

IMMERSIVE 3D EXPERIENCE

Comfortably sitting in seat 3F, John is watching one of his favourite operas. This close he can see all the details of the set, costumes, and the movements of the music director as he skilfully conducts the orchestra by careful gestures of his baton. He is immersed in the scene, capturing all the details. Then all of a sudden, the doorbell rings. Annoyed, John has to stop the video to see who it is. This could be the mainstream TV experience of the future.

This scene is called free-viewpoint technology that is part of my research at the University of Malta (UoM). Free-viewpoint television allows the

user to select a view from which to watch the scene projected on a 3D television. The technology will allow the audience to change their viewpoint when they want, to where they want to be. By moving a slider or by a hand gesture, the user can change perspective, which is an experience currently used in games with their synthetically generated content — synthetically generated by a computer game's graphics engine.

Today we are used to seeing a single viewpoint. If there are multiple perspectives we usually don't have any control over them. Free-viewpoint technology will turn this idea on top of its head. The technology is expected to hit the market in the near future, with some

companies and universities already experimenting with content and displays. New autostereoscopic displays do not need glasses (pictured next page), these displays 'automatically' generate a 3D image depending on which angle you view them. A clear example was the promise made by Japan to deliver 3D free-viewpoint coverage of all football games as part of their bid to host the FIFA World Cup in 2022. The bid was unsuccessful, which might delay the technology by a few years.

Locally, my research (and that of my team) deals with the transmission side of the story (pictured). For free-viewpoint to work, a scene needs to be captured using many cameras. The more »



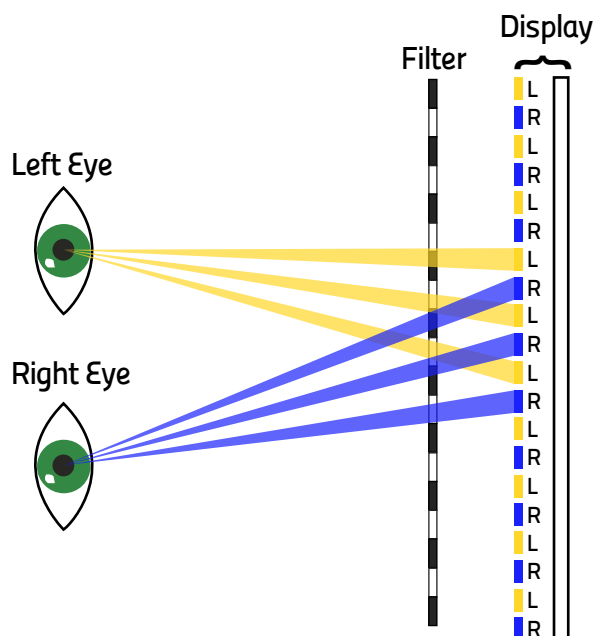
Student plan

Free Calls & SMS to all Vodafone
plus 2Gb Internet on Your Mobile

vodafone.com.mt/student

power to you

The autostereoscopic display technique without glasses. Depending on the position of the viewer, the filter directs the left image to the left eye and the right image to the right eye



cameras there are, the more freedom the user has to select the desired view. So many cameras create a lot of data. All the data captured by the cameras has to be transmitted to a 3D device into people's homes, smartphones, laptops and so on. This transmission needs to pass over a channel, and whether it is fibre cable or wireless, it will always have a limited capacity. Data transmission also costs money. High costs would keep the technology out of our devices for decades.

My job is to make a large amount of data fit in smaller packages. To fit video in a channel we need to compress it. Current transmission of single view video also uses compression to save space on the channel so that more data can be transmitted and save on price. Note that, for example for high definition we have 24 bits per pixel and an image contains 1280 by 720 pixels (720p HD standard), that's nearly 100,000 pixels for every frame. Since video is around 24-30 frames per second the amount of data being transmitted every minute starts escalating to unfeasible amounts.

Free-viewpoint technology would be another big leap in size. Each camera would be sending their own video, which is the same amount of data as we are now getting. If there are ten cameras, you would need to increase channel size by a factor of ten. This makes it highly expensive and unfeasible. For the example above, the network operator needs ten times more space on the network to get the service to your house, making it ten times more expensive than single view. Therefore, research is needed to drastically reduce the amount of data that needs to be transmitted while still keeping high quality images. These advances will make

the technology feasible, cheaper, and available for all.

So the golden question is, how are we going to do that? Research, research, and more research. The first attempts by the video research community to solve this problem were to use its vast knowledge of single view transmission and extend it to the new paradigm. Basic single view algorithms (an algorithm is computer code that can perform a specific function, like Google's search engine) compress video by searching through the picture and finding similarities in space and in time. Then the algorithms send the change, or the error vector, instead of the actual data. The error vector is a measure of imperfections and how it is used by computer scientists to compress data is explained below.

First let us look at the space component. When looking at a picture, it is quite clear that some areas are very similar. The similar areas can be linked and the data grouped together

into one reference point. The reference point has to be transmitted with a mathematical representation (vector) that explains to the computer which areas are similar to each other. This reduces the amount of data that needs to be sent.

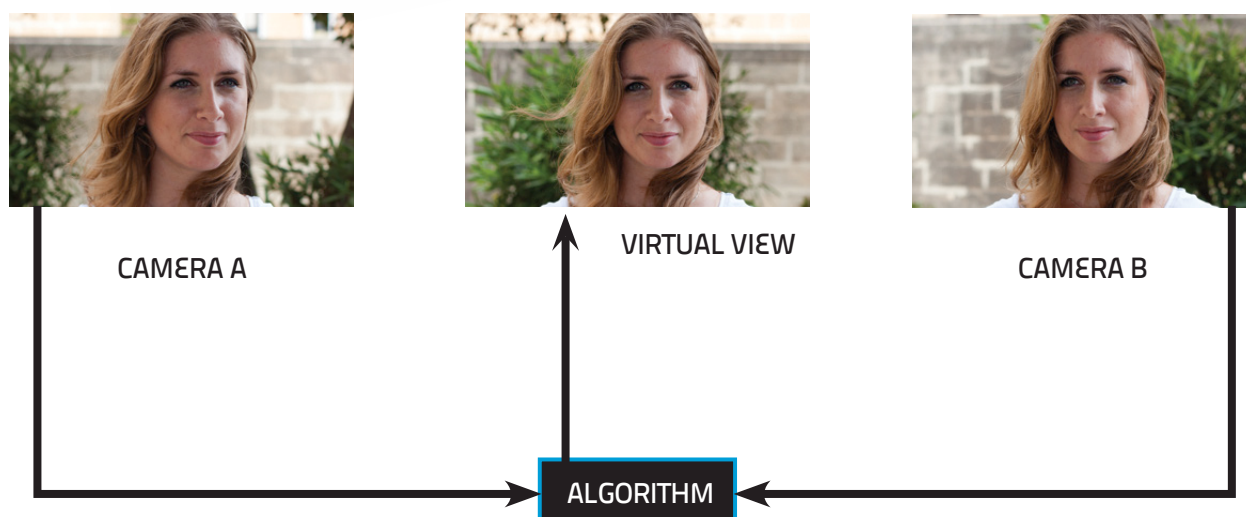
Secondly, let us analyse the time aspect. Video is a set of images placed one after another and run at 25 or 30

frames per second that gives the illusion of movement and action. To make a video flow seamlessly images that are right after each other are very similar. If we have two images the second one will be very similar to the first, with only a small movement of some parts of the image. Like we do for space, a mathematical relationship can be calculated for the similar areas from one image to the next. The first image can be used as a reference point and for the second we transmit only the vector that explains which pixels have moved and by how much. This greatly reduces the data that needs to be transmitted.

The above techniques are used in single view transmission, with free-viewpoint technology we have a new dimension. We also need to include the space between cameras shooting the same scene. Since the scene is the same there is a lot of similarity between the videos of each camera. The main difference is that of angle and the problem that some objects might be visible from one camera and not from another. Keeping this in mind, »

“For free-viewpoint to work, a scene needs to be captured using many cameras”

From the pixel information in the picture captured by Camera A and that of Camera B, we can construct an image in between the two cameras



a mathematical equation can be constructed that explains which parts of the scene are the same and which are new. A single camera's video is used as a reference point while its neighbouring cameras only transmit the 'extra' information. The other camera can compress their content drastically. In this way the current standard can be extended to free-viewpoint TV.

Compressing free-viewpoint transmissions is complex work. Its complexity is a drawback, mobile devices simply aren't fast enough to run computer power intensive algorithms. Our research focuses on reducing the complexity of the algorithms. We modify them so that they are faster to run, need less computing power, and still keep the same quality of video, or with minimal losses.

We have also explored new ways of reconstructing high quality 3D views in minimum time, using graphical processing units (GPUs). GPUs are commonly used by high-end video games. Video must be reconstructed with a speed of at

least 25 pictures per second. This speed must be maintained if we want to build a smooth continuous video in between two real camera positions (picture). A single computer process cannot handle algorithms that can achieve this feat; instead parallel processing (multiple simultaneous computations) is essential. To remove the strain off a main processing unit in a computer processing can be offloaded to a GPU. Algorithms need to be built that use these alternative processing powers. Ours show that we can obtain the necessary speeds to process free-viewpoint 3D video even on mobile devices.

Since free-viewpoint takes up a large bandwidth on networks, we researched

whether these systems can feasibly handle so much data. We considered the use of next generation mobile telephony networks (4G). Naturally they offer more channel space, we wanted to see how many users they can handle at different screen resolutions. We showed that the technology can be used only using a limited number of cameras. The number of users is directly related to the resolution used, with a lower resolution needing less data and allowing more views or users. This research came up with design solutions for the network's architecture and broadcasting techniques needed to minimise delays.

The road ahead is steep and a lot of work is needed to bring this technology to homes. My vision is that in the near future we will be consuming 3D content and free-viewpoint technology in a seamless and immersive way in our homes and mobile devices. So for now sit back and imagine what watching an opera or football match on TV would look like in a few years' time. ●

"The road ahead is steep and a lot of work is needed to bring this technology to homes"

The transmission chain in free-viewpoint technology showing four cameras. The encoder is where our algorithms are being implemented

