Virtual Reality So Far!

CSC 561: Multimedia Systems

Omowunmi Olalude

V00912070

December 08, 2019

1 Introduction

1.1 What is virtual reality?

The idea of virtual reality is to replace our reality with some new virtual, computer generated environment. Unlike conventional user interfaces, the user is immersed in this new reality. The users are immersed and able to interact with 3D worlds rather than viewing a screen in front of them. By simulating as many senses as possible, such as vision, hearing, touch, even smell, the computer is transformed into a gatekeeper to this artificial world. The only limits to near-real VR experiences are the availability of content and cheap computing power.

1.2 Project Overview

This project intends to take a deep dive into the evolution of virtual reality as we know it and understanding the different compression techniques for 3D images and video. This project will also evaluate how these compression techniques fare in different situations.

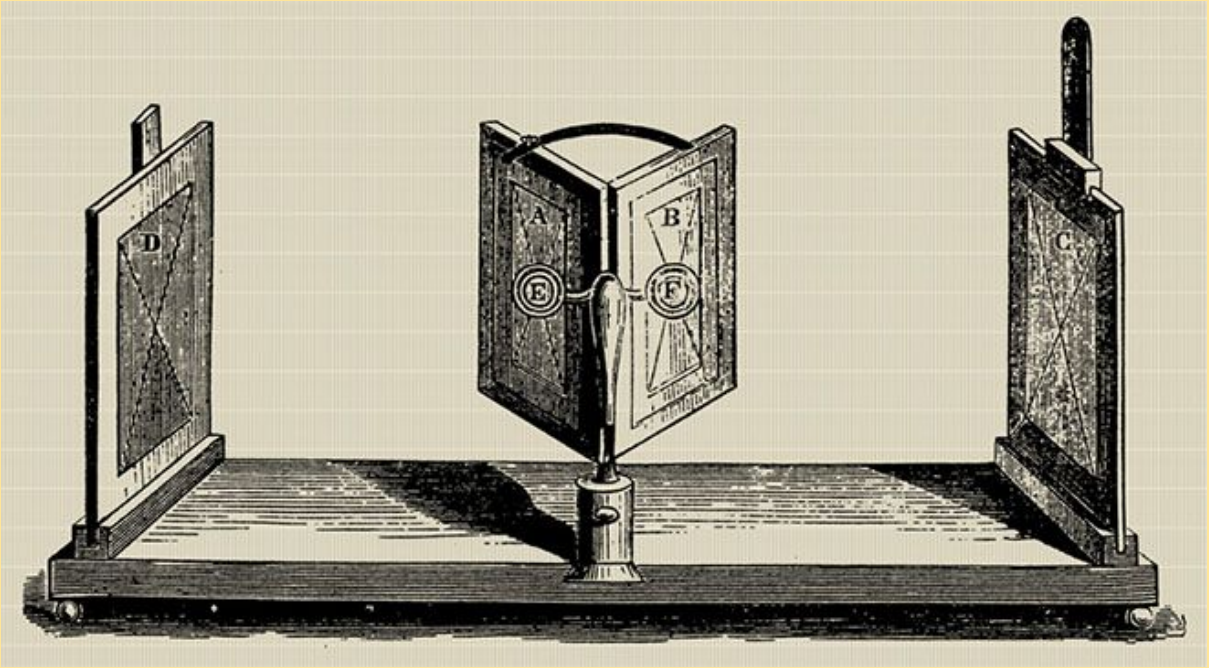
2 History of VR

We can trace attempts at virtual reality back to the 360-degree murals (or panoramic paintings) from the nineteenth century. The purpose of these paintings were to fill the viewer's entire field of vision, making them feel present at some historical event or scene.

However, the exact origins of virtual reality are disputed, partly because of how difficult it has been to formulate a definition for the concept of an alternative existence[1]

2.1 In 1938

Sir Charles Wheatstone was the first to describe stereopsis in 1838 and was awarded the Royal Medal of the Royal Society in 1840 for his explanation of binocular vision, a research which led him to construct the stereoscope.[2]
The research showed that the brain combines two photographs (one eye viewing each) of the same object taken from different points to make the image appear to have a sense of depth and immersion (3-dimensional).
This technology allowed Wheatstone to create the earliest type of stereoscope. It used a pair of mirrors at 45 degree angles to the user's eyes, each reflecting a picture located off to the side.

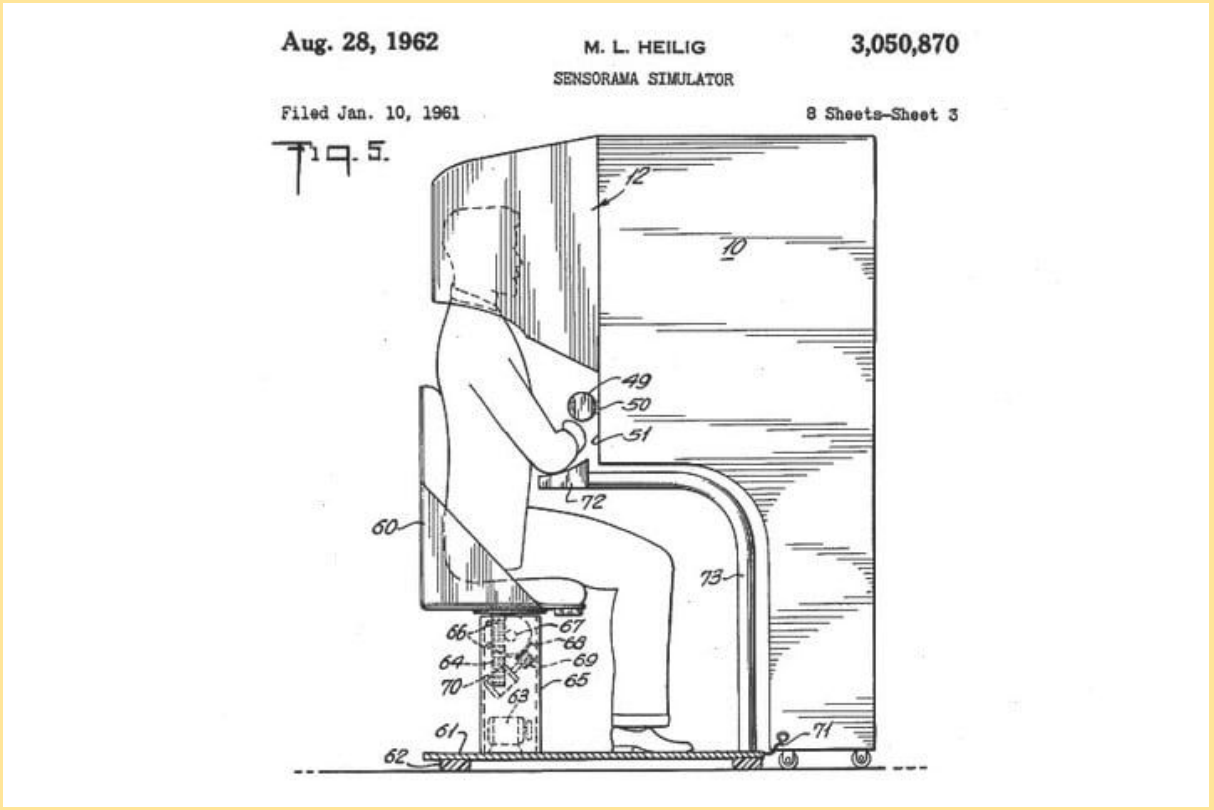


The Wheatstone mirror stereoscope.

2.2 1956

The filmmaker Morton Heilig created Sensorama, the first Virtual Reality machine (patented in 1962). It was a large booth that could fit up to four people at a time. It combined multiple technologies to stimulate all of the senses: there was a combined full colour 3D video, audio, vibrations, smell and atmospheric effects, such as wind.
This was done using scent producers, a vibrating chair, stereo speakers and a stereoscopic 3D screen. Heilig thought that the Sensorama was the "cinema of the future" and he wanted to fully immerse people in their films. Six short films were developed for it.

The Sensorama



Heilig's sensorama

A few years later, in 1960, he honed a version of this idea into a patent for the world's first head-mounted display, promising stereoscopic 3D images, wide vision, and true stereo sound. Heilig patented the Telesphere Mask (1960) which was the first head-mounted display (HMD). This provided stereoscopic 3D images with wide vision and stereo sound. There was no motion tracking in the headset at this point.

Neither technology ever materialized in his lifetime, but they both helped lay the groundwork for the VR revolution to come.

Telesphere Mask

2.3 The Sword of Damocles (1968)

In 1968, a computer scientist Ivan Sutherland created another HMD (Head Mounted Display). Sutherland was one of the most important figures in the history of computer graphics, having developed the revolutionary "Sketchpad" software that paves the way for tools like Computer-Aided Design (CAD).

Sutherland, with his student Bob Sproull, created the first virtual reality HMD, named The Sword of Damocles. This head-mount connected to a computer rather than a camera and was quite primitive as it could only show simple virtual wire-frame shapes.

These 3D models changed perspective when the user moved their head due to the tracking system. It was never developed beyond a lab project because it was too heavy for users to comfortably wear; they had to be strapped in because it was suspended from the ceiling.(Hence the "Sword of Damocles" name, a reference to the Roman myth about a sword which hangs above a person's head to teach them responsibility.)

The Sword of Damocles[3]

2.4 1991

In order to address the increasing popularity of virtual reality among the everyday people, the British company W Industries developed a community space VR gaming system called Virtuality in 1991. The multiplayer setup consisted of real time stereoscopic renderings viewed through a head mounted display with handheld joysticks.[3]

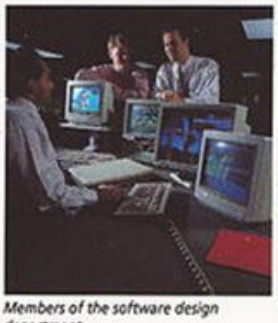
VIRTUALITY

Head 4 Head Leisure System used in larger arcades, bowling alleys and theme parks around the world

Legend Quest, a dungeons and dragons game, winner of Cyberedge Journal's Virtual Reality software of the year award 1993

Example of Virtual Reality centres now in over 20 countries

Virtual Reality in a restaurant in Covent Garden UK



Members of the software design department

Touch glove which allows users to feel Virtual Reality objects

2.5 In 1993

Sega launched its VR headset dubbed as Sega VR in 1993 advertised as the future of gaming with the added dimension of immersivity. The device incorporated cutting edge technologies like LCD screens for vision, head movement tracking, and stereo headphones to provide an immersive gaming experience. However, the product failed to gain popularity among the users because of the ergonomic issues.

Sega VR googles

2.6 Oculus (2010)

In 2010, 18-year-old entrepreneur Palmer Luckey created the first prototype of the Oculus Rift. Boasting a 90-degree field of view that hadn't been seen previously in a consumer and relied on a computer's processing power to deliver the images. This new development boosted and refreshed interest in VR. It raised $2.4 million on Kickstarter a couple years later, before the company was purchased by Facebook for $2 billion in 2014.

Luckey's decision to sell the company before shipping any prototypes to Kickstarter backers stirred up controversy from early supporters.

Oculus Rift

2.7 Google Cardboard

Google Cardboard is a platform developed by Google engineers David Coz and Damien Henry as a low cost solution for experiencing virtual reality content[4]. The basic housing of the device is a cardboard cutout which incorporates a couple of lenses and a slot for attaching a mobile phone. The lens components enables stereoscopic view of the visual content displayed on the phone. The users were encouraged to make there on cardboard viewers with the instructions provided by the developer. The much grounded innovation from Google which was launched in 2014, paved the way for the development of similar head mounted displays which made use of a mobile phone to render images.

The google cardboard

2.7.1 The technology behind the google cardboard

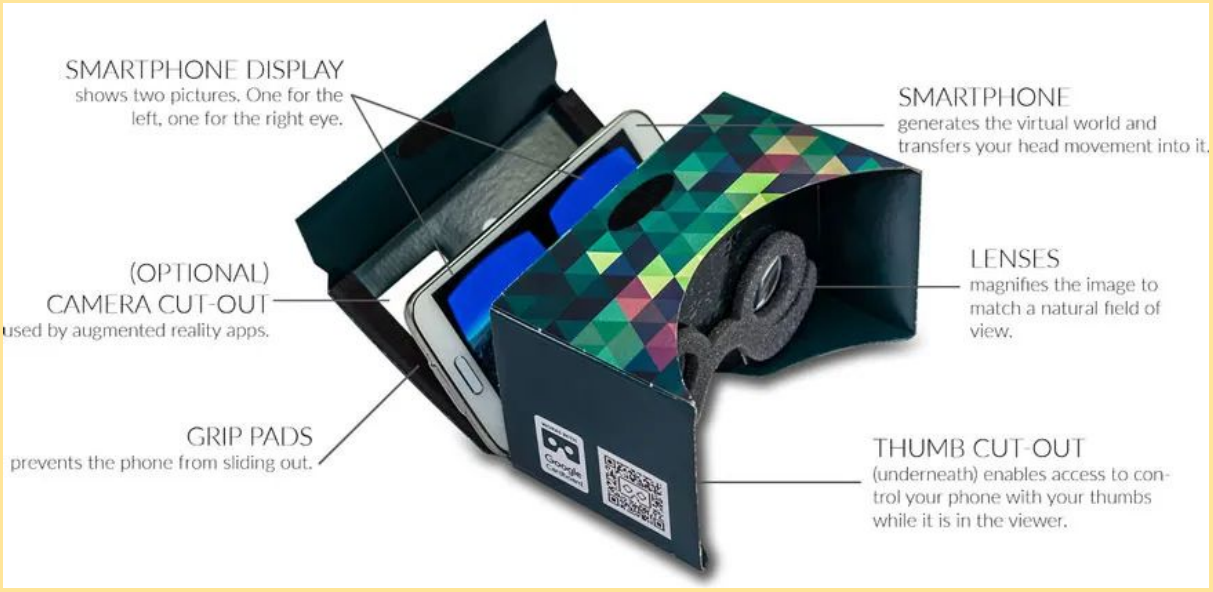
Google Cardboard is a VR construct that allows users to block out external visual stimuli and focus on their device's screen[4]. Once you have it, you fold it into shape, pop your mobile device into it, cue up the VR-friendly app of your choice and hold it up to your head to start your VR experience.

Originally made famous by Google and first presented at the 2014 I/O Developer Conference, Google Cardboard is synonymous with virtual reality glasses using a smartphone[5]. Cardboard ensures cohesion, comfort and light sealing, and is fully compatible with Google Android & Apple iOS, and is suitable for people who wear eyeglasses.

You use your own smartphone as a display and sensor. You insert or push it into the Cardboard, and it generates the virtual reality. The sensors (gyroscope) simultaneously detect head movements and translate them into the virtual world.

If you look up, the image follows upward, and if you look to the left, the image follows to the left, etc. This creates the illusion of being in the middle of things (immersion). Google Cardboard is also called 360° glasses because you can turn in all directions.

You simply hold Cardboard in front of your eyes. The lenses act like a magnifying glass, magnifying the image so that almost the entire field of view (field of view) is filled out. A slightly mutually offset image is shown on the display for both eyes, which conveys the spatial impression (3D).The principle is always the same, and thus many versions of the basic Google Cardboard idea have been developed since 2014.



Google cardboard

2.8 Virtual reality prospects

The global virtual reality market was valued at USD 3.13 billion in 2017 and is expected to reach USD 49.7 billion by 2023, at a CAGR of 58.54% over the period 2018-2023. Virtual reality blurs the line between digital and physical worlds, thereby generating a sense of being present in the virtual environment for the consumer.

Several multinational corporations such as Sony and HTC are venturing into this market space. The launch of commercial virtual reality headsets is expected to accelerate the growth of the market. Technological advancements in VR are expected to generate a plethora of VR solutions with diverse capabilities, which allow consumers to experience utmost immersion[6].

The boundaries which distinguish virtual reality from related technologies like augmented reality are diminishing, leading to the emphasis on a broader canvas of mixed reality technology[7]. Leading tech-companies like Microsoft, Google and Apple are developing augmented reality systems and software applications which are readily available for public consumption. Microsoft developed Hololens, a head mounted device which can augment 3D content on our surrounding real world view. Startup ventures like Magic Leap and Avegant are currently working on devices which uses light field technology for experiencing mixed reality applications.

3 The process of creating a Virtual reality environment

3.1 How Does Virtual Reality Work?

In order for the human brain to accept an artificial, virtual environment as real, it has to not only look real, but also feel real. Looking real can be achieved by wearing a head-mounted display (HMD) that displays a recreated life size, 3D virtual environment without the boundaries usually seen on TV or a computer screen. Feeling real can be achieved through handheld input devices such as motion trackers that base interactivity on the user's movements. By stimulating many of the same senses one would

use to navigate in the real world, virtual reality environments are feeling increasingly more like the natural world.

3.2 Key Components in a Virtual Reality System

1. PC (Personal Computer)/Console/Smartphone

Virtual Reality content, which is what users view inside of a virtual reality headset, is equally as important as the headset itself[8]. In order to power these interactive three-dimensional environments, significant computing power is required. This is where PC (Personal Computer), consoles, and smartphones come in. They act as an engine to power the content being produced.

2. Head-Mounted Display

A head-mounted display (also called HMD, Headset, or Goggles) is a type of device that contains a display mounted in front of a user's eyes. This display usually covers the user's full field of view and displays virtual reality content. Some virtual reality head mounted displays utilize smartphone displays, including the Google Cardboard and Samsung Gear VR. Head-mounted displays are often also accompanied with a headset to provide for audio stimulation.

3. Input Devices

Input devices are one of the two categories of components that provide users with a sense of immersion (i.e. convincing the human brain to accept an artificial environment as real). They provide users with a more natural way to navigate and interact within a virtual reality environment.

Some of the more common forms of virtual reality input devices include: Joysticks, Force Balls/Tracking Balls, Controller Wands, Data Gloves, Trackpads and so on.

3.3 VR Video Formats Explained

Not all VR video experiences are created alike. Here are some of the VR video formats out there.

3.3.1 Monoscopic 360 Video

Mono 360 video was the first and is the most prominently used format for immersive video today[9]. A mono 360 video is usually a 2:1 aspect ratio equirectangular video container, common resolutions include 3840x1920, 4096x2048, 5760x2880, and 7680x3840.

Monoscopic 360 image

Mono means the video is a single channel, but it does actually display to both eyes in the VR headset. A stereo 360 video has 2 channels, with slightly different perspective to give you the perception of depth, like a 3D film in the theater.

Although many 360 videos are a minimum of 4K, content can often still look very blurry. When viewing 360 video content, the viewer is only seeing a small slice of the 360 footage at a given time within their field of view. This means that a 3840x1920 360 video is actually only displaying about 1280x720 in the viewing portal at a given time. This is why VR video content sometimes looks like television from the 1990s. For this reason, every pixel counts!

3.3.2 Stereoscopic 3D 360 Video

A stereo 3D 360 video contains 2 video channels within the same video container, one for each eye. Each view has a slightly different perspective, giving the viewer a sense of depth of field, and separating objects from foreground to background.

Stereo formats for 360 video platforms are typically side by side, or top/bottom, containing that identical video content from a slightly different perspective within the same video file. Specialized cameras are used to capture both perspectives at the same time.

Because both channels are stored in the same video container, this means that essentially your resolution is cut in half to the end viewer. To compensate for this, stereo 360 videos should be delivered at double the resolution of mono, which can be challenging for most streaming platforms and hardware. Common stereo resolutions may include 3840x3840, 5120x5120 and even 7680x7680. For lower end hardware, it's typically delivered in 3840x2160 resolution and both "stretched" stereo channels are crammed into this container, but at this resolution we lose a tremendous amount of detail.

In this example frame, we see 2 independent stacked videos per each eye in the video container, for a 3D 360 video at 4096x4096 resolution.

But how does anyone play a 7680x7680 video file on any device? Thankfully, companies like Pixvana and Visbit have developed what we call viewport biased 360 players. These platforms are specifically designed to display only the pixels in your field of view at the highest resolution, while conserving bandwidth and processing power by showing the pixels you aren't seeing in lower resolution. When you move your head, the resolution will seamlessly adapt.

3.3.3 VR180 or 180 3D Video

Very recently, popular video platforms like YouTube and Facebook have begun to support VR video content in 180 degrees. VR180 is a video file containing 2 channels of video for the left and right eye, but for only the front facing 180 degree field of view.

3.4 Production, Encoding, Transmission and Decoding of 3D images

3.4.1 Production

3D movies/images trick your brain by bringing images projected onto a flat cinema screen to life in full three dimensional glory. As human beings, we have incredible depth perception. Because our eyes are slightly set apart, each eye has a slightly different perspective of what we're looking at. Therefore, our retinas (layer of our eyes upon which light is received) form two different 2-dimensional images, which are instantly pieced together by our brain to form a 3-dimensional picture of the world around us. This is known as stereopsis or stereoscopic vision[9].To create a similar effect, 3D films are captured using two lenses placed side by side, just like your eyes (or by producing computer generated images to replicate the same effect).

In old fashioned 3D films, footage for the left eye would be filmed using a red lens filter, producing a red image, and footage for the right eye would be shot using a blue filter, resulting in a blue image. Two projectors then superimposed the images on the cinema screen.

3D glasses with blue and red filters ensured viewers' left and right eyes saw the correct image: the red filter would only let red light through to your left eye, and the blue filter would only let blue light through to your right eye. Your brain would then combine these two slightly different images to create the illusion of 3D. Unfortunately, this meant that old fashioned 3D films couldn't make full use of colour.

To get around this problem, modern 3D films use polarised light instead of red and blue light.

4 Production of 3D images

4.1 Production

3D movies/images trick your brain by bringing images projected onto a flat cinema screen to life in full three dimensional glory. As human beings, we have incredible depth perception. Because our eyes

are slightly set apart, each eye has a slightly different perspective of what we're looking at. Therefore, our retinas (layer of our eyes upon which light is received) form two different 2-dimensional images, which are instantly pieced together by our brain to form a 3-dimensional picture of the world around us. This is known as stereopsis or stereoscopic vision[9].To create a similar effect, 3D films are captured using two lenses placed side by side, just like your eyes (or by producing computer generated images to replicate the same effect).

In old fashioned 3D films, footage for the left eye would be filmed using a red lens filter, producing a red image, and footage for the right eye would be shot using a blue filter, resulting in a blue image. Two projectors then superimposed the images on the cinema screen.

3D glasses with blue and red filters ensured viewers' left and right eyes saw the correct image: the red filter would only let red light through to your left eye, and the blue filter would only let blue light through to your right eye. Your brain would then combine these two slightly different images to create the illusion of 3D. Unfortunately, this meant that old fashioned 3D films couldn't make full use of colour.

To get around this problem, modern 3D films use polarised light instead of red and blue light.

4.2 Compression

4.2.1 Multiview Video Coding (MVC)

Multiview Video Coding (MVC, also known as MVC 3D) is a stereoscopic video coding standard for video compression that allows for the efficient encoding of video sequences captured simultaneously from multiple camera angles in a single video stream.[10] The basic approach of most MVC schemes is to exploit not only the redundancies that exist temporally between the frames within a given view, but also the similarities between frames of neighboring views. By doing so, a reduction in bit rate relative to independent coding of the views can be achieved without sacrificing the reconstructed video quality.

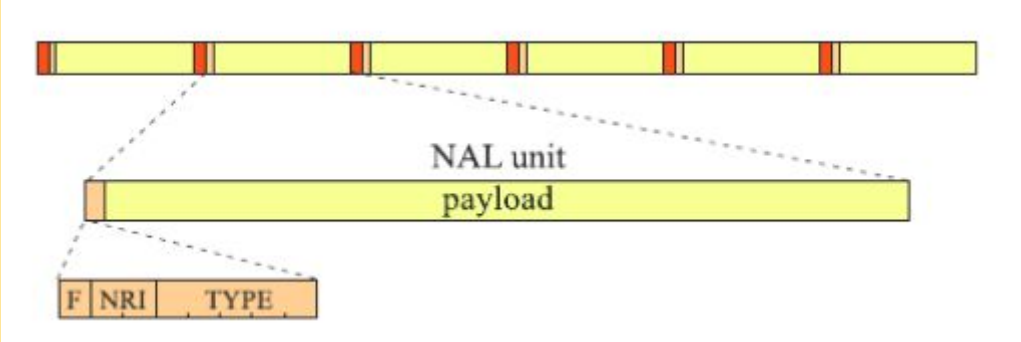
4.2.1.1 Extensions of the H.264/MPEG-4 AVC Standard

The MPEG−4 AVC/H.264 standard is currently the most important one in the area of video coding [10]. It is based on the hybrid motion compensation and transforms coding algorithm like many of its predecessors. Significant enhancements of the classic algorithm have been implemented in this standard to improve its coding efficiency. The H.264/AVC encoder is divided into two layers: the video coding layer (VCL) and the network abstraction layer(NAL). The VCL processes video data: each frame of the input sequence is partitioned into a set of macroblocks, each macroblock is temporally or spatially predicted and its prediction error is then transform coded. The VCL generates a stream of encoded macroblocks organized into slices. The slice covers a part (or entire) frame and can be parsed independently from other slices. The NAL formats the outputstream of the encoder as a series of packets called NAL units.The set of consecutive NAL units decodable into a singleframe is called an access unit (AU). Each NAL unit is composed of one byte header followed by its payload (see below).The header contains three fields describing the payload:

- F – error flag (1 bit), NAL units with this field set to 1should not be processed;

- NRI – NAL unit priority (2 bits), this field should be set to 0 in all NAL units not used as a reference by otherNAL units. The higher value of this field the more important the NAL unit is for the video sequence recon−struction;
- TYPE – NAL unit type, values 1 ÷ 23 are restricted to be used only within the H.264/AVC standard. Values 0 and 24 ÷ 31 may be used for other purposes, e.g., in trans−mission.The length of a NAL unit is not encoded in its header

Therefore, NAL units must be prefixed with start codes(e.g., defined in the Annex B of the H.264/AVC standard)or encapsulated in additional data structures (e.g., transmission protocol packets) to allow their separation.



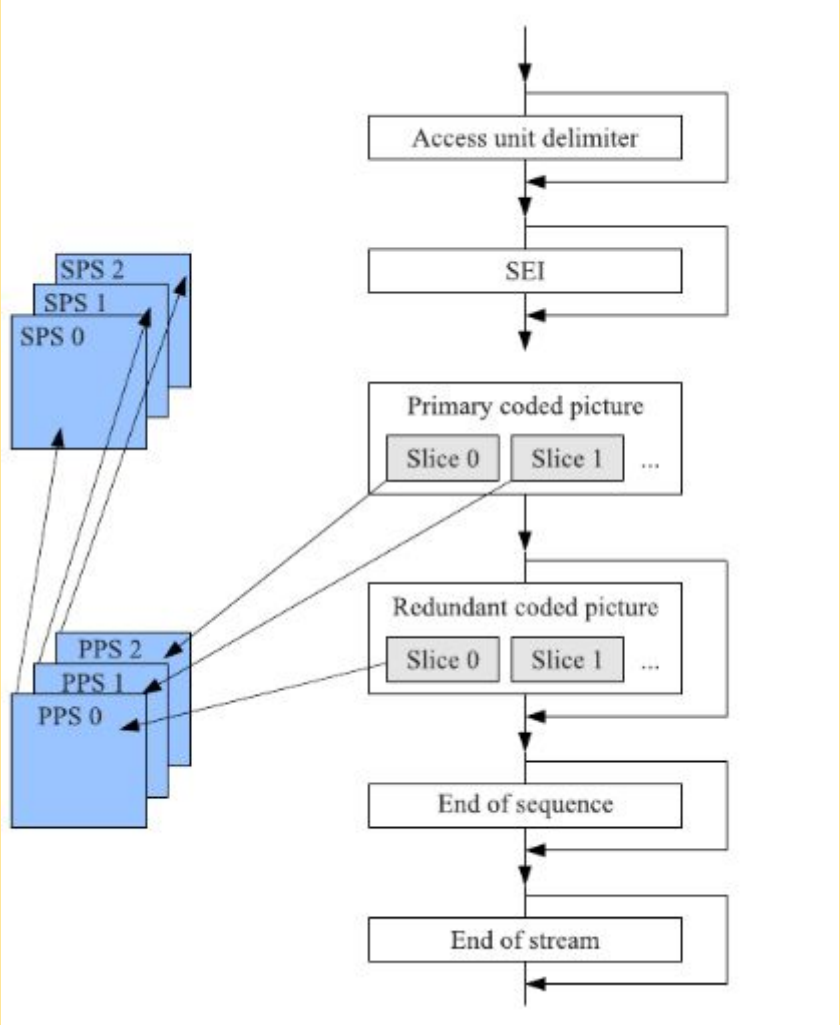
NAL units in the H.264/AVC bitstream.

There are two main NAL unit categories:

1. VCL containing encoded video data, the following types have been defined:
    - coded slice of instantaneous data refresh (IDR) pictures (TYPE = 5). IDR picture allows to start the decoding process. The first frame in the video sequence is always encoded in an IDR mode. IDR pictures may be repeated in the video sequence to allow stream switching or recovery from transmission errors;
    - coded slice of non−IDR picture (TYPE = 1);
    - coded slice data partition (TYPE = 2, 3, 4). Data partitioning is an error resilience tool available in the H.264/AVC standard. More detailed description can be found in Ref. 12;
2. Non−VCL carrying associated additional information,the most important types are:
    - sequence parameter set (SPS, TYPE = 7), contains in frequently changing information relating to the entire video sequence;
    - picture parameter set (PPS, TYPE = 8), contains in frequently changing information relating to one or more pictures in a video sequence;
    - supplemental enhancement information (SEI,TYPE = 6), provides information supporting the decoding process or displaying the reconstructed video sequence;
    - access unit delimiter (TYPE = 9);–end of sequence (TYPE = 10);
    - end of stream (TYPE = 11).

    The syntax of the H.264/AVC bitstream is more flexible with respect to the previous video coding standards. The only required syntax element in every access unit is theVCL NAL unit containing at least one slice of the primary coded picture (see figure below). In certain profiles of the H.264/AVC standard primary coded data may be followed by VCL NAL units with a redundant representation of this picture.

Each slice header directly refers to the PPS and indi−rectly to the SPS, both parameters' sets must be known to the decoder to allow slice processing. Usually, non−VCL NAL units with SPS and PPS are transmitted before anyVCL NAL unit in the same channel (”in−band”). However,SPS/PPS NAL units may be transmitted in an additional,more reliable channels as well (”out−of−band”). SEI mes−sages may be very useful in the VCL NAL unit processing,but they are not necessary to decode the access unit. Similarly access units delimiters are not required to detect the beginning of a new frame in the encoded video sequence.The frame boundaries can be derived from slice headers inVCL NAL units, however it is a resource consuming pro−cess. The H.264/AVC bitstream processing may be significantly simplified if it contains access unit delimiters. Multi−ple SPSs and PPSs may be defined and used by an encoder in the same bitstream. It is only required that the VCL NALunits using different SPS should be preceded by the end of a sequence NAL unit. The end of a stream NAL unit may follow the very last access unit in the entire sequence



Access unit representation in the H.264/AVC bitstream.[13]

4.2.1.2 High efficiency video coding

The MPEG−4 AVC/H.264 video coding standard is commonly used in many multimedia applications. However, the user requirements are still growing and more effective compression tools are demanded. The standardization process for the new video coding standard has been initiated. The standard,called

high efficiency video coding (HEVC) is expected to provide significantly higher compression effectiveness with respect to the MPEG−4 AVC/H.264. It is especially important for the becoming more popular HD and emerging Ultra−HD(4k × 2k and more) video applications (e.g., digital cinema).

The proposals submitted are based on the traditional hybrid video coding algorithm. Several new compression tools have been proposed to fulfil the requirements for the higher efficiency :

- variable size coding units: frame partitioning into a fixed size macroblock structure used in the previous standards has been replaced by the partitioning into a variable size(from 8 × 8 up to 64 × 64) structure of coding units(CU). The CU structure adapts to the picture characteristics – smaller CU are used in regions with many details,larger CU in uniform regions;
- quad−tree partitioning of coding units into variable size prediction units (PU) and transform units (TU). The prediction mode is selected at the PU level, spatial trans−form is applied to the TU. The structure of PU may be independent from the structure of TU;
- adaptive loop filters: used in addition to the deblocking filter to improve image quality;
- adaptive interpolation filter for the samples in the sub−pixel positions to improve the quality of the predictions signal;
- entropy coding with the use of probability interval parti−tioning: unit interval is divided into small number of probability intervals to decouple the probability modelling form entropy coding.

The HEVC standard is still being developed. It has reached the Committee Draft level, the standardization process is expected to be completed in the beginning of 2013. It is worthwhile to mention, that similarly as for the MPEG−4AVC/H.264, the scalable and multiview extensions for the HEVC have been proposed

4.2.1.3 3D video

Multiview 3D scene representation with fixed number of views allows for the only limited implementation of the fee viewpoint television concept. On the contrary, multiview video plus depth (MVD) representation allows for the synthesis of any view within the certain range providing much better viewing experience. Although MVD representation may be encoded with the use of existing tools – depth maps may be handled as an unrelated monochrome image –better results may be achieved if views and depth maps are encoded jointly. The MPEG initiated a standardization process for the 3D video (3DV) based on the MVD representation in the beginning of 2011. The goal is to create an effective compression tools allowing for a synthesis of high quality views. Another requirement is compatibility with the existing coding standards, as well as an ability to extract data required for the presentation on traditional mono or stereo displays. The depth map extraction techniques are beyond the scope of the call, the supplied test material for the proposal evaluation contains both views and depth maps.The proposals submitted in the response for the call fall into two categories: ACV−based and HEVC−based. It has been decided that 3DV standard will be developed in two parallel tracks, one for each of the above categories.The following coding tools have been selected for theAVC−based test model:

- inter−view prediction based on view−synthesised picture motion vector prediction for texture coding with the use of associated depth information;

- joint inter−view depth filtering to denoise depth data−depth range−based prediction to compensate inconsis−tencies of depth maps generated for different views re−use of texture motion information in depth coding;
- slice header prediction to remove redundancy between texture slice headers and depth slice headers.The HEVC−based test model is based on the following coding tools:
- dependent views encoding with the use of disparity compensated prediction
- inter−view prediction with the use of a view−synthesized picture;
- adaptive filter to remove artefacts created by the view synthesis process;
- inter−view motion prediction with the use of depth information;
- inter−view residual prediction based on the depth information;
- bit assignment (QP adjustment) for the texture component based on the depth information to improve perceptual quality;
- non−linear depth map representation to improve the quality of the synthesised views;
- modified motion information coding to preserve sharp edges in depth maps;
- depth maps intra prediction modes providing better representation of edges;
- motion parameter inheritance – reuse of texture motion information in depth coding.

5 What can be improved

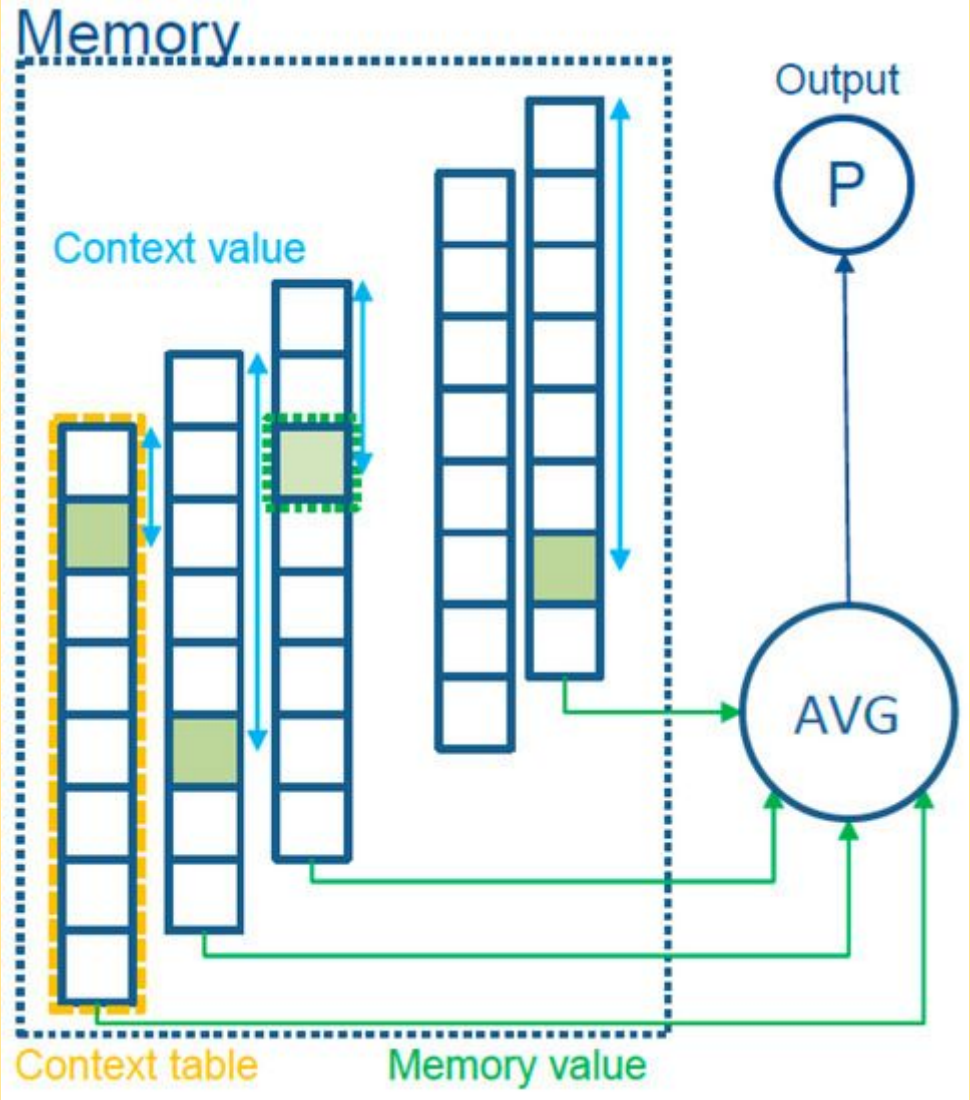
I studied a paper that attempted and succeeded in providing a lossless compression algorithm for these images.[14] I found it very interesting and worth delving into because not all 3D images or videos will allow for lossy compression. Images like medical images are less tolerant to losses and this fueled my interest in lossless compression of 3D images.

In this paper they gave a detailed description of the state-of-the-art lossless compression software PAQ8PX applied to grayscale image compression. They proposed a new online learning algorithm for predicting the probability of bits from a stream. The idea behind the algorithm is to encode probabilities in a memory-like structure. The probabilities are accessed by using a set of keys computed on a known context. Resilience to noise (since lossless compression for photographic images will mostly have to deal with noise) would be handled by allowing that not all the keys will find a match in the memory.

5.1 Lossless Image Compression with Contextual Memory

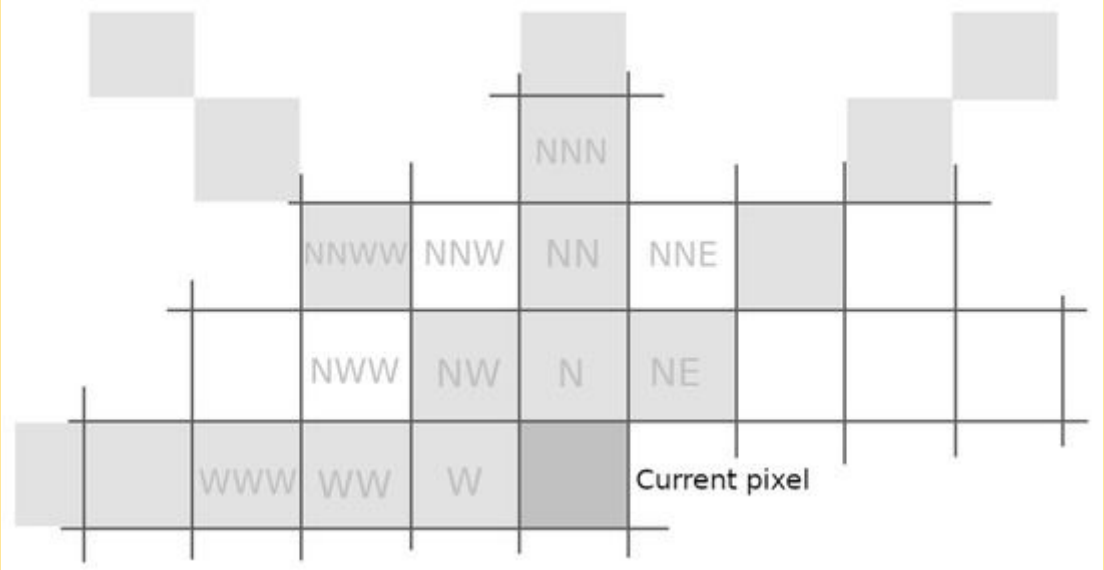
Context Modeling

Lets start by defining the main terms used below: The word context represents the region of the image that participates in the prediction mechanism. “Context value” is the numeric value of the context, either a direct value or a function of that value, which will be used as an index in the memory structure. The algorithm makes no assumption about the memory structure, but they provide some implementation details. The output of indexing the memory is the “memory value”.



Block scheme of the proposed method.

They chose a simple model for contexts for predicting the bits of the pixels. They used rays in four directions and with various lengths, and the quantized derivatives along the rays. Since the pixels of the image are predicted from left to right, top to bottom, the only information we can rely on are known pixels, which means the directions are to the west, 45 degrees north-west, north, and 45 degrees north-east. The rays are depicted as gray background in Figure 4. We choose rays of varying lengths from length 1 to 7, but use this as a parameter. The derivative with respect to the intensity value is computed as the difference between the consecutive pixels of the ray and quantizing is done by masking the lower order bits from the derivative. We use three levels of quantizing, each cutting out one more bit than the other. The current pixel participates in the contexts only with the currently known bits.



Contexts as rays.

In order to compute the context value from the contexts, we use a hashing function. We chose Fowler–Noll–Vo hash function (FNV) that is a non-cryptographic one byte at a time, designed to compute fast and with a low collisions rate. It is found to be particularly suited for hashing nearly identical strings.(Reference to this paper is [14])

As an optimization, since we know that we will need to compute hashes for rays, we exploit the fact that FNV computes one byte at a time hashes and pass as input only the longest ray and output all the intermediate results. We apply the same optimization for the quantized derivatives of the rays.

Description of the Contextual Prediction

Model Prediction

To make a prediction, they propose the following algorithm (simplified from the original proposed algorithm, which had a probability refinement phase):

1. We obtain a value from the memory for each context. One way to do that is to index the hash of the "context value" in a table
2. We average all the obtained "memory values"
3. Convert the average into a probability using the sigmoid function

$$p = \sigma\left(\frac{k}{n}\sum_{i=0}^{n} v_i\right),\ v_i = M\left[i\right]\left[hash\left(c_i\right)\right]$$

p is the output probability (that a bit is one),
n is the number of input contexts,
c_i is the *context value* of the i-th context,
v_i is the *memory value* from the memory M for context i,
k is some ad-hoc constant
σ is the sigmoid function.

Interpretation of Values

Logistic regression is a way of combining probabilities when they are fed as inputs to the algorithm. Using stretched probabilities as inputs (applying logit function to them), logistic mixing becomes optimal for minimizing wasted coding space (Kullback–Leibler divergence) because the weighting becomes geometric.

$$\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_0 x_{k,i}$$

where $x_{j,i}$ is a probability, becomes

$$\beta_0 + \beta_1 t_{1,i} + \beta_2 t + \cdots + \beta_0 t_{k,i}$$

where

$$t_{j,i} = \text{logit}(x_{j,i})$$

The update formula for minimizing the relative entropy is:

$$\beta_j = \beta_j + \alpha * t_j * (y_i - p_i)$$

The set of weights carries a part of the predictive part of the ensemble, and they get updated to better represent the potential of individual components. In the case of PAQ8, the components gather statistics independently and the network independently mixes the statistics. Adding more weights to the mixer can result in improved predictive power since the model can better discriminate between the contexts. But what if instead of separating the mixer from the statistics we move the mixing information towards the components? How can we pass the mixing information to the weak learners of the ensemble?

So far, we know that the memory value vi is taken from a memory structure. The index in the memory is computed based on the context value. But the feature is the context, not the memory value. We can assume that the memory value is

vi=βi∗ ti

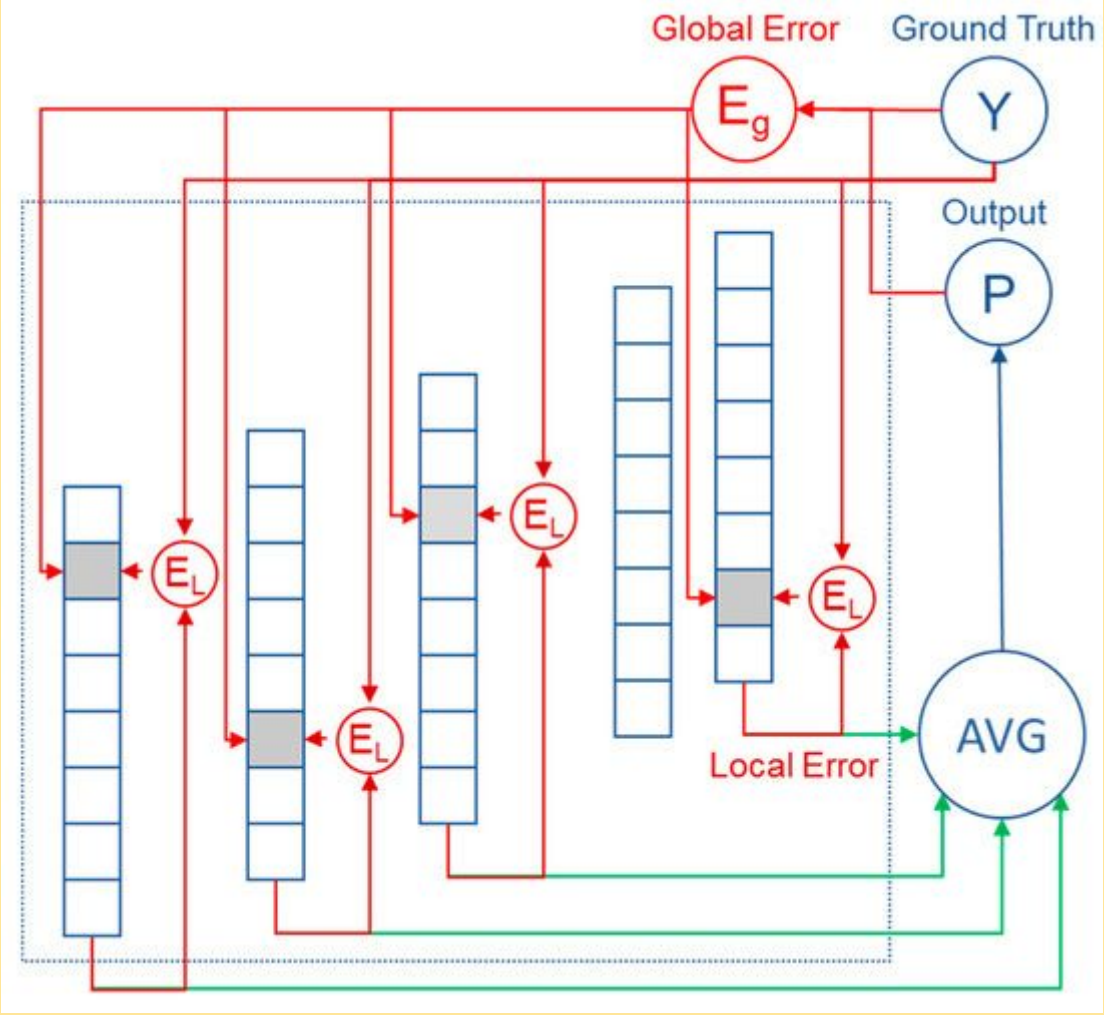
with ti a stretched probability and β the weight of the probability in the ensemble. Computing the output probability resembles logistic regression, with the main difference being that we apply averaging. The average in itself is a weighted stretched probability

$$\frac{k}{n}\sum_{i=0}^{n}(\beta_i * t_i)$$

which is converted into a regular probability by applying the sigmoid function.

Updating the Model

In order to pass mixing information to the weak learners, we propose a dual objective minimization function (as depicted below):



Block scheme for the proposed update algorithm.

- In respect to the output of the network–global error
- In respect to the output of the individual nodes (side predictions)–local error

Like PAQ8, we use reinforcement learning. Since we do not know the true value of the probability that a bit is 0 or 1 in a given context, we cannot use supervised learning. We backpropagate the binary outcome in the network and try to minimize the cumulative logistic loss in an online manner. The square loss can be also used, but we are trying to minimize the wasted coding space.

If we wanted to minimize the square loss, the formula would be

$$E_g = \beta_g (p - y) * p * (1 - p)$$

with E_g as the global error, β_g the global error learning rate, p is the output probability of the entire network, and y the binary ground truth.

The local error is computed for each *memory value* in a similar fashion to the global error.

$$E_l = \beta_l * (p_i - y), \quad p_{i=}\sigma (k_v * v_i)$$

with E_l as the local error, β_l the local error learning rate, p_i is the output probability for the i-th context (side prediction) multiplied by an ad-hoc value k_v, computed as the sigmoid of the *memory value* v_i, and y is the binary ground truth.

All the "memory values" are then updated by subtracting the local and global errors:

$$v_i = v_i - E_l - E_g$$

Instead of updating weights of the mixture, we update directly the values that contribute to the average. We have no layer to separate the context weights from the input probabilities, making the method different from the context mixing algorithm.

Memory Implementation and Variations[14]

The algorithm makes no assumption on how to organize the memory structure. We describe here potential implementation and give more details to the implementation we chose to use. The proposed implementations are based on hash tables since they give fast retrieval, given the fact that the context values are computed by hashing series of pixel values.

6 Conclusion

6.1 Application of Virtual Reality

Virtual Reality provides a way for human computer interaction where the user can involve and interact with the simulated reality in the same way he/she behave with the real world. There are multidisciplinary fields and sectors of applications of VR from Design to Architecture andTourism to Entertainment and so on.

● Medical Applications

VR can be used for the treatment of patients with mental disorders of phobia, as a meditation technique for overcoming stress and anxiety, and so on. The users are allowed to confront their fears in a controlled environment which is otherwise impractical to achieve in a real scenario. Intuitive, alternative therapies based on VR are nowadays used for mental and physical rehabilitation.[15]

● Design and Architectural Visualization

A huge amount of gathered information is associated with any kind of research and study methodologies. The primitive goal of data visualization techniques is to develop some visual representations out of these in order to make them perceptible and easily accessible to the users.[15] VR can be used as an intuitive medium to where the users can assess the information and interact to

the visualizations. The feeling of physical presence and sense of space makes VR effective in experiencing architectural walkthroughs and virtual tourism. Modelling of three dimensional objects in real time is another application of VR systems.

● Training and Simulation

Industries like aviation and mining require effective methodologies to practically train their professionals for getting acquainted with the technicalities involved. The high stakes of risk and capital involved make these mock training exercises inevitable. First such applications of VR were explored in the form of flight simulators which helped the pilots to overcome practical difficulties. VR systems are nowadays used for training of surgeons, mining workers, soldiers etc.

● Tele Operation

Tele Operation is the technology which allows conducting remote operations without being physically present in the actual space. VR can be used to achieve possibilities of tele operation where the user can carry out operations remotely from a controlled space using a remote control system. VR seems to be the only effective solution when the distant environment is hazardous to human life or impractical to work (Remote bomb diffusion, Remote astronomic explorations).

●Entertainment

Entertainment is the domain which effectively brought VR to the masses. The added dimension of immersion achieved through VR games makes the collaboration tailored for the purpose of entertainment. Storytelling in VR is being explored in the form of 360 videos which is now gaining popularity.

●Education

Education is another domain which can make use of VR technology to enhance the effectiveness of teaching and learning. Virtual environments can be a better way of representation of information to the students as it provides the possibility of understanding through interaction. Virtual field trips (museums, historic sites), Simulation based training (Medical students) are some of the methodologies that are being implemented today. VR can also be used for education sector catering to those with special needs.

● Performing Arts

VR redefines the possibilities of performance based artists by adding attributes of interactivity. Wearable accessories are effectively used by the artists to enhance the experience delivered through simulations which interacts them in the desired manner.

7 References

[1]”History Of Virtual Reality” https://www.vrs.org.uk/virtual-reality/history.html

[2]”History of VR - Timeline of Events and Tech Development” https://virtualspeech.com/blog/history-of-vr

[3]http://www.dsource.in/course/virtual-reality-introduction/evolution-vr/sword-damocles-head-mounted-display

[4]”"What is Google Cardboard?': Everything you need to know about Google's low-tech VR experience” https://www.businessinsider.com/what-is-google-cardboard

[5]”WHAT IS POP! CARDBOARD?” https://mrcardboard.eu/what-is-google-cardboard-how-does-google-cardboard-work/

[6]”Virtual Reality Market by Technology Advancements and Future Prospects 2019 to 2023

”https://www.marketwatch.com/press-release/virtual-reality-market-by-technology-advancements-and-future-prospects-2019-to-2023-2019-12-05?mod=mw_quote_news
[7]”Vr and Prospects” http://www.dsource.in/course/virtual-reality-introduction/evolution-vr/vr-and-prospects
[8]”The Ultimate Guide to Understanding Virtual Reality (VR) Technology” https://www.realitytechnologies.com/virtual-reality/
[9]”Science Of 3-D Movies: How Do Images On A Flat Screen Pop Out?” https://www.scienceabc.com/humans/science-of-3-d-movies-how-things-on-a-flat-screen-appear-to-emerge-depth.html
[10]”Multiview Video Coding” https://en.wikipedia.org/wiki/Multiview_Video_Coding
[11]Vetro, A.; Wiegand, T.; Sullivan G.J. “Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard” January 2011
[12]Information technology – coding of audio−visual objects –Part 10: advanced video coding, ISO/IEC 14496−10:2008
[13]A. BUCHOWICZ “Video coding and transmission standards for 3D television – a survey” 2013
[14]A. Dorobanţiu,B. Remus “Improving Lossless Image Compression with Contextual Memory” June 2019
[15]”Applications of VR” http://www.dsource.in/course/virtual-reality-introduction/applications-vr