# Animation Content in Frame Analysis

Goran J. Zajić, Milena D. Vesić, Ana M. Gavrovska *Member, IEEE,*
and Irini S. Reljin, *Senior Member, IEEE*

*Abstract* — **Animation frame content is analyzed in this paper. The animation found in frame/video content is usually characterized by specific color distribution and texture. The existence of animation within a frame can be described by carefully selected features. We analyzed both fully and partially animated frames. Here, we present our initial results related to automatic animation interpretation using different types of video sequences.**

*Keywords* — **Video, frame, animation, features, content interpretation, color, texture.**

## I. INTRODUCTION

THERE is a rapid development of internet based technologies and digital media. The production and exchange of video contents has become very popular among the digital media users all around the world. Enormous amounts of video files are daily distributed over the internet and stored without any kind of labeling. Searching such a large video database becomes a real challenge. Furthermore, retrieving desired video file and video content information is of importance. Therefore, the development of a system for automatic classification and labeling of video content is valuable. This field of science is still very interesting for researchers.

Automatic classification and labeling of video content is a very important task in the indexing process of large and heterogeneous collections of video material. This task may be realized either on a global or local level. Globally, videos may be classified into one of several main genres, for instance: movies, cartoons, music, news, sports, and documentaries. Also, the global classification includes the classification into more fine-grained sub-genres, such as: specific types of sports (football, hockey, etc.), movies (drama, thriller, etc.) and similar. Locally, video sequences/segments can be classified according to specific concepts in the video content, such as: outdoor, indoor scenes, action segments, etc.

Video genre classification includes two steps: feature extraction and classification. The selection of features and the feature extraction from the source are very important steps because these steps mainly determine the success of classification and labeling process. Different sources of information have been used for the genre classification task [1]. Text based information is one of the least common approaches for feature extraction. The text could be retrieved from scene text or the transcripts of dialogues, as well as using other external sources. An example is a Bag-of-Words model [2]. This approach has some disadvantages, such as a high error rate due to automatic recognition and high computational cost. Features based on audio information are more widely used than the text. Extraction of these features can be performed in both time [3] and frequency domain [4]. On the other hand, visual elements are the most commonly used source for the feature extraction in video genre classification.

The visual information should be analyzed from various standpoints. Both static and dynamic visual information can be exploited using: time dependence [5]-[6], objects [7], motion [8], color, texture and similar descriptors [9]. The compressed domain may also be useful for such a task, as in [10] where the MPEG coefficients were applied.

In the genre classification task multimodal approaches provide high accuracy results, as reported in [11]-[12]. The use of different types of information from different information sources is considered to increase reliability of the classification. In such approaches different multiclass classifiers are applied [13].

In this paper animation frames are analyzed. An attempt to extract the visual information related to animation is made. Image time series are extracted from video sequences of different genres, both real and animated. Each frame is described with a feature vector which consists of specific color and texture features. The analysis is performed within a multifractal framework. The spectrum slope is calculated for each spectrum. This is followed by the comparison of the obtained values. The results show that the real frames can be differentiated from the animated frames in a multifractal singularity spectrum domain, using adequate postprocessing.

Goran J.Zajić is with the ICT College of Vocational Studies, Belgrade, Zdravka Čelara 16, 11000 Belgrade, Serbia (phone: 381-11-3290828; e-mail: goran.zajic@ict.edu.rs).

Milena D. Vesić is with the ICT College of Vocational Studies, Belgrade, Zdravka Čelara 16, 11000 Belgrade, Serbia (phone: 381-11-3290828; e-mail: milena.vesic@ict.edu.rs).

A. M. Gavrovska, is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia; (phone: 381-11-3370143; e-mail: anaga777@etf.rs; anaga777@gmail.com).

Irini S. Reljin is with the School of Electrical Engineering, University of Belgrade, Bul. kralja Aleksandra 73, 11120 Belgrade, Serbia (phone: 381-11-3370143; e-mail: irinitms@gmail.com).

The paper is organized as follows. After the Introduction, in Section II a multifractal framework is briefly described. In Section III the image features for frame content description are presented. Section IV describes the animation frame analysis performed in this paper. The obtained results of the analysis are presented in Section V. Concluding remarks and future research directions are given in Section VI.

## II. MULTIFRACTAL FRAMEWORK

Multifractal analysis (MA) represents a useful tool for describing regularity of a structure (presented by a variable, a function or measurement results). It is formulated and developed by Benoit B. Mandelbrot [14]. It has been used for the representation of different irregular objects and phenomena, and has become a part of various problem-solving algorithms and procedures in many fields of science [14].

Multifractal formalism is based on a self-similarity characteristic of the analyzed variable structure. This characteristic arises from highly non-uniform distributions in the variable, which are the results of nonlinearity of a system [15]. MA can be used for studying the variable structure dynamics, its long-range dependency, and for the variable characterization within a multifractal framework.

Local and global characteristics of a variable are described by the statistical properties of irregularities found in the variable structure. The description of local and global characteristics can be made in the form of their generalized dimensions (local property) and their singularity spectrum (global property).

There are many algorithms for determination of MA parameters. One of the most common is a box-counting method, where the calculation of MA parameters is based on the following procedure. The time series obtained by the measurements can be covered by the squares with side dimension $l$ (grid). The number of squares which contain some particular measurement value is represented by $\mu_i(l)$. This variable can be interpreted as a probability describing the measuring value $\mu$ positioned in the $i$-th square. Relation between the measuring value and the dimension of squares corresponds to the power-law,

$$\mu_i(l) \simeq l^{\alpha_i}, \qquad (1)$$

where $\alpha_i$ is a rough Holder's exponent [14] and represents the fractality of the structure. In the boundary process, when $l \rightarrow 0$, the value of $\alpha$ will converge to the common value for a given measure $\mu$. Parameter $\alpha$ describes the local regularity of the structure, and it depends on the position within the structure. There is a possibility that the same value of $\alpha$ may exist on different positions in the structure. So, it is interesting to calculate the distribution of $\alpha$ along the measured structure, which provides a representation of a global regularity called multifractal spectrum. The representation is noted as $f(\alpha)$. The algorithm for the calculation of multifractal singularity spectrum can be explained on the example of box-counting method. If $N(\alpha)$ is the number of squares (points in a boundary process) and the value of singularity intensity is in the range [$\alpha$, $\alpha$+d] with probability $P_i$, then $f(\alpha)$ represents the fractal dimension of the structure described by value $\alpha$, equation (2).

$$N(\alpha) \simeq l^{f(\alpha)} \qquad (1)$$

Based on the presented formalism, a multifractal measure can be defined in the form of the Hausdorff dimension distribution of exponent $\alpha$ [15].

Generally, the multifractal singularity spectrum $f(\alpha)$ has a form of a parabola. This parabola can be characterized by: width of the curve, starting and ending position of the curve, slopes on these positions, etc. Signal structure singularities are rapid changes in a time or spatial series. The presence and intensity of the singularities affect the shape of multifractal singularity spectrum [16].

## III. FRAME CONTENT DESCRIPTION

Two kinds of features are used in this paper for image content description. For the purpose of analysis, both color and texture descriptors are applied. Image color distribution is represented by a global color histogram. On the other hand, image texture is described with a wavelet texture feature. The procedure of image feature extraction is briefly explained.

### A. Global color histogram

A global color histogram based descriptor can be considered as a very important image feature. The presence of an animation has a strong influence on the histogram, where an animation is characterized by a specific color distribution. The descriptor is extracted from video frames in HSV (hue-saturation-value) color space, with H:S:V=18:3:3 quantization per channels. The color descriptor dimension is 162 coordinates (18x3x3 coordinates).

### B. Wavelet Texture Grid Descriptor

A wavelet texture grid descriptor is used since an animation has a different distribution of edge elements from real content. Wavelets are considered very useful for describing combined textures, as well as for image differentiation based on the amount of details. Such a descriptor is extracted locally by dividing an image into 4x4=16 regions. The Haar wavelet transform using the fourth decomposition level is applied to each region. The components of the feature are obtained from each region, and they represent the variances of the high frequency sub-bands (12 sub-bands). The descriptor dimension is 192 coordinates (16x12 coordinates).

## IV. ANIMATION FRAME ANALYSIS

The animation frame analysis is realized in four steps. The block diagram of the animation frame analysis procedure is presented in Fig. 1.

In the first step frames are extracted from a video sequence. The color and the texture image features are extracted from the frames in step two. MA is applied to the obtained feature vectors in step three. In the fourth step the analysis of multifractal singularity spectra is performed.

Decoding phase in the process of analysis is an important part of procedure because the errors in decoding may produce additional graphic elements in the frame structure. Thus, this can affect the final results. Any type of video format can be used for the analysis. Images, i.e.

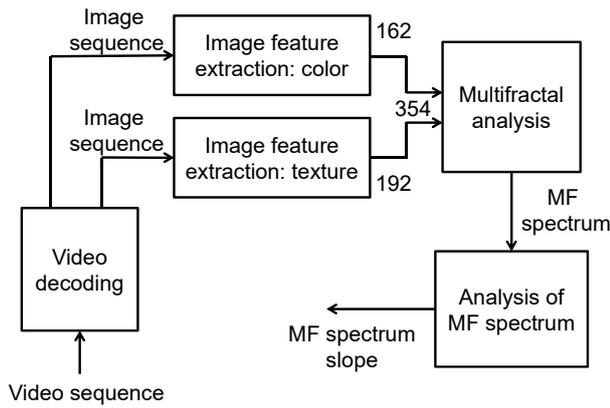frames, are extracted from a video sequence in JPEG image format [17].



Fig. 1 Block diagram of the frame analysis

The features are extracted separately from each frame. After the extraction, features are concatenated in a feature vector with a dimension of 354 coordinates (162+192 coordinates). The feature vector is calculated for each frame. The feature vectors are then stored in a feature matrix, which is normalized by columns. In the procedure each column is normalized with a maximum of column values. Normalization of a feature matrix provides an adequate preprocessing for the clusterization of feature vectors [18].

MA performs the calculation of multifractal singularity spectrum for each image/frame feature vector. MA method based on the Legendre transform [19] and box-counting algorithm [20], is applied for the multifractal spectrum calculation. Parameters for the spectrum calculation are extracted for the real frames used in the simulation. A typical singularity spectrum is shown in Fig. 2. Analysis of a multifractal singularity spectrum is based on the slope calculation on the right side of the spectrum curve as in [16]. In Fig. 2 the slope descriptor is illustrated with dashed lines.
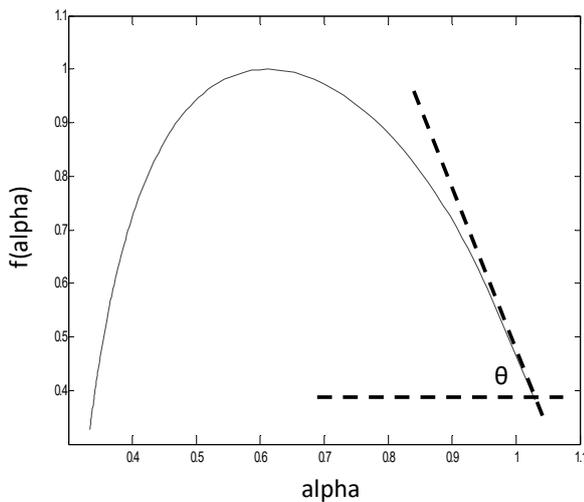


Fig. 2 Typical multifractal singularity spectrum calculated using the Legendre transformation with ending slope θ (dashed lines)

Long-range dependency of a measured structure is shown on the right side of a singularity spectrum. Therefore, in this area the differences between the animated and the real frames are expected. We assume that a specific color and texture distribution of animation frame has a low slope θ value. Real frames have a rustic color and texture distribution in comparison to animated frames and this should affect the θ value.

## V. EXPERIMENTAL RESULTS

For the purpose of the experiments three video sequences are used: real movie "Good Year" (filmed in 2006, director Ridley Scott), partially animated movie "Who Framed Roger Rabbit" (filmed in 1988, director Robert Zemeckis) and fully animated movie "Little Mermaid" (filmed in 1989, directors Ron Clements and John Musker). The examples of real movie frames (without animation) and fully animated frames used in the experiments are shown in Fig. 3. The examples of real movie frames (without animation) and partially animated frames used in the simulation are shown in Fig. 4. The frames presented in Fig. 4 are taken from the selected partially animated movie sequence (movie "Who Framed Roger Rabbit").



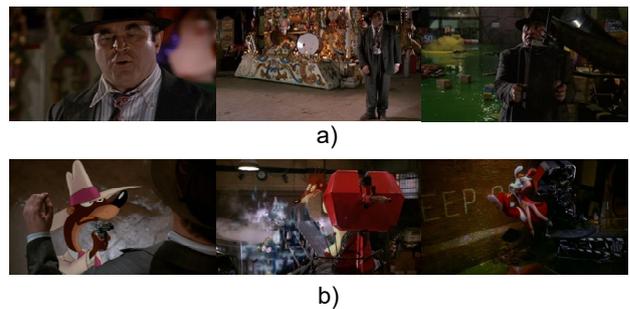Fig. 3 Examples of: (a) real movie frames and (b) fully animated frames



Fig. 4 Examples of: (a) real movie frames and (b) partially animated frames from the selected movie sequence

For a real and fully animated movie ten frames are picked randomly from each movie. Analysis is performed for the selected frames. The multifractal singularity spectra for real and fully animated frames are shown in Fig. 5(a). It is noticeable that singularity spectra for real and animated frames have different shapes. A few spectra calculated for real frames have the same shape as the animated frames, but their slopes are greater than in the case of the animated ones. Two spectra for real frames are overlayered by other real spectra in Fig. 5(a).

The results of MA for the real and the partially animated frames are presented in Fig. 5(b). Singularity

spectra have different forms for these two frame groups. The partially animated frames are very similar on the right side of spectra. They are also quite different from the fully animated frames as presented.

For each analyzed singularity spectrum a slope in the ending point on the right side of spectrum is calculated. The values of slope are presented in Fig. 6. As it is shown in the figure, animated frames (fully or partially) have a generally low slope value, where real frames have considerably higher slope values.

In the process of calculation, a slope angle is calculated as a positive value. As it is shown in Fig. 6, a few slope values of fully animated frames are close to the slope values belonging to the analyzed real frame. Generally, there is a big difference between the slope values of real and animated frames. This distinction obtained in the experiment could provide a good framework for the classification between real and animated frames.
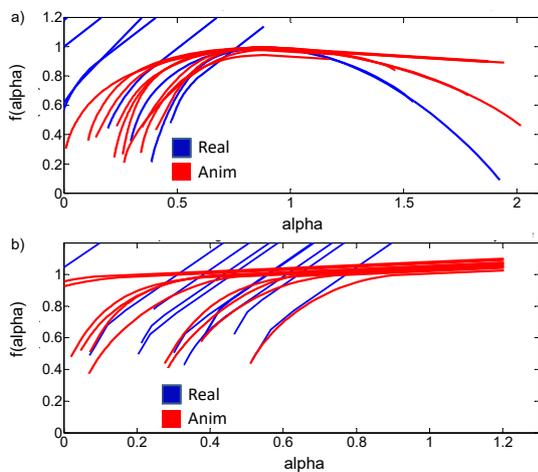


Fig. 5.Multifractal singularity spectra of: (a) real and fully animated movie frames, and (b) real and partially animated frames from the selected movie sequence
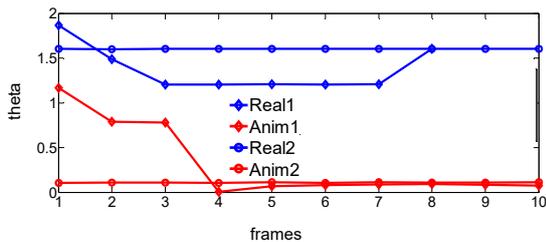


Fig. 6 Descriptor value comparison of: (a) real and fully animated movie frames, and (b) real and partially animated frames from the selected movie sequence

In the second experiment a local analysis is performed. A frame is taken from the group of partially animated frames (movie sequence "Who Framed Roger Rabbit"). This frame is divided into 4x4 regions, as presented in Fig. 7.

For each of the presented regions the methodology from the first experiment is applied. The feature vector is calculated similarly to the global approach, where the feature matrix is constructed here without normalization. The normalization is excluded because of the small number of regions. In Fig. 7 the regions are labeled with numbers 1-16.
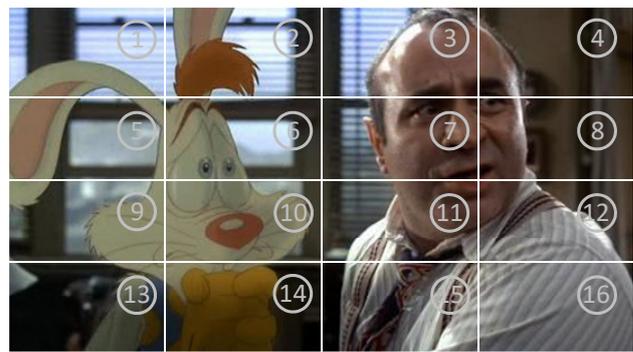


Fig. 7 A local analysis of a frame based on the partition (4x4 regions)

For each region, the parameter θ is calculated based on the multifractal spectrum. The obtained values are presented in Fig. 8, where the regions are classified into two groups: the regions with animation (Anim) and the regions without animation (Real). In the Anim group there are eight regions labeled by [1, 2, 5, 6, 9, 10, 13, 14], where the rest of them are in the Real group ([3, 4, 7, 8, 11, 12, 15, 16]).
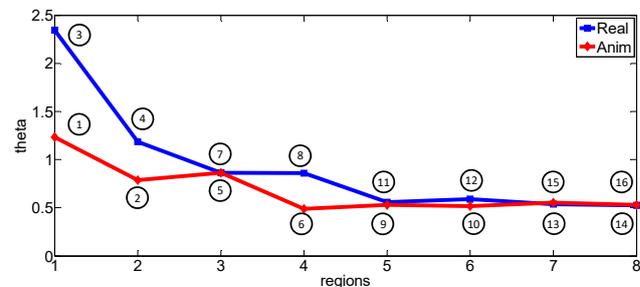


Fig. 8 The obtained parameter values for different regions from a partially animated frame

It can be noticed that several parameter values are higher. Namely, the higher parameter values are obtained for regions 3, 4, 7 and 8 in comparison to the other regions from the Real group. The difference is found because of the coherent texture found in regions 11, 12, 15 and 16. Such texture is found in Area 1 and Area 2 marked in Fig. 9. Additionally, the multifractality in the regions is affected by the animation adjustment to the real frame in the Area 1. The frame processing applied for avoiding abrupt changes/edges in animation-real part transition affects the texture structure.
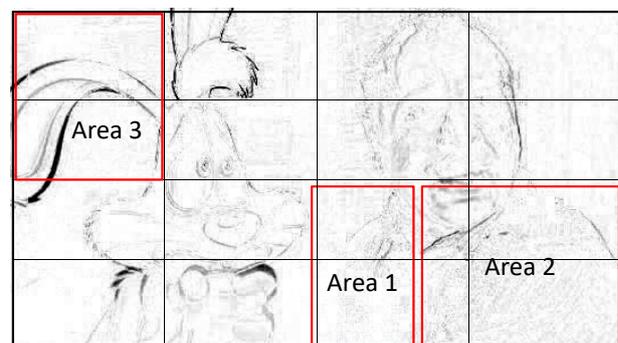


Fig. 9 The specific areas found within the frame

The specific multifractal behavior is noticed in Area 3 as a consequence of the background having different distribution characterization from the object. In this area, regions 1 and 5 have a greater amount of details related to texture in comparison to the other animated regions. Fig. 9 is obtained as a negative of the module of the difference between the analyzed and the following frame, multiplied by a factor of three. Also, noise contamination can be noticed here, as a video decoding error. Noise has a significant influence on the multifractality of the regions.

The presented results are preliminary. Extensive research is still to be done on a larger database of frames, to prove the credibility of results. Nevertheless, the presented frame analysis and the obtained results are very promising. They are promising in the simplification of the detection procedure for commercials, cartoons and other animated structures in video sequences.

## VI. CONCLUSION

In this paper the analysis of video frames with animated structure is performed. For the analysis, the color and texture based features are extracted from randomly picked frames in video sequences. Image features of the frames are processed by a multifractal analysis. The slope of multifractal singularity spectrum is calculated for each feature vector. Frame slope values are compared, and the comparison results show that in the calculated slope domain it is possible to successively differentiate the real frames from the animated ones. Future work will be oriented towards extending the number of singularity spectrum parameters used for frame analysis and testing the procedure using a larger frame database.

## REFERENCES

[1] D. Brezeale, D.J. Cook, "Automatic Video Classification: A Survey of the Literature," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(3), pp. 416- 430, 2008.

[2] G. Forman, "An Extensive Empirical Study of Feature Selection Metrics for Text Classification," *Journal of Machine Learning Research*, 3, pp. 1289-1305, 2003.

[3] Z. Rasheed, M. Shah, "Movie Genre Classification by Exploiting Audio-Visual Features of Previews," *IEEE Int. Conf. on Pattern Recognition*, 2, pp. 1086-1089, 2002.

[4] J. Fan, H. Luo, J. Xiao, L. Wu, "Semantic Video Classification and Feature Subset Selection under Context and Concept Uncertainty," *ACM/IEEE Conference on Digital Libraries*, pp. 192-201, 2004.

[5] R. Lienhart, "Reliable Transition Detection in Videos: A Survey and Practitioners Guide", *Int. Journal of Image and Graphics*, 1(3), pp. 469-486, 2001.

[6] L. A. Duarte, O. AB Penatti, J. Almeida. "Bag-of-Genres for Video Genre Retrieval." *arXiv preprint arXiv:1506.00051* (2015).

[7] G. Wei, L. Agnihotri, N. Dimitrova, "TV Program Classification based on Face and Text Processing," *IEEE Int. Conf. on Multimedia and Expo*, 3, pp. 1345-1348, 2000.

[8] A. Bovik, "The Essential Guide to Video Processing," Academic Press, ISBN 978-0-12-374456- 2, 2009.

[9] D. Brezeale, D.J. Cook, "Using Closed Captions and Visual Features to Classify Movies by Genre," *Int. Workshop on Multimedia Data Mining*, 2006. 35

[10] H. Wang, A. Divakaran, A. Vetro, S.-F. Chang, H. Sun, "Survey of Compressed-Domain Features used in Audio-Visual Indexing and Analysis," *Journal of Visual Communication and Image Representation*, 14(2), pp. 150-183, 2003.

[11] B. Ionescu, K. Seyerlehner, C. Rasche, C. Vertan, P. Lambert. "Video genre categorization and representation using audio-visual information." *Journal of Electronic Imaging* 21, no. 2 (2012): 023017-1.

[12] , H. K. Ekenel, T. Semela. "Multimodal genre classification of TV programs and YouTube videos." *Multimedia tools and applications* 63, no. 2 (2013): 547-567.

[13] A. AM Hamed,, R. Li, Z. Xiaoming, C. Xu. "Video genre classification using weighted kernel logistic regression." *Advances in Multimedia* 2013 (2013): 2.

[14] B B Mandelbrot, "The fractal geometry of nature", New York, W. H.Freeman and company, 1977.

[15] B A Chhabra, C Meneveau, V R Jensen, R K Sreenivasan, "Direct determination of the f(a) singularity spectrum and its application to fully developed turbulence", *Physical Review A,* Vol. 40, No. 9, Nov. 1, 1989.

[16] A.Gavrovska, G. Zajić, I. Reljin, and B. Reljin, "Classification of Prolapsed Mitral Valve versus Healthy Heart from Phonocardiograms by Multifractal Analysis", *Computational and Mathematical Methods in Medicine*, Vol. 2013, Article ID 376152, 10 pages, 2013.

[17] Gregory K. Wallace,"The JPEG still picture compression standard."*Communications of the ACM* 34, no. 4 (1991): 30-44.

[18] G. Zajić, N. Kojić, V. Radosavljević, M. Rudinac, S. Rudinac, N. Reljin, I. Reljin, B. Reljin, "Accelerating of Image Retrieval in CBIR System with Relevance Feedback", *EURASIP Journal on Advances in Signal Processing, Spec. Issue on Knowledge Assisted Media Analysis for Interactive Multimedia Applications*, 2007:1-13, Article ID 62678, 2007.

[19] R. Lopes, N. Betrouni, "Fractal and multifractal analysis: a review." *Medical Image Analysis*, 13(4):634-649, 2009.

[20] FRACLAB: A fractal analysis toolbox for signal and image processing, http://fraclab.saclay.inria.fr/, 2015.

[21] G. J. Zajić, M. D. Vesić, A. M. Gavrovska and I. S. Reljin, "Animation frame analysis," *Telecommunications Forum Telfor (TELFOR), 2015 23rd*, Belgrade, 2015, pp. 732-735.