https://www.researchgate.net/publication/312284285_Innovative_developments_in_HCI_and_future_trends. [Accessed: 3 Oct. 2019].