

# Projection Methods for 360-degree Video

Yilin Wang

Hong Kong Polytechnic University 11 Yau Choi Road, Hung Hom, Kowloon, Hong Kong, 999077, China

**Abstract:** This paper discusses the methods used for projecting 360-degree video onto a two-dimensional plane to reduce the transmission bandwidth. The two primary methods examined are Equirectangular Projection and Cube Projection. Equirectangular Projection is known for its oversampling problem at high latitudes, leading to uneven pixel distribution. To address this, a region adaptive smoothing technology is introduced to optimize video quality and save bandwidth. Cube Projection, which maps spherical video onto the faces of a cube, is analyzed for its ability to maintain pixel uniformity and reduce geometric distortion. The paper further discusses the solutions for correcting motion compensation errors in Cube Projection, emphasizing the importance of multi-facet and cube face extensions in eliminating geometric distortions.

**Keywords:** 360-degree Video; Equirectangular Projection; Cube Projection; Region Adaptive Smoothing; Motion Compensation; Geometric Distortion; Panoramic Video; Video Coding; Virtual Reality (VR).

## 1. Introduction

Virtual reality (VR) is a kind of computer simulation technology. Users can experience the realistic scenes and sounds brought by VR devices such as headwear devices as if they are in the real world. While immersive image or video is obtained by capturing video within 360° at the same time by panoramic camera, so in the process of watching, the viewer can see all-round video and select the part they want to see by changing the observation point. However, the existing coding and storage technologies do not support panoramic video processing, so video need to be projected on a two-dimensional plane. In fact, during spherical projection process there is a serious oversampling problem in the high latitude area [3], so a lot of bits are needed for transmission which results in a lot of waste. In order to reduce the transmission bandwidth of panoramic video, the common processing methods are Equirectangular Projection and Cube Projection and so on. Although these projection methods can reduce the transmission bits of panoramic video, the bandwidth and playback complexity of panoramic video are still much larger than traditional video. This essay mainly discusses two parts. The first part is how to use the Equirectangular Projection technology to get 360 degree video and how to use a region adaptive smoothing technology to solve the problems brought by this projection technology.

The second part is talking about cube projection and how to correct the motion compensation to eliminate the geometric distortion caused by this projection method.

## 2. Equirectangular Projection

Equirectangular Projection is a kind of projection method that divides the sphere by longitude and latitude, mapping them into vertical lines and horizontal lines respectively with equal spacing so as to form a rectangular grid with equal spacing.

All latitudinal and longitude grids intersect at 90°, in this projection, the poles are represented as lines passing through the top and bottom of these grids. This kind of projection is between equal angle and equal product projection, so it has the problem of both angle deformation and area deformation. The distortion will be distributed over the whole image. When the equator is projected onto the image, the distortion will be minimized.

In panoramic image and video, the realization idea of Equirectangular Projection is to save the data on each latitude line with the same number of sampling points, so as to get the corresponding rectangular video on the two-dimensional plane. The specific methods are as follows:

### 2.1. Process:

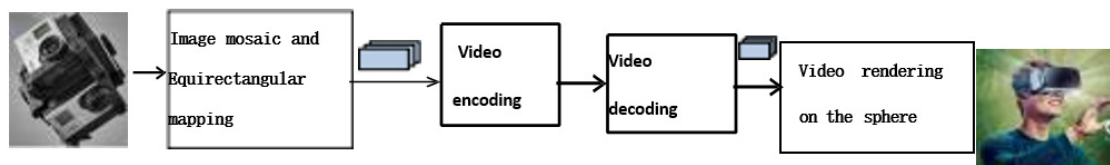


Figure 1. process of 360° video from shooting to rendering (Picture from <https://ss0.bdstatic.com> searched by baidu.com)

**Step1:** Capture 360 Degree Video—At each t time, you can capture the longitude and latitude points of the scenery around the camera position in all directions by fixing the camera position ( $V_x, V_y, V_z$ ) and get the corresponding all-optical function. Then get the X and Y coordinates of the plane point according to the mapping relationship between the all-optical function and the two-dimensional image. In addition, longitude and latitude points can be sparsely sampled along

the directions of  $V_x, V_y$  and  $V_z$  without fixing the camera position. At this time, the all-optical function records the angle of latitude and longitude and the position of the camera at t time [1]. Each image corresponds to Equirectangular Projection mapping of different all-optical functions (fixed or unfixed camera position), which transforms the latitude and longitude of the earth into Cartesian grid, and forms 360° video through two-dimensional image sequence in time.

**Step2:Image Mosaic And Equirectangular Projection Mapping**—For example, using seven cameras, five of them monitor the front, back and side of the scenery and one monitors the top and one monitors the bottom. Firstly, seven images taken by seven cameras are aligned and then these images are stitched together. These images are the vertical field of view generated by simulating the chessboard image. In order to handle the splicing process better, the camera uses different color coding for these chessboard images, such as using black to represent top part and blue to express bottom part. Finally, the three-dimensional video is spliced and output by using the VideoStitch software [1].

**Step3: Video Decoding and Encoding**—The three-dimensional video output from the splicing process is compressed into two-dimensional video by using standard video codecs (such as h.264/hevc, etc.) [1]. At this time, two-dimensional video can be streamed, downloaded and decoded normally.

**Step4: Video Rendering On Sphere**—After decoding, the texture of mosaic image is mapped to three-dimensional virtual sphere. When using the three-dimensional video viewer KolorEyes to view the spliced image, users will feel that they are standing in a space with ceiling, floor and five walls, and the video viewer will present a screenshot of the corresponding output image.

## 2.2. Problem:

In the process of image mosaic and Equirectangular Projection mapping, it can be observed that the top and bottom images occupy more pixels than other direction images. This shows that Equirectangular Projection method will lead to the problem of uneven pixel density distribution. This phenomenon also exists in the world map, for example,

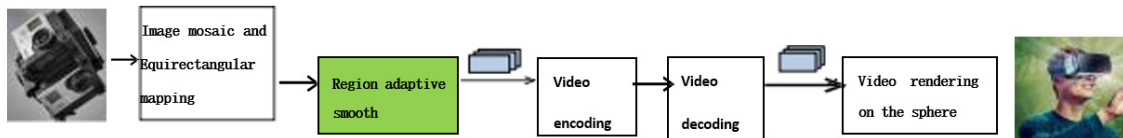


Figure 2. 360 ° video processing with area self-smoothing technology (Picture from <https://ss0.bdstatic.com> searched by baidu.com)

Fixed smoothing refers to the fixed image top and bottom. By changing the percentage of the smoothed top image and bottom image(pSmooth), we can observe the effect of different pSmooth values on the bit saving rate. We can use x264 (QP = 27) to calculate the saved bit rate, and use GearVR to determine the corresponding pSmooth value when the perceived quality drops [1]. Firstly, we collect 360 videos of different categories for data analysis: natural video and computer animation. The experimental results show that the perceptual quality of natural video decreases slightly when psmooth value is 15%, while that of computer animation video decreases slightly when pSmooth value is 20% [1].

Variable smoothing is to control the smoothness by increasing the variance of Gaussian smoothing filter when moving to the top and bottom of the image. First, the relationship between the resolution of the sphere and X is obtained when the image is in Y coordinate, that is, when y decreases to the bottom, X also decreases; then the relationship between the variance of the Gaussian smoothing filter and Y is obtained, that is, when y moves from the center of the image to the top of the image, the variance value becomes larger, corresponding to the smoothing value from small to very strong[1]. When the variable smoothing value

it will be concentrated in the south, the north pole and other top parts where the image quality is relatively high, in contrast to that the image quality near the equator will be the worst part. So this is different from our habit of watching movies. Specifically speaking, if the theme of the video we shoot is in the middle of the picture, but the camera focuses on the ceiling and the ground, which causes the theme part to be fuzzy, the unimportant part to be clear, and the essence to be inverted.

On the other hand, after stitching with the Equirectangular Projection method, it can be observed that the picture quality of the horizontal line in the center of the picture is relatively poor, and the straight line will be distorted into a curve, that is to say, the original straight-line moving object will become non-linear moving after projection. It will reduce the feasibility of dynamic compensation, resulting in the difficulty of compression and the waste of effective resource storage so that making the image quality does not look high under the same video capacity.

## 2.3. Solution:

Before video coding, the top and bottom image regions are needed to carry out adaptive smooth pretreatment. Because adaptive smoothing filter can remove noise and increase details at the same time, so as to achieve the best enhancement effect of the image. Moreover, the pixels of the top and bottom images are relatively high, even if the image is smoothed, the resolution of the image will not be reduced and because the change coefficient of the smooth area is relatively small, many bit rates can be saved. Among them the adaptive smoothing technology is divided into fixed smoothing and variable smoothing. As shown in Figure 2.

K increases, it not only enhances the image at the bottom and top of the smoothing, but also the image at the center of the slightly smoothed image can also be enhanced. Therefore, the variable smoothing can save more bit rates than the fixed smoothing, so that the 360 degree video can be more effectively compressed.

## 3. Cube Projection

Cube Projection is a kind of projection method that maps the panoramic image of sphere to the faces of cube, each face has a 90 °×90 ° field of vision, and corresponds to the images of different directions of panoramic sphere. Its mapping method is simpler, only need to paste each face to the corresponding cube face. Cube projection can achieve the uniform distribution of pixels. Basically, the pixels at the two poles and at the side are the same, and this projection reduces the redundancy of the two poles, so it reduces a certain amount of data. In addition, the straight line can be kept straight in the projection of cube project. Compared with the Equirectangular Projection, it is more conducive to video coding, and the distortion is much smaller.

### 3.1. Process:

When storing the projected spherical video, in order to effectively utilize the existing video compression technology, firstly expand the projection of the cube sphere and map it to multiple areas in a texture. For the projection mode of multiple faces, rectangular frames of different sizes can be designed to expand the cube face and place it on a two-dimensional plane image: non compact layout and compact layout.

Non compact layout: the image aspect ratio is 4:3. Because the adjacent areas in the plane image are also the same adjacent faces on the cube, so each face of this layout keeps a good image continuity, which is conducive to video coding. However, due to the increase of six invalid areas in the  $4 \times 3$  layout, it is more like a lying cross, as shown in Figure 3. These invalid pixels account for half of the total number of plane image pixels, so it will increase the complexity and time-consuming of coding, resulting in the poor coding efficiency.

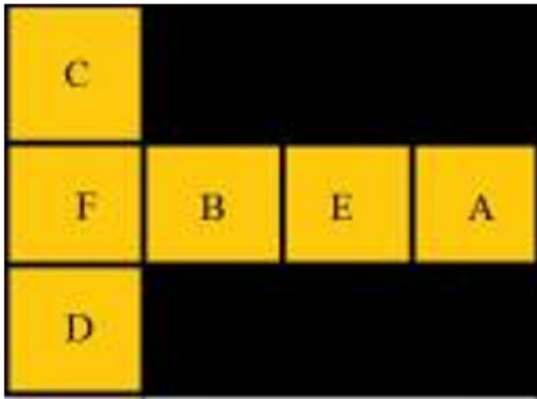


Figure 3. cube projection\_4x3 layout

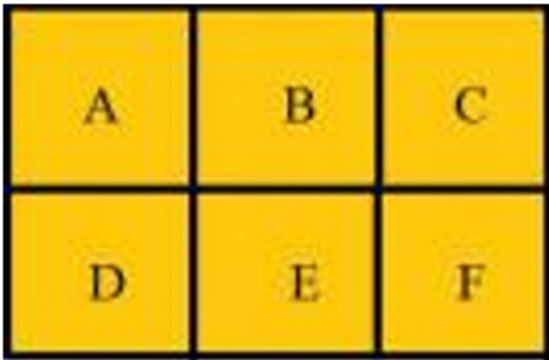


Figure 4. cube projection\_3x2 layout

Compact layout: in order to avoid the influence of invalid area, the image aspect ratio is 3:2, six faces are arranged in two lines, three faces in each line, as shown in Figure 4. The multi-dimensional data of these areas are connected with each other, so there will be no blank space in the generated video [2], and the entire plane pixel is an effective pixel. The number of discontinuities in the image distribution is reduced to only one, which maximizes the image continuity [2]. In figures 3 and figure 4, ABEF is the side, C is the top, and D is the bottom.

### 3.2. Problems:

Because the straight line in the projection of cube project is easy to bend at the boundary between two faces, that is, the boundary between two crossing faces will have geometric

deformation, so when the motion is carried out at the edge of the face, the motion vector is not a straight line and the motion compensation will fail during the encoding. For example, in the 360 degree video of face image taken with pinhole camera, the edge of the face is most likely to deform. The intensity of deformation is related to the angle of the edge. The steeper the angle is, the more severe the deformation is, and vice versa.

### 3.3. Solution:

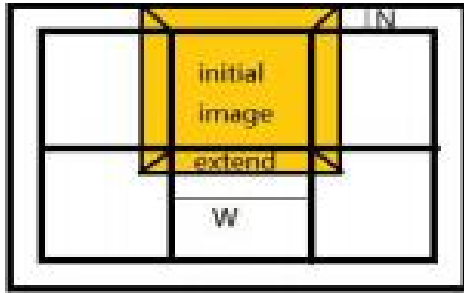
Because 360 degree video has rotational symmetry, the content removed from the frame will appear in another frame again, rather than disappear [2], so this can be used to correct the curvature of the image across the polyhedron boundary, project the part of the connected face to the prediction block image so that improving the motion compensation and correctly dealing with the geometric distortion. The specific solution steps are divided into multi-facet extension and cube face extension, and put them integrated into the video coding scheme.

Multi-facet extension refers to the extension of the area of a cube by using the data of its adjacent facets. To achieve this, only the adjacent facets need to be projected onto the plane to be expanded to generate the current front facet. The projections of adjacent facet and the current front facet are homography, and they all share the same camera center as the center of the sphere performing the stitching phase.[2] To be more precise, how to transfer the image of adjacent faces to the current front image should include three points. The first is the correction matrix reflecting the projection relation between the three-dimension points on the adjacent and the current surface. Even if the distance between the points on the adjacent surfaces and the center of the sphere is not known, but the correction matrix can be used to make the points on the adjacent surfaces be projected onto the current surface one by one. The second is to change the distance from the adjacent plane of the coordinate system to the current plane. According to the angle between the two planes, rotating around the common edge, the three-dimension points on the adjacent facets are represented as the three-dimension points corresponding to the current front facet. Last, the third is to represent the 3D points on the current front facet as the points on the two-dimension according to a certain transformation relationship. In this way, through the above three points, we can get the complete homography.

The expansion of cube refers to how to use the expanded facet to correct the geometric deformation of intersection from one face to another based on the fact that 360 degree video has been mapped to each face of cube. Even in panoramic video, these lines are straight, but when projected to a plane, there will always be some curvature at the edge of the facet. On the premise of generality, the resolution of different cube faces is the same [2]. Using the correction matrix, the projection of adjacent facet on the current front (rotating around the y-axis) is obtained, and the expanded image does not have these deformations. Through this method, we can get the extended image of other corresponding connecting surfaces, which can be observed as a plane with all extended sides. Moreover, the farther the distance from the extended region, the smaller the sample density of the extended region.

Lastly, we need to put them integrated into the video coding scheme. Generally, the compact cube  $3 \times 2$  layout is adopted. A reference picture for modifying the picture expansion method used for motion compensation is shown in Figure 5.

Because the extended facet cannot be overlapped, the extended facet of each part on each frame have corresponding reference pictures, so it has a total of 6 reference pictures. The surrounding area of the original image in Figure 5 expands the data projected from the connection part, and the size can be configured. For the symmetry of panoramic video, we commonly select a picture with the same size as the picture frame extension used for motion compensation so that the content removed from the image can be extended to the adjacent facet more conveniently. As for how to locate the location of the accessed block, the encoder and decoder can be derived from the CU location in each figure [2].



**Figure 5.** The reference image is divided into a layout with an image aspect ratio of 3:2. The orange area extends the reference image to the top of the cube face. [2]

In the existing experimental results, we can divide the sequence of static camera and non-static camera video to evaluate the effect. Without affecting the average PSNR, the speed of static camera video is increased by 0.06% [2]. For non-static camera video, not only the speed is increased to about 2%, but also the PSNR is increased, so the motion compensation effect is better. For this method, we need to consider the size of its memory requirements and the complexity of calculation. By setting and calculating the width  $W$  of the picture and the size  $N$  of the extended area, we can ensure that the extended area is only need to generate once and the low complexity of calculation is realized. [2]

## 4. Results

At present, there are more and more panoramic video applications, especially in the VR field. In order to make users have a better experience, panoramic video needs to be more

processed. When the panoramic video is projected to the plane by using the existing encoding and storage technology, we need to apply the appropriate compression method according to the properties of the different images and video shots: not only reducing the transmission bit rate, but also ensuring the quality of the image. Through the above introduction and analysis, the common video projection methods are Equirectangular Projection and Cube Projection and they have their own disadvantages. However, at present many experts are committed to eliminate this disadvantage and put forward many solutions. For example, the equirectangular projection has the problem of uneven pixel density distribution after projection. Taking advantage of the uniqueness of panoramic video, a region adaptive video smoothing technology can save about 20% of the bit rate [1], and at the same time, it does not need to add additional video decoder algorithm. While the method of cube projection has the problem that the boundary between the two crossing facets after the projection is distorted, and the current solution is the extension of multi-facet and cube. It is a method of panoramic video motion compensation which is used to predict the blocks between the two crossing surfaces with reference pictures, and the complexity of coding is reduced by setting some variables so that eliminating geometric distortion. Generally speaking, the straight line of the cube projection in the internal plane can keep straight, but equirectangular projection could not ensure that. So Cube Projection is more conducive to video coding and motion compensation, and the distortion is also much smaller.

## References

- [1] M. Budagavi, J. Furton, G. Jin, A. Saxena, J. Wilkinson, A. Dickerson, "360 degrees video coding using region adaptive smoothing," in Proceedings of the IEEE International Conference on Image Processing(ICIP), 2015, pp 750 -754.
- [2] J. Sauer, J. Schneider, M. Wien, "Improved motion compensation for 360 video projected to polytopes," in Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), 2017, pp. 61 -66.
- [3] WANG Yueming, WANG Ronggang,GAO Wen,"Efficient Projections for Immersive Video Coding and Transmission" in ZTE TECHNOLOGY JOURNAL,Dec 2017 Vol.23 No.6,pp. 50 -55.