Arbitrary Meshes to Simulate Natural Video Objects

Olena Osharovska
Television and Radiobroadcasting Department
O.S. Popov's ONAT
Odesa, Ukraine
osharovskaya@gmail.com

Нерівномірні Сітки для Моделювання Натурних Відео Об'єктів

О.В. Ошаровська
кафедра телебачення та радіомовлення,
ОНАЗ ім. О. С. Попова
Одеса, Україна
osharovskaya@ gmail.com

***Abstract*—The paper considers the option of creating three-dimensional mesh objects from images natural scenes. The image contains both areas with a smooth change in brightness and chromaticity, and areas of detail. To divide the scene into sections with different details, the gradient method was used. A comparison is made with respect to the digital stream being created when the luminance component is represented by vertices of an equidistant uniform mesh and with an adaptive no equidistant mesh. For the case of an adaptive mesh, Delaunay triangulation is used for the threshold values of the gradient module. It is shown that the gain in compression of a luminance signal by an adaptive mesh increases when coding images of high and ultra-high definition.**

***Анотація*—У статті розглядається можливість створення тривимірних сітчастих об'єктів із зображень природних сцен. Зображення містить обидві області з плавним зміною яскравості і колірності і області дуже щільною детальністю. Щоб розділити сцену на ділянки з різними деталями, був використаний метод градієнта. Проведено порівняння щодо цифрового потоку, коли компонент яскравості представлений вершинами рівномірної сітки і з адаптивним не рівновіддаленою сітки. Для випадку адаптивної сітки, тріангуляція Делоне використовується для порогових значень модуля градієнта. Показано, що виграш в стисненні сигналу яскравості за допомогою адаптивної сітки збільшується при кодуванні зображень з високої і надвисокої чіткості.**

Keywords—Image; scene; meshl; triangulation; detail; gradient; bitrate

Ключові слова—Зображення; сцена; сітка; тріангуляція; детальність; градієнт; швидкість цифрового потоку

I. Introduction

In this paper, we will focus on methods for creating a high-quality image on the transmitting side. Object-based approach MPEG 4 and JPEG 2000 audiovisual codecs operate in a frame- and block-based fashion. A scene is composed and rendered at the sender site. Video frames and audio are coded, multiplexed, and transmitted to the receiver. At the receiver site, the transport stream is demultiplexed, and video and audio data are decoded, synchronized, and presented as defined by the sender site. MPEG-4, however, codes objects separately, using different coding tools for video, speech, or music. A description of the scene, called binary format for scenes (BIFS), dynamically describes the scene and allows objects to be animated. Thus, scene composition is shifted to the decoder. This object-based approach also permits choosing an appropriate coding tool for each audio, video, and graphics object. From the point of view of transmission systems, stereo is still used with an additional depth signal [1, 2].

II. The two-dimensional mesh model

The advantage of mesh-based processing is that it can always provide better performance for both spatial and temporal interpolation. Therefore, it has been found being more effective to be used for inter-view prediction since it can provide a coarse representation of the disparity field with adequate models degenerated from three-dimensional mesh. On the other hand, mesh-based video coding is a new technique for motion estimation and compensation.

There are several assumptions for the two-dimensional mesh model. First, there is no disparity discontinuity, that is, the surface of the object is continuous. Second, no occlusion occurs. With these two assumptions, the two-dimensional mesh model can provide the same performance as that of the three-dimensional mesh model. There are two kinds of two-dimensional mesh: regular mesh and arbitrary mesh [2D mesh]. With arbitrary mesh, the delicate surface structure can be represented with small patches, and the smooth surface can be represented with large patches. Therefore, the performance of arbitrary mesh is better [3, 4].

Consider the creation of three-dimensional mesh of objects from the real scene. The image of dolphins in the sea contains practically flat areas without a shallow structure, as well as ripple structures on the sea and clouds in the sky (Fig. 1).

Fig. 1. The image of dolphins in the sea

Select from the original image individual objects: dolphins, sea spray, smooth sea surface. There are several methods for selecting objects; they are based on splitting the entire image into separate sections or clusters, differing either in color, or in texture or semantic content. We used a gradient method for isolating the contours of individual clusters. Fig. 2 shows the image of clusters.

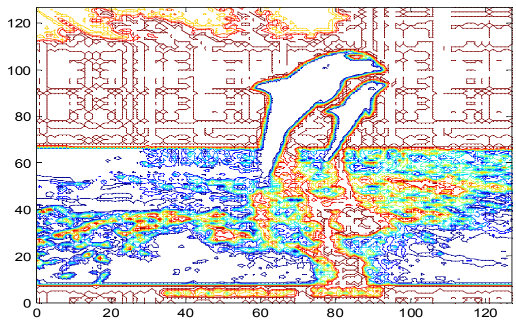

Fig. 2. The contours of individual clusters

Using program Matlab, a portion of mesh model was selected and vertices of these clusters were identified.

Analyzing the gradient mask, one can conclude that there are objects with different details. Our task is to build a three-dimensional image. In accordance with the details of individual regions, we apply a grid partition, and on even areas the grid will be uniform and with a large step, and on regions with detailed use of a non-equidistant grid. A three-dimensional model is obtained based on the orthogonal grid and the values of the gradient of the luminance signal at each point. In the next step, each three-dimensional object is approximated by the Delaunay triangulation. The number of triangles on objects of different detail is differ hundreds of times.

Let us examine the dependence of the number of mesh vertices on the choice of gradient threshold steps. Table 1 shows the dependencies for the lowest resolution image obtained because of thinning with interpolation of the original image to a size of 128 by 128 pixels. Denote the number of vertices for a uniform mesh as T1 and for an uneven mesh as T2.

TABLE I. DEPENDENCE OF THE NUMBER OF MESH VERTICES ON THE CHOICE OF GRADIENT THRESHOLD STEPS 128×128.

Gradient threshold step	T1	T2
0.1	32 258	3 256 800
0.25	32 258	524 228
0.5	32 258	232 562
0.75	32 258	131 072
1	32 258	57 800

Tables II – IV show the simulation results for the images 256×256, 512×512, 1024×1024.

TABLE II. DEPENDENCE OF THE NUMBER OF MESH VERTICES ON THE CHOICE OF GRADIENT THRESHOLD STEPS 256×256.

Gradient threshold step	T1	T2
0.1	130 050	13 107 200
0.25	130 050	2 097 152
0.5	130 050	524 288
0.75	130 050	232 562
1	130 050	131 072

TABLE III. DEPENDENCE OF THE NUMBER OF MESH VERTICES ON THE CHOICE OF GRADIENT THRESHOLD STEPS 512×512.

Gradient threshold step	T1	T2
0.1	522 242	52 428 800
0.25	522 242	8 328 608
0.5	522 242	2 097 152
0.75	522 242	930 248
1	522 242	524 288

TABLE IV. DEPENDENCE OF THE NUMBER OF MESH VERTICES ON THE CHOICE OF GRADIENT THRESHOLD STEPS 1024×1024.

Gradient threshold step	T1	T2
0.1	2 093 080	211 226 044
0.25	2 093 080	33 554 438
0.5	2 093 080	8 328 608
0.75	2 093 080	3 726 450
1	2 093 080	2 097 152

For example, in the area of an equidistant mesh where the gradient remains practically unchanged, 373 triangles have left for the description of image objects with a gradual change in the gradient of the image. Moreover, in a non-equidistant mesh where brightness jumps occur, 6275 triangles, a minimum triangle size of 3 picture elements, and a maximum triangle of 126 picture elements were required.

We estimate the digital luminance signal stream at the standard progressive scan the original image. Based on the image size M×N ultra-high definition, progressive scan with number of frames in a second n, we will take a bit digital luminance signal is equal to 10. Then the upper cut-off frequency can be determined by the formula (1)

$$f_{max}=M \times N \times n/2, \qquad (1)$$

$$f_{max}=3840 \times 2160 \times 50/2=297.4 \text{ MHz}$$

The sampling frequency of the luminance signal of UHDTV is $fs_Y = 432$ MHz, and the speed of the digital stream will be 4320 Mbps for the luminance signal. The number of vertices and the bit depth of the codes for the three coordinates of each vertex can estimate the flow created by triangulated objects. We can assume that the number of vertices of V is approximately twice as large as the triangles of T. In our example, the minimum dimension of the side of the triangle is two decomposition elements and we can limit ourselves to the twelve-digit code for the vertices for all three coordinates. In this case, a rough estimate of the digital flow can be represented by the formula (2)

$$V_{\Sigma}=V_{sm}+V_{det}=[3 \times m \times n \times 2 \times (T_{sm}+T_{det})] \qquad (2)$$

$$V_{\Sigma}=[3 \times 12 \times 50 \times 2 \times (373+6275_{t})] = 23.93 \text{ Mbit/s}.$$

where V_{sm} – the bit rate for the polygons on flat areas;
V_{det} – the bit rate for polygons in areas of great detail;
T_{sm} – number of triangles on flat areas;
T_{det} –number of triangles;
m – bit code for vertices;
n – number of frames transmitted per second.

Moreover we found values of bitrate in each case. The result you can see in the Table V.

TABLE V. BITRATE OF DIFFERENT IMAGE SIZE AND MESH CODING METHODS

Image size	Bitrate, Mbit/s
128×128	0,38
256×256	0,38
512×512	1,50
1024×1024	6,00
2048×2048	18,76
4096×4096	23,93

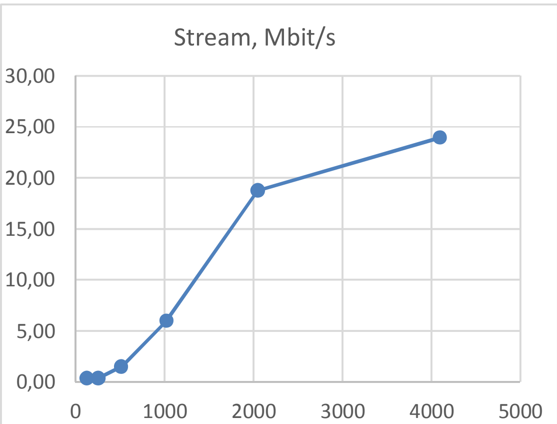

Fig. 3. A plot of the bit rate of the image size for mesh coding methods

Even with this approximate analysis it is shown that the transition from the progressive scan, when information is transmitted about each picture element to an object representation allows you to get the original digital stream which is much less for ultra-high-definition images. From the graph in Fig. 3, the dependence of the digital bit stream on the size of the images it is seen that the gain from the transition to object-oriented representation of scenes appears only for images with the size larger than 2K.

III. The Some Aspects of Assessing the Quality of Mesh Models

There is no doubt that the introduction of 3D TV can have a lasting success only if perceived image quality and viewing comfort, at least comparable to a conventional high-definition TV.

The main emphasis in our work is made on the concept of an image quality model, which is a key component of the development of an object-oriented television system. This concept can be defined on the basis of formally perceived attributes related to this system, and how these attributes can be integrated to predict image quality.

To date, the most accurate ways to assess the perceived quality of 3D are subjective assessment methods that are obtained by statistical processing of the results of examinations. The disadvantage is that these studies are time consuming. In addition, universal procedures for assessing the quality of stereoscopic images do not yet exist; they adapt the existing 2D standards that should be used correctly for subjective evaluation of 3D content.

In the subjective examination of TV stereo images, along with resolution, color reproduction and other parameters, which are usually checked in monoscopic TV systems, other characteristics specific to stereo systems are also analyzed, including resolution of the depth of the image, movement in its depth, distortion of the type of effect miniaturization (unnatural size of objects) and the effect of "cardboard" (unnatural thickness of objects), the comfort of viewing images, etc.

In multimedia applications, such as games and simulators, object-oriented methods for creating the bulk effect have become more widespread. Such objects require painstaking work to create, and, for the most part, artificially created objects, or even transformed from actually shot stereo. When creating objects from two cameras, there are many difficulties in matching the two types for the left and right eyes, since it is difficult to ensure the identity of the cameras, and the angle of incidence of the illumination is different. The latter circumstance leads to the need to identify points on two images, and by identifying by color, brightness, shadows; there is no possibility of achieving high reliability.

In search of an ideal 3D model of quality, on the one hand, there were numerous studies that tried to determine the quality of stereoscopic 3D only with respect to 2D components of stereoscopic content, as in the case of images and video. On the other hand, there were also studies in which stereoscopic 3D quality was regarded as part of a more complex multidimensional concept. The factors listed, such as reproduction depth, stereoscopic deterioration and visual comfort, must be taken into account when developing a model for evaluating 3D quality.

We accept as input polygon-mesh models of 3D-objects, which is a set of triangles in 3D-space. A mesh consists of a set of vertices V. In addition to three-dimensional coordinates, each vertex contains information about the color represented in the RGB space; the point cloud in space specifies a discrete source object. Each image corresponds to an associated tensor. For each vertex, we want to estimate the color and the color difference from the neighboring vertices. The collection of colors of peaks is a color model to be investigated. According to our estimates, nonlinear spline approximation problems that minimize the color divergence compared to the original continuous object obtained from the camera for each vertex are applicable to the color model. A particular task, to be optimized, is to solve the problem of metamerism of colors.

When using polygon-mesh models of objects, especially those created on the basis of real ones, it is necessary to pay attention to the effect on the overall quality of the change in the reproduction of textures. The establishment of the type of the reflecting surface plays an important role, namely: diffuse or specular. In the case of specular reflection, glare, ghosting, significantly masking other distortions.

The manifestation of the aggregate of distortions in the subjective assessment is the subject of investigations, since deterioration stack with different impact on the final score. One should also take into account the masking effect of some distortions on others. Neither of the existing image bases included information on the spectral composition of the primary colors, nor does it include the characteristics in which color system the primary colors are represented.

One of the fundamental problems of image analysis is the creation of an adequate mathematical description of images that convey their content, meaning. In other words, this description should reflect only significant (from the point of view of the problem being solved) image features, and not depend on non-essential details. The methods of morphological analysis are, therefore, a step towards solving the problem of describing the semantics of images

At the heart of the methods of morphological analysis lies the mathematical concept of form. The form of the image is understood as the maximum invariant of image transformations, to which it undergoes changes in observation conditions, changes in the parameters of recording equipment,

Contextual or content description of images is a certain set of features that characterize this image. Therefore, it is important what source of descriptions is used to obtain them, since it determines what methods and means of analysis and representation of images can be applied. The main problem of content image retrieval is the translation of the content of images from the low-level representation of features (color, texture, shape, position, etc.) into the semantic plane, that is, the system's understanding of the content of the image at the level of the image, scene, situation and events.

The task of creating the most objective description in the form of a semantic annotation involves enumerating all possible objects potentially interesting to the user with support for objective data of different types, in particular: - the actual relative size of the object in the image, for example, the part of the image area occupied by the object in relation to The area of the whole image or other object; - perspective - reducing the size of objects when moving away from the observer; - knowledge about the device, relative dimensions (for example, in comparison with the size of a person) and the functionality of objects. The concepts of ontology are elements of the semantic annotations of images represented in the XML formalism.

IV. Conclusion

The development of the means of production of television programs is faced with the tasks of an adequate assessment of the quality of images and sound.

Large screens allow viewers to view the smallest details. For their unmistakable transfer, it is suggested to use the division of the scene into separate objects. Depending on the details of these objects, a different number of polygons represents them.

It is proposed to use 2D polygon-mesh models, and the size of the triangles and their number depend on the detail. Investigations of the model based on the gradient of the luminance signal have shown the possibilities of winning in the volume of the transmitted data due to the redistribution of the reference points for areas with different details.

Further research is aimed at developing a criterion for assigning images to a particular detail group.

References

[1] Draft New Recommendation ITU-R BT.[3DTV-REQS] Performance requirements for the production, international exchange and broadcasting of 3DTV Programmes.

[2] Recommendation H.264 (10/16): Advanced video coding for generic audiovisual services

[3] ISO/IEC, JTC1/SC29/WG11, "Coding of Moving Pictures and Audio - MPEG-4 Visual Standard, ISO / IEC FIS 14446-2, 2000.

[4] A. Murat. Tecalp and all Two-Dimensional Mesh-Based Visual-Object Representation for Interactive Synthetic/ Natural Digital Video // Proceedings of the IEEE vol 86, No 6, june 1998, pp. 1029 – 1051.

[5] K. Mamou, N. Stefanoski, H. Kirchhoffer K. Muller, T. Zaharia, F. Preteux, D. Marpe, J. Ostermann. The New MPEG-4/FAMC Standard For Animated 3D Mesh Compression // 3DTV-CON'08, May 28-30, 2008, Istanbul, Turkey. pp. 97-100