Associative Relations in the Assessment of TV Images

Oleg Gofaizen, Olena Osharovska, Mikola Patlayenko, Volodymyr Pyliavskyi
Dept. Radio and Television Broadcasting,
O.S. Popov Odessa National Academy of Telecommunications
Odessa, Ukraine
Email: oleg.gofaizen@gmail.com; osharovskaya@gmail.com; nick_msa@ukr.net; v.pilyavskiy@ukr.net

***Abstract*—With the creation and international standardization HDTV and UHDTV systems becomes an actual task of specification test signals to be used in the construction and maintenance video path and equipment of this systems. In this paper suggestions are made on the associative relations in the assessment of TV images. The semantic relation between the objects of the frame will be the most valuable information that requires transmission with the highest priority. The proposals can be considered as the initial to new step of progress of metrological provision technical implementation HDTV and UHDTV systems**

Keywords—3DTV, mesh, video path, image quality, quality assessment, semantic relation, date base.

I. Introduction

The paper contains proposals on three-dimensional (3D) mesh that is currently used in many multimedia applications such as digital entertainment, medical imaging, and digital television system. In the process of visualization 3D-mesh, undergo some operations lossy compression, simplification, turns, putting watermarks. Since end-users, in most cases, people, it is important to obtain indicators that can accurately evaluate the perceptual distortions introduced by such operations [1]. Classic metrics simple geometric distances (eg, mean square error and the Hausdorff distance) [2, 3], are not consistent with human visual perception and, therefore, is unable to predict the visual difference between a pair of reference and deformed mesh object Material of this work is a continuation of previous studies in this area [5]. It should be noted that the parameters of color for HDTV systems have been proposed in [5]. However, sub-sequent analysis done by the authors has shown that to achieve compatibility with existing measurement equipment proposed set of parameters should be modified in terms of time characteristics, and this modification is made in this work.

II. Mesh object composition

TV mesh objects are combining several structures, in particular:

- polygonal mesh cage;
- a set of matching texture;
- a set of normals to landfills;
- the color information for each polygon;

Each of these sets in its own is coded and transmitted. Effectively given polygonal mesh model is suitable for television broadcast only if the object-oriented construction of the scene. Trips to mesh compression quality assessment, usually based on a comparison of the starting mesh with it image distorted copy. Graphical representation of the mesh image is represented in Fig. 1.

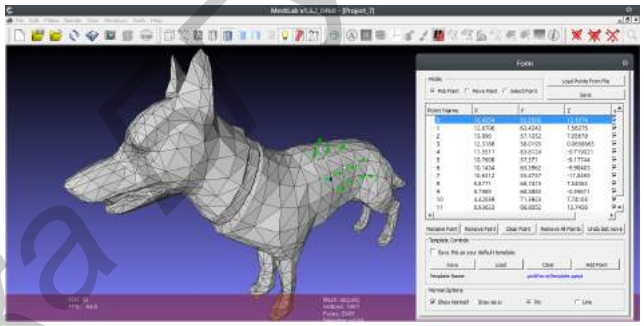

Figure 1. Mesh representation of the image

It can be assumed that by specifying the description of the object of the mesh set of semantic relations, with the subsequent search in the database corresponding to images of objects, allowing you to replace them distorted.

III. The semantic relations in the TV frame

P.A. Florensky had shown that human perception of three-dimensional images are always carried out in time, it is always discrete, and it has a certain rhythm. Then, any message must be configured or "stapled" of closed ones in the space-time events and objects. The audience watching the static frame for 3 - 4 seconds and then if nothing happens begin to lose interest in it, or closely begins to learn the details. A frame is a minimum unit of the television language. It is like words in a language, respectively, the frame carries meaning. The semantic relation between the objects of the frame will be the most valuable information that requires transmission with the highest priority.

Multimedia space can adequately replace the digital image signal or audio images of their "verbal" representation and vice versa. The associative memory search is carried out without the use of direct or indirect addressing but according to the content of information Search for data on the basis of their content is based on a limited comparison of some important parts of the words stored in the memory cells. The procedure is advantageously carried out by comparing the rules of fuzzy logic, which allows to find similar, but not exactly the word, wondering interval criterion relevant key information.

However, this understanding of the associative memory represents, in essence, only the fact of the relationship between the data and has nothing to do with the very mechanism of information storage. Therefore, to refer to such information storage mechanism uses the term "memory content addressed". In contrast, the unit you can store data integrity in a particular cell, having only information on direct connections a given cell with the other - so we come to understand the semantic networks. These principles are also applicated for indexing and searching the modern databases. The concept of linguistic variable and its application to accept approximate solutions capable of dynamically generate the object based on the association of a text description and audio or video image. [4] The search in the database can be organized on a linguistic change, to open access to the association by the file storing the image or sound. Sami linguistic variables is desirable to be formed by use of a multilingual interface. It can be argued that the associative sample in a narrow sense is equivalent to pattern recognition.

IV. The process of image verbalization

Image verbalization occurs in the following sequence: the image contours are monitored, after the corresponding differences of brightness; color segmentation performed. Then, the approximation obtained for contours and color boundaries.

For example, here, the image search procedure comprising similar dominant color ranges in the database. Comparison is made on the basis of the shortest distance from Manhattan found the dominant color in the selected color coordinate system of a predetermined set of basic colors classified by brightness, saturation and hue.

The resulting color is used as an identifier for the organization of the search procedure in the image database. Determine the association of colors from a core set of primary colors for the formation of the characteristics of the visual similarity of colors and visual contrast of colors. Generate index (metadata) of the image to the required basic color in RGB format, and / or name and / or the tag color. Searches for images in the information database.

In the search result list form index images coinciding with the presented index and / or indices in which the primary colors are present in the list of visual associations or similarities in color contrast to the list of primary color images in the search index.

V. Databases of images

Closest in technical essence is a method of extracting digital images from a database of images [5,6,7], wherein the presented image is calculated histogram and histograms of images stored in the database, then the image similarity measure is calculated using a method of assessing the similarity of two distributions probabilities. Then, the search results are ranked according to the degree of similarity of the input image. [8]

Currently several universities developed image database as the high and ultrahigh color definition, and three-dimensional. Base images of various objects created by members of Intelligent Systems Lab Amsterdam lab. The database contains images of 1,000 items (one-thousand small objects. Each object presented three series of images. The first series includes 24 frontal images of the object taken in different lighting direc and light color. Each object was recorded frontally, with five included bulbs. lighting colors changed from 2175K to 3075K. White balance in the cells was set to 3075K, whereby illuminated objects from reddish to white light. The third series includes 72 images taken at different positions of the camera on the subject. The database also includes a stereo pair for 750 objects. Each image in the database is represented in the color and grayscale form in three different resolutions: 768 $\times$ 576, 384 $\times$ 288 and 192 $\times$ 144 pixels. (Tabl.1)

Table I. ALOI Media database 1000 different objects

Amsterdam Library of Object Images	Processing method
24 frontal image of the object	different directions of illumination
12 Frontal image of the object	different light color
72 images relative to the object	at different positions of the camera
stereo pair for 750 objects	in color and grayscale form

Image database for testing three-dimensional object recognition algorithms developed at New York University. The base is designed for experiments with recognition of three-dimensional objects. It contains 50 images of toys belonging to 5 main categories: the four-legged animals, human figures, airplanes, trucks and cars. The images were taken by two cameras at 6 different lighting conditions, with 9 slopes and 18 azimuths.

Image database of persons trained at Yale University. The database contains frontal images of faces. The base consists of 165 images of 15 people. Each person is represented by a series of 11 images: with three different light; with glasses and without glasses; six different facial expressions. The base of natural texture images, created Massachusetts Institute of Technology. (Tab.2).

Table II. Media databases

Dataset	Type of images
The NORB Dataset, New York University	four-legged animals, human figures, airplanes, trucks and cars
Yale Face, Yale University	images of 15 people with six different face impressions
Massachusetts Institute of Technology	homogeneous texture. 15 texture scenes
Notre Dam datasets	iris recognition

VI. Conclusions

The paper proposes parameters for priority construction of 3D images as a task of searching for images according to the systems by form and a specific color. Image attributes that describe the location, color, texture, category image and other are called an index or image metadata. Image index is stored in the information database. In this picture index must include, along with the tag name of the color, which should be as accurate as possible to characterize the color of the object in the image. Semantic Coding images, audio, images and video moves us to a more general understanding of the world.

In the field of high-definition and ultra-high television pressing issue is to evaluate the image quality based on the properties of human vision. Researchers have proposed

dozens of criteria correlated with the human visual system and allowing to predict the quality of the images. Due to the fact that holding subjective assessment appears long and expensive procedure in recent years are developed color image data bases, which contain typical reference image and distorted images. The image distortion uniquely identified objective metric of distorting factors and the result of the subjective evaluation of this particular type of distortion. An example of such a database is the database TID2013 of television images, developed jointly by Ukrainian and Finnish researchers [6]. The database contains information about the 24 types of distortions including distortion arising from the compression of digital images. Each of the distortions is represented on five levels. Few types of the distortions from the database TID2013.

- additive Gaussian noise;
- spatially correlated Noise;
- quantization noise;
- image denoising;
- JPEG transmission. Errors;
- contrast change;
- change of color saturation;
- chromatic aberrations.

Of course, a single database cannot foresee all possible distortions. Moreover, the feature of the television program is to have a combination of several distortions simultaneously. 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. We should also take into account the effect of masking some other distortion. None of the existing image database did not include information on the spectral composition of the primary colors, and does not include features about color system and its coordinates.

From the point of view of the image search capabilities for associative connections in these bases are not provided until the appropriate descriptors, or other verbal descriptions. The other branch of image databases associated with the collection of color volume images in which the images are presented by the primary colors red, green and blue, as well as depth.

Michael Firman studied in [7] the area of creating a set of image data into eight categories:

- semantics;
- evaluation of the object pose;
- tracking camera positions;
- reconstructing of the scenes;
- tracking of the moving objects;
- human actions;
- identification of persons.

Recent research in the field of computer graphics and 3D TV has led to the creation of larger and more ambitious RGBD data sets, and the number of new databases of images every year shows no signs of. Semantic video tags were added, depth reconstruction was used to capture the entire surface of the object, and generating algorithms entire scene were used to create synthetic data plausible. Synthetic data have attracted relatively little attention to the problems associated with the transfer of the depth of the TV cameras. However, such artificial data may offer many advantages.

Table III. EXAMPLES OF ACTIONS

Examples of actions	Year
Get up, enter room, stand up, mop the floor	2011
Two people interacting	2012
Drink, eat, reading the book	2012
Typical gaming actions	2012
Arm gesture	2012
American Sign Language	2012
Older people performing actions	2013
Scratch head, cross arms	2013
Brushing teeth, drinking, talking on couch	2013
Jumping, bending, punching	2013
Mopping, sleeping, [U+FB01]nding objects	2014
Discussion, smoking	2014
Humans falling over	2014
Italian hand gestures	2014
Humans interacting with computer game	2014
Boxing, forward bend	2015
Sneezing, staggering, punching	2016

Early face datasets focused on the method of acquisition or tended to be quite small. The field has now expanded to include datasets for identity recognition, pose regression, and those where the expressions or emotions are to be inferred. As front-facing depth cameras become installed in laptops and tablets we expect this area of research to continue to gain further attention. Actions being performed include sign language, Italian hand gestures and common daily actions such as standing up, drinking and reading. Examples of actions [7] are given in tab. 3

Three datasets of humans falling over reflect an interest in use of RGBD sensors for monitoring vulnerable humans in their daily lives. The manipulation action is unique in providing semantic segmentation of objects as they are manipulated. Many of these datasets suffer from being filmed in the confines of an office or laboratory, with researchers performing the actions. Filming real people at work and home would help prevent dataset bias and provide a more believable baseline for activity and gesture recognition.

Tracking datasets feature videos of dynamic worlds, where the aim is to detect where an object is in each frame. Like datasets of actions, datasets designed for human recognition typically film people performing activities such as walking. However, the aim now is to recognize the identity, gender or other attributes about the subjects, rather than the activity they are performing.

Aside from a single sequence [6] it is know of no RGBD datasets captured from dynamic scenes with ground truth dense geometry. One option is to use deformable meshes provided for face datasets or fabrics, which can be synthetically re-rendered to give dense correspondences between frames re-render data. Datasets of humans with motion capture data also give a very sparse dense geometry with correspondences. The open challenge for the [U+FB01]eld of dense reconstruction is to directly capture an RGBD dataset of deforming objects with ground truth geometry and correspondences between frames.

In recent years, the number of range scanners and surface reconstruction algorithms has been growing rapidly. Many researchers, however, do not have access to scanning facilities or dense polygonal models. The purpose of this repository is to make some range data and detailed reconstructions available to the public. [9].

The models in this archive are fairly widely used in the graphics, visualization, and vision communities. Things people have done with these models include simplification, multi-resolution representation, curved surface fitting, compression, texture mapping, modeling, deformation, animation, physically-based simulation, texture synthesis, and rendering. The Stanford Bunny is particularly widely used, as surveyed by Greg Turk. This repository might use 3D models and range datasets, and remember that several of these artifacts have religious or cultural significance. Aside from the buddha, which is a religious symbol revered by hundreds of millions of people, the dragon is a symbol of Chinese culture, the Thai statue contains elements of religious significance to Hindus, and Lucy is a Christian angel; statues like her are commonly seen in Italian churches. Keep your renderings and other uses of these particular models in good taste.

To describe the complex volume of real objects meshes need to move in the direction of reducing the computing power, as in the classic form of each vertex of the object is subjected to a mathematical transformation, which increases on the order of the computational cost. To reduce the computational complexity of polygonal - grid model proposed shift in the spectral region by dividing the object in the field with further subband filtering by using wavelet transforms.

To increase the processing speed of a real object Daubechies 4, Wavelet Coiflets 2, Symlets 4, Discrete Meyer, Biorthogonal 2.4, Biorthogonal 4.4: - Wavelet conversion of these types were used. The test object for further processing are not subjected to additional distortions. To increase the rate of the algorithm is necessary to filter out non-significant factors (which are not actually affected, site restoration). We determine the threshold analytically to give a depth of wavelet – transformation.

REFERENCES

[1] M. Corsini, M.C. Larabi, G. Lavoue, O.Petrik, L.Vasa, K. Wang *Perceptual metrics for static and dynamic triangle meshes*. [*Proc. of Eurographics State-of-the-Art*]. Rep. 135-157, 2012

[2] P. Cignoni, C. Rocchini, R. Scopigno *Metro: measuring error on simplified surfaces*. [*Proc. of Eurographics State-of-the-Art*]. Comput. Graphics Forum 17(2):167-174, 1998

[3] N. Aspert, D. Santa-Cruz, T. Ebrahimi *MESH: measuring errors between surfaces using the Hausdor- distance*. Proc. of IEEE Int. Conf. on Multimedia Expo. (2002) 705-708.

[4] Gofaizen Oleg, Osharovska Olena, Patlayenko , Volodymyr Pyliavskyi *Signals for Assessment Image Quality in HD and UHD TV Video Path*. 8th International Conference Ultrawideband and Ultrashort Impulse Signals (UWBUSIS). (5-11 Sept. 2016)

[5] O.I. Eremeev, N.N. Ponomorenko, V. V Lukin *Ocenka tochnostnyx xarakteristyc metodom verificacii metric vizualnogo kachestva na osnove specializirovanyx baz izobrajeniy*. [*Radioelektroni i komputerni sistemy*]. (2): 48-57, 2013

[6] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen. Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi Marco Carli , Federica Battisti, C.-C. Jay Kuo *Color image database TID2013: peculiarities and preliminary results*. [*EUVIP 2013*]. (1): 106-111, 2013

[7] Michael Firman RGBD Datasets: Past, Present and Future, `http://www.cs.ucl.ac.uk/staff/M.Firman//RGBDdatasets/`

[8] ITU-R *Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a reduced bandwidth reference*. Recommendation ITU-R BT.1885: 47, 2011.

[9] The Stanford 3D Scanning Repository, `https://graphics.stanford.edu/data/3Dscanrep`

АССОЦИАТИВНЫЕ СВЯЗИ В ОЦЕНКЕ ТВ ИЗОБРАЖЕНИЙ

О.В. Гофайзен, Е.В. Ошаровская,
Н.А. Патлаенко, В.В. Пилявский

С созданием и международной стандартизаций систем телевидения высокой и сверхвысокой четкости становится актуальной задача спецификации испытательных сигналов, которые будут использоваться при построении и техническом обслуживании видео трактов и оборудования таких системы. В области телевидения высокой и сверхвысокой четкости актуальным вопросом является оценка качества изображений с учетом свойств человеческого зрения. Исследователями были предложены десятки критериев, коррелированных со зрительной системой человека и позволяющих предсказывать качество изображений. В связи с тем, что проведение субъективной оценки представляется длительной и дорогостоящей процедурой, в последние годы разрабатываются базы данных цветных изображений, в которых собраны эталонные типичные изображения и искаженные изображения. В искаженных изображениях однозначна определена объективная метрика искажающего фактора и результат субъективного оценивания конкретно этого вида искажения. Примером такой базы является база телевизионных изображений TID2013, разработанная совместно украинскими и финскими учеными [6]. В этой базе собраны сведения о 24 типах искажений, в том числе и искажения, возникающие при сжатии цифровых изображений. Каждое из искажений представлено на четырех уровнях. Конечно, в одной базе данных невозможно предусмотреть все возможные искажения. Более того, особенностью телевизионной передачи является наличие совокупности нескольких искажений одновременно. Проявление совокупности искажений при субъективной оценке является темой исследований, так как ухудшения складываются с разным влиянием на конечную оценку. Следует учитывать также и маскирующее действие одних искажений на другие. С точки зрения возможностей поиска изображений по ассоциативным связям в таких базах пока не предусмотрены соответствующие дескрипторы или другие вербальные описания.

Другая ветка развития баз данных изображений связана с коллекцией объемных цветных изображений, в которых изображения представлены основными цветами красным, зеленым и синими, а также глубиной. В данной работе сделаны предложения об ассоциативных отношениях при оценке качества телевизионных изображений. Семантические отношения между объектами кадра и будет наиболее ценной информацией, которая требует передачи с наивысшим приоритетом. Можно предположить, что, указав описание объекта множествами смысловых отношений, с последующим поиском в базе данных соответствующих изображений объектов, позволит заменить их на приеме неискаженными. В статье предлагаются приоритетные параметры описания 2D и 3D-изображений для дальнейшего поиска изображений. Атрибуты изображения, описывающие местоположение, цвет, текстуру, категорию изображения и другие называют индексом или метаданными изображения. Индекс изображения хранится в информационной базе данных. Поиск данных на основе их содержания основано на ограниченном сравнении некоторых важных частей метаданных, хранящихся в ячейках памяти. Процедура поиска проводится путем сравнения по правилам нечеткой логики, что позволяет найти подобные, но не точно такие изображения, задаваясь преимущественно интервальными критериями соответствия. Эти предложения могут быть рассмотрены в качестве исходных для нового шага технической реализации метрологического обеспечения систем телевидения высокой и сверхвысокой четкости.