Asian Journal of Convergence in Technology **Volume V Issue I**
ISSN NO: 2350-1146 I.F-5.11

Mood Detection through Aesthetic Assessment of Videos using Deep Learning

Madhura Phatak

Assistant Professor
School of ComputerScience and Engineering
Dr.Vishwanath Karad MIT World Peace University
Pune,India
Madhura.pathak@mitpune.edu.in

Shruti Vinit Asarkar
School of Computer Science and Engineering
Dr.Vishwanath Karad MIT World Peace University
Pune,India
asarkarshruti@gmail.com

***Abstract*— In human interaction, emotion detection plays an effective role. Monitoring and predicting various human's feelings (happy, sad, anger, fear, etc.) is a challenging task. Human interaction and carrier of feelings amongst humans are accomplished mainly through five senses: touch, smell, taste audio, visual. Considering Visual sense, images and videos are important gradients in day-to-day life. It can elevate/ depress the mood of a person. Digital contents of multimedia are image, audio, video, text, and so on. The usage of internet is tremendously increasing, so Internet bandwidth and storage space, video data has been generated, published, and spread robustly, and becoming an important of today's big data. This has encouraged the development of advanced techniques for a wide scope of video understanding applications including online advertising, Cinematography, video retrieval, video surveillance, video data on Social sites, etc. However, it is easy to convey a story to a viewer of video, since a video is worth of thousands worth. And this story actually creates a mood. This work is to detect the mood of aesthetically pleasing videos that reflect on a person's mood.**

Keywords*— *Video aesthetics, Deep Learning, DCNN

I. Introduction

In recent years, image and video aesthetics assessment have grabbed much attention. Evaluation methods of image aesthetics massively depend on the fruitful aesthetic feature. The traditional method extracts hand-crafted features for aesthetic evaluation. But nowadays, research is more focused on Deep Convolutional Neural Network for more accurate aesthetic assessment.

Mood detection is an essential component. Mood detection applications can be found in different domains. In this study of mood detection, we will detect the mood of videos and images with an aesthetic assessment of video. This assessment of videos will classify videos into pleasing and non-pleasing videos. Pleasing videos will helps to elevate the mood of the person. We are going to present a computational approach to evaluate Video Aesthetics.

A video is nothing but gathering images altogether. According to the Oxford dictionary, "Video is the recording, reproducing, or broadcasting of moving visual images." Images having some visual features. Visual features are those which are directly affected a person's visual perception. Low-Level features (Color of an image, Texture, Intensity Edges), Middle-Level features (Objects), High-Level feature (Rules of photography, Conceptual rules (eg, Aestheticism)). There are few rules of photography including, Depth of Field (Main focus on subject and background is blurred), Color Contrast (Also known to be Opposing colors), Rule of Thirds (Divide image into equal parts such that, they reveal primary composition elements near the intersection of the line).

It is a very tough challenge to evaluate video aesthetics. But this aesthetic assessment of videos is useful in many cultures, including cinematography, to show beautiful high content, revenue generation in the advertisement world. To improve user satisfaction in many applications such as search & recommendation, video aesthetics can be used as a useful idea.

A. Image Aesthetics

Image aesthetics assessment is a try to define the "beauty" of an image. Beauty is defined to those, qualities that give satisfaction to the human senses that provides an intuitive experience of pleasure, meaning or satisfaction. Image aesthetics is a subjective field. An arrangement of a composition of the image, that affects the visual senses of the viewer. Image aesthetics is an approach to, what is pleasing and what is non-pleasing to the human senses--especially here we are considering visual sense.

B. Significance of Image Aesthetics

For platforms such as media content, that shows beautiful high-quality content of images and videos. With Social websites and focusing on the 'selfie' trend, where we people generating a huge amount of images as well as video data. Websites like Facebook, Instagram, Hike where appreciation is based on people's "Likes" and this "Likes" belongs to Aesthetically Pleasing Images and video contents. For revenue generation in Advertisement world image as well as video aesthetics is applicable.

C. Approaches for Image Aesthetics Assessment

An image aesthetic may rely on both the scenes and depicted objects. Developing a model to provide automatic photo and video quality or say aestheticity assessment. There have been many pieces of research and studies focusing on the estimation of image aesthetics. The aim is to classify images automatically into pleasing and non-pleasing images. The very first approach is that defines a set of image features that they assume to affect the aesthetic quality of photographs, and then design some mathematical models to extract them. This Hand-crafted features, however, is not only difficult but also does not enough to account for the full and composite nature of image aesthetics. Therefore, it could lead to incorrect assessments.

Another significant approach is to automate the process of picking and modeling image features as well as estimating the aesthetic image involves deep learning. Particularly, as compared to a conventional approach, by using the Deep Convolutional Neural Networks (DCNN) that have been trained on a large-scale image database to get more specific photo quality assessment. The set techniques based on self-operating feature modeling generally perform well in terms of concluding whether the given image is aesthetically pleasing or not [6].

D. Video Aesthetics

Video aesthetics is extracting aestheticity from Video. Video aesthetic is very useful for improving user gratification in many applications like search and recommendations. Aesthetic assessment of videos is useful in many applications, including Cinematography, UI Design, Advertisement world, and Social Media Websites too [14]. Existing research has mostly concerned about constructing hand-crafted features for estimation of video aesthetics. Here, we present an operational approach to evaluating video aesthetics. In this framework, we will adopt Deep Learning to predict aesthetic quality. In this case, a thousand features are evaluated.

II. LITERATURE SURVEY

Following is a summary of existing work.

In past years, the volume of images and videos grows explosively through a social network. So, recently Aesthetic assessment of images as well as videos has caught a lot of attention of researchers. Aesthetic assessment of videos can help people to pick out or filtrate beautiful or say pleasing videos from the crowd. Aesthetic evaluation is an abstract field. Aesthetic evaluation of videos includes some visual features. Visual features are those which are directly affecting to a person's visual appreciation. This visual feature includes color, texture, shape, pixel-level feature, edge detection. Various properties or attributes of an image affect on the photo, like Low-level feature, Middle-level feature, High-level feature. Low-level features are a slight description of an image including color, texture, intensity, edges, etc. In [1] for aesthetic evaluation, there is a use of a Multi-scene Deep Learning Model (MSDLM). This model includes eight-layer deep convolutional neural network(DCNN), from which five layers are convolutional, and other three layers are fully connected layers. Here a novel deep neural network and pre-training strategy is adopted. Due to use of DCNN approach, the accuracy of image aesthetic evaluation is increased. In this work, AVA(Aesthetic Visual Analysis) and CHUKPQ are used. AVA dataset consists 255,000 images[3]. There is a lot of data noise in the datasets (AVA and CUHKPQ). A lot of data noise is reduced in this from this datasets.

As the number of layers is increased, complexity is increased. In [2], to reduce the complexity of DCNN, Global Average Pooling (GAP) is used. Here two approaches are used, in which 1) fine-tunes a standard CNN with a newly discovered GAP layer. 2) with individual GAP operations, reduce the dimensionality of convolution layers, and then extract global and local CNN codes. An experiment shows the comparable accuracy results within different methods. The accuracy of the GAP is 76.32%, which is maximum than other methods. The complexity of training and testing is substantially reduced. Also, GAP layer is used in producing CAMs(Class Activation Maps) for visualizing location in photographs. This will contribute towards an aesthetic quality of photos. Image aesthetics may depend on objects as well as scenes. To forecast image aesthetics scores in [3], Deep Convolutional Neural Network is used. By directing the image aesthetics prediction as a regression problem, make improvements in an authorized CNN architecture, to classify both targets or objects and scene.

Aesthetic assessment can be done with handcrafted features extraction as well as with Deep Learning. To predict the aesthetic image descriptor, earlier algorithms use a wide variety of techniques such as SVM, Neural Network or Random forest [4]. To assess image aesthetics, study and compare the selection of algorithm that used handcrafted features. By combining all of the features together and attain performance close to that of the model trained on learned CNN features, we can achieve an additional improvement in aesthetic prediction accuracy. For various tasks such as classification, regression, categories, feature elimination is performed. [5] uses the architecture of DCM (Deep Chatterjee's Machine) for image aesthetics assessment that leads to superior performance. By using DCM images are classified into a high-quality image and low-quality image. For Binary rating prediction, accuracy is compared with a different method. For low-quality images, accuracy by using DCM is correctly identified with accuracy result as 76.80% and for a high-quality image, accuracy is 76.04%. Hence DCM leads to superior performance compared to another state-of-art model. [6] focuses on Deep Neural Network approach. This approach is used to estimate image aesthetics. The focus of the paper is more on automatic feature learning. To absorb effective aesthetic features, convolutional neural network(CNN) is used. However, image aesthetics depends on a combination of local views(for example sharpness and noise levels of an image) and global views(i.e. rule of thirds). To take parallel inputs from two columns, the double-column neural network architecture is developed. In this double column neural network, one of the columns takes a global view and another column takes the local view of the image. These two columns further aggregated after some layers of transformation and then mapped to the label layer. This double column approach is applied to the generic image aesthetic problem. For content-based image aesthetics, a network adaption approach is proposed. For network related approach, attributes associated with images such as a style of an image and semantic attributes are explored. By leveraging these two style and semantic attribute respectively, performance is boosted. The overall result shows DCNN(Double Column Neural Network) achieved more accuracy (72.9%) as compared to SCNN(Single Column Neural Network) for global input as well as local input. [7] aims to improve the accuracy of photo quality assessment. The main focus is on the classification of an image into high quality and low-quality image, aesthetic attributes identification, automatically extract the high-level features, using pre-trained DCNN. In [8], low level and high-level features are combined in this paper. Humans usually perceive only a few silent regions in the photo. To integrate low level and high-level visual signals sparsity-constrained graphlet ranking algorithm is used. For silent graphlets discovery, this framework integrates multiple visual/ semantic features. For duplicating a human gaze shifting path, these discovered paths are connected into AVP(Actively Viewing Path). To learn the distribution of AVP from aesthetically pleasing training photos, GMM is employed.

[9] aims to study the properties of videos from the perspective of aesthetics. For videos, a variety of features are designed. Furthermore, the performance of these features in the application of video detection is examined. The experiment in this paper shows, this set of feature can be used to achieve high professional and amateur video classification rate. For professional video classification and amateur video classification, the accuracy rate is 97.3% and professional video detection rate is 91.2%. This paper is more focused on video features, such as camera motion and shot length. In [10] authors designed some low-level features for a visual aesthetic impression. They observed that this low-level feature affects the aesthetics of videos. With low-level features, authors are using some high- level features. The high-level feature is those which includes Rules of Photography, Conceptual rules. This high-level feature is also affecting the aesthetics of videos. Generic architecture that explains in the paper is shown below:

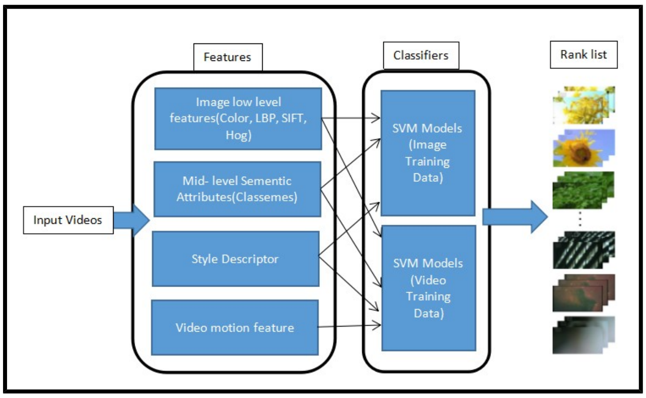

Figure 1 : Generic architecture for Video Aesthetics

As shown in the above figure, low-level image feature, mid-level semantic attributes, style descriptor, Video motion feature are extracted. To classify images and videos, SVM is used. Resultant will shows the ranking of image and videos according to aesthetic pleasingness. Similar to this summary of a literature survey is shown in the following table I.

Table I : Literature Survey

Sr. No	Papers search (Year of publication)	Authors	Strong Points	Accuracy
[11]	Video Aesthetic quality assessment by combining semantically independent and dependent features (2011)	Chun-Yu Yang, Hsin-Ho Yeh and Chu-Song Chen	1. Assessing the Video aesthetic quality, for that combine semantically independent and dependent features. 2. Semantically independent features include Motion Space(MS), Hand Shaking(HS), Color Harmonic, Composition. Semantically dependent features include Motion Direction Entropy(MDE), Color Saturation and Value, Lightness. 3. A result of this experiment shows that, in order to distinguish the assessment performance between semantically independent features and semantically dependent features, it can be seen that semantically independent features exceed the semantically dependent features in terms of aesthetic quality assessment.A result of this experiment shows that, in order to distinguish the assessment performance between semantically independent features and semantically dependent features, it can be seen that semantically independent features exceed the semantically dependent features in terms of aesthetic quality assessment.	Sementic-dependent: 69 $\pm$ 2.2%, Semantic-independen : 74 $\pm$ 1.5%
[12]	Video Aesthetic Quality Assessment by Temporal Integration of Photo- and Motion-Based Features. (2013)	Hsin-Ho Yeh, Chun-Yu Yang, Ming-Sui Lee, & Chu-Song Chen	1. For accessing the aesthetic quality of videos, a new method is presented in this paper. 2. Two processes are there, namely- for extracting the aesthetic features from for each frame in a video, this first method combines both photo-based and motion-based visual clues. 3. A temporal-order-aware framework integrates frame-based features. This increases the evaluation accuracy. 4. Remarkable accuracy difference is the result of this paper.	On dataset, Telefonica- 81.5 $\pm$ 1.9% And on ADCC- 81.1 $\pm$ 1.3%
[13]	Towards a Comprehensive Computational Model for Aesthetic Assessment of Videos. (2013)	Subhabrata Bhattacharya, Behnaz Nojavanasghari, Tao Chen	1. Novel Aesthetic model is proposed, in which psycho-visual statistics are extracted from multiple levels. 2. For evaluating the beauty of broadcast quality videos, this novel model is proposed. 3. Each video is split into shots, and for each shot, uniformly key-frames are selected. 4. Features are selected at three levels, namely cell, frame and shot.	
[14]	A Novel Feature Set for Video Emotion Recognition. (2018)	Shasha Moa, Jianwei Niua, Yiming Sua, Sajal K. Das	1. In this paper, Social media analysis is done. 2. Given approach in this paper is for effective feature extraction in Video affecting recognition system is built. 3. Performance of emotion recognition is improved.	

[15]	Video Aesthetic Quality Assessment using Kernel Support Vector Machine with Isotropic Gaussian Sample Uncertainty (KSVM-IGSU). (2016)	Christos Tzelepis, Eftichia Mavridaki, Vasileios Mezaris, Ioannis Patras.	1. The kernel SVM with Isotropic Gaussian Sample Uncertainty (KSVM-iGSU), is an extension of the standard kernel SVM that exploits the uncertainty of input data in order to achieve better classification results. 2. Video aesthetic quality assessment method combines the representation of each video according to a set of photographic and cinematographic rules, with the use of a learning method that takes the video representation's uncertainty into consideration.	0.6814

Thus as given in Table I, related work in the area of video aesthetics involves including video and approaches used include machine learning algorithms ,deep learning methods and DCNN.

A. Research Gaps.

Earlier work concentrated on extracting the handcrafted image and video characteristics or generic image descriptors to construct a statistical model for aesthetic assessment. Although, the efficiency of these approaches is restricted by researchers' understanding of the aesthetic rules. Nowadays, researchers start to apply Deep Learning strategy in the aesthetic assessment of images as well as videos.

Till now work on image aesthetic assessment is done, some researchers use handcrafted feature extraction technique or some prefer Deep Learning. But there have been few works on aesthetic assessment using video aesthetics by using Deep Learning. Almost all studies have been using, handcrafted feature extraction for aesthetic assessment of videos, which is not really accurate to rely on.

Thus there is a need for mood detection through aesthetic assessment of videos using Deep Learning. Also here we are considering high-level features. Since the high-level features are unique for aesthetic sense.

III. Proposed System Architecture

The proposed system overview consists of the following steps as shown in Figure 2. In our proposed work we present a computational approach to evaluate video aesthetics by using Deep Learning approach. As per the paper referred above [15], the first step is to split the video into one long video and ended up having a few common frames at the start or end of each video. Or each video is divided into its shots using the shot detection method. Then, for each video, estimate the mean duration of its shots, and, considering that the shot transitions can be either abrupt or gradual, we estimate for each of these transition types their span as a percentage of the whole video's span. This results in a 3-element video-level vector. In our proposed work, we are using Video key-frame extraction. Subsequently, one key-frame per second is taken out from the original raw video sequence (regardless of shot boundaries), and photo- and motion-based characteristics are taken out for each one of them. Characteristics of photos include the simplicity, colorfulness, sharpness, and pattern and overall aesthetic quality values, which are taken out, rely on the still-image aesthetic quality evaluation. Motion-based features include:

a) compute likeness between consecutive frames (cross-correlation between these frames),
b) compute the variety of motion directions(motion direction entropy),
c) compute the steadiness of the camera during the capturing process (hand-shaking), and
d) A measure which can determine the difference between three divisions of shots: focused shots, panorama shots and static shots (shooting type).

Extracting these above features from images as well as video and used them to construct a classifier can help to pick delightful images from unappealing.

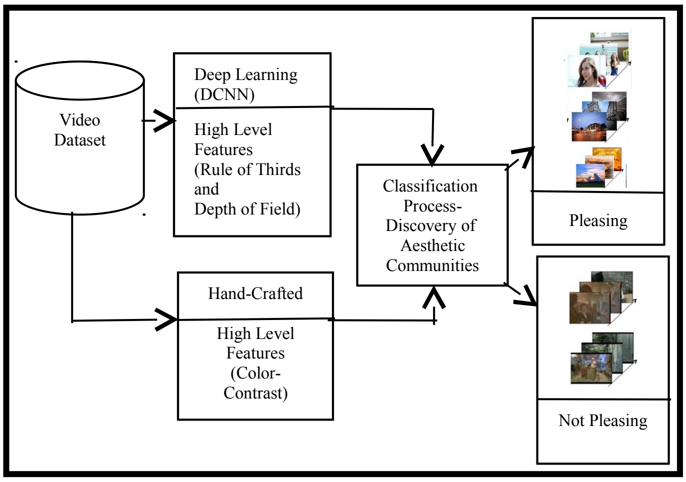

Figure 2: Proposed system Generic Architecture for video aesthetics

A. Deep Learning:

A Deep Convolutional Neural Network is trained to learn aesthetic features automatically. The DCNN can concurrently assess the aesthetic quality of a video. As automated feature extraction makes deep learning models, highly accurate. For computer vision tasks such as object classification, these deep learning models are used. DCNN approach enhances the accuracy of Aesthetic assessment. DCNN is used to automatically learn features for aesthetic quality categorization.

B. Hancrafted Technique:

In computer vision applications, such as object detection and image classification, these handcrafted features have been used for more than a decade. By using the information present in the image itself, "Hand Crafted" features refers to properties obtained using various algorithms. For example, two simple features that can be taken out from images are edges and corners. A basic edge detector algorithm works by finding areas where the image intensity "suddenly" changes. For identifying object and computer vision, "handcrafted features" were commonly used with "traditional" machine learning approaches like Support Vector Machines. This technique is used to extract features of color contrast images.

C. Classification Process:

In this phase, we will report the results for the different classification methods employed. It will discriminate between pleasing and non-pleasing videos. A challenge in the video aesthetic quality assessment problem, similarly to many video classification tasks, is that video representation techniques usually introduce uncertainty in the input that is subsequently fed to the classifiers. The basic metric for classification problems is to measure accuracy. Thus during classifier training, unpredictability needs to be taken into consideration.

The resultant stage is, the video is classified into pleasing and non-pleasing. That is it detects the mood of an image. Does really that image is pleasing or not? This aesthetically pleasing videos will reflect the person's mood.

IV. Algorithmic Approach scope

A. Video key-frame extraction:

The video is nothing but a set of static keyframes (motionless images) which maintains the overall content of a video with minimum data. In our proposed work, we are extracting Video key-frame. For that, we are using dynamic Delaunay graph clustering via edge pruning strategy [17]. Basics of K-means Clustering is used, in the case of Delaunay graph clustering.

B. Color extraction:

For extracting color from the image, SIFT is used [10]. There are few tools also available for Automatic Color Extraction. Color scheme Extraction is one of the tools used for extracting color from an image. This is an open source tool, which uses Scikit learn the framework.

C. Segregation of Object from Background:

For segregating object from the background, SVM is used. Object-based segregation is based upon an object identification model.

V. Future Scope

Our distant Goal is to use Human Aura Scanning to enhance the mood. Aura Scanning to detect what is the impact of pictures or videos on human's mood, so the mood is enhanced. In this work, we are choosing a similar type of video content, which is of common resolution, uniform background, and a single object. In future work, we can include different video clipping types, like multiple objects.

VI. Conclusion

In this study, we have reviewed state-of-the-art deep learning techniques for video aesthetic assessment. In contrast to handcrafted features that are costly to design and have limited generalization capability, the essence of deep learning for video classification is to derive robust and discriminate feature representations from raw data through exploiting massive videos with an aim to achieve effective and efficient classification, which could hence serve as a fundamental component in video aesthetic assessment. This approach of assessment of videos is focusing on accuracy. In case of Deep Learning, accuracy is more than that of Handcrafted Feature Extraction mechanism. In Deep Learning our focus is on DCNN, for accuracy in the aesthetic assessment result. Here I am going to use, unique feature combination of color contrast (handcrafted extraction) and high-level features of Rule of Thirds and Depth of Field for Mood Detection through aesthetic assessment of Videos using Deep Learning.

References

[1] Weining Wang, Mingquan Zhao, Li Wang, Jiexiong Huang, Chengjia Cai, Xiangmin Xu, "A multi-scene deep learning model for image aesthetic evaluation", 2016, ACM journal Image communication, Vol.47, Issue C, Pages 511-518.

[2] Magzhan Kairanbay, John See, Lai-Kuan Wong, Yong-Lian Hii, "Filling the GAPS: Reducing the Complexity of Networks for Multi-Attribute Image Aesthetic Prediction." 2017 IEEE International Conference on Image Processing.

[3] Simone Bianco, Luigi Celona, Paolo Napoletano, and Raimondo Schettini, "Predicting Image Aesthetics with Deep Learning", 2018 Springer international Journal.

[4] Michal Kucer, Alexander C. Loui, David W. Messinger, "Leveraging Expert Feature Knowledge for Predicting Image Aesthetics", 2018 IEEE Transactions on Image Processing, Vol. 27, No. 10.

[5] Zhangyang Wang, Ding Liu, Shiyu Chang, Florin Dolcos, Diane Beck, Thomas Huang, "Image Aesthetics Assessment using Deep Chatterjee's Machine" 2017 IEEE International Joint Conference on Neural Networks(IJCNN).

[6] Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, and James. Z. Wang, "Rating Image Aesthetics Using Deep Learning", 2015 IEEE Transactions on Multimedia, Vol. 17, No. 11.

[7] Hui-Jin Lee, Ki-Sang Hong, Henry Kang, and Seungyong Lee, "Photo Aesthetics Analysis via DCNN Feature

Encoding", 2017 IEEE Transactions on Multimedia, Vol. 19, No. 8.

[8] Luming Zhangy, Yue Gaoy, Chao Zhangz, Hanwang Zhangy, Qi Tian, Roger Zimmermann, "Perception-Guided Multimodal Feature Fusion for Photo Aesthetics Assessment", 2014, Proceeding of the 22nd ACM international conference on Multimedia Pages 237-246.

[9] Yuzhen Niu and Feng Liu, "What Makes a Professional Video? A Computational Aesthetics Approach", July 2012 IEEE Transactions on Circuits and Systems for Video Technology, Vol. 22, issue:7

[10] Yanran Wang, Qi Dai, Rui Feng, Yu-Gang Jiang, "Beauty is Here: Evaluating Aesthetics in Videos Using Multimodal Features and Free Training Data", 2013, Proceeding of the 21st ACM international conference on Multimedia Pages 369-372.

[11] Chun-Yu Yang, Hsin-Ho Yeh and Chu-Song Chen, "Video Aesthetic quality assessment by combining semantically independent and dependent features", 2011 IEEE International conference on Acoustics, Speech, and Signal Processing(ICASSP).

[12] Hsin-Ho Yeh, Chun-Yu Yang, Ming-Sui Lee, and Chu-Song Chen, "Video Aesthetic Quality Assessment by Temporal Integration of Photo- and Motion-Based Features", Dec 2013 IEEE Transactions on Multimedia, Vol. 15, issue: 8

[13] Subhabrata Bhattacharya, Behnaz Nojavanasghari, Tao Chen, "Towards a Comprehensive Computational Model for Aesthetic Assessment of Videos", 2013, Proceeding of the 21st ACM international conference on Multimedia Pages 361-364.

[14] Shasha Moa, Jianwei Niua, Yiming Sua, Sajal K. Das, "A Novel Feature Set for Video Emotion Recognition", 2018 Neurocomputing 291, 11-20.

[15] Christos Tzelepis, Eftichia Mavridaki, Vasileios Mezaris, Ioannis Patras, "Video Aesthetic Quality Assessment using Kernel Support Vector Machine with Isotropic Gaussian Sample Uncertainty (KSVM-IGSU)", 2016 IEEE International Conference on Image Processing (ICIP).

[16] Loris Nanni, Stefano Ghidoni, Sheryl Brahnam, "Handcrafted vs Non-Handcrafted Features for computer vision classification", 2017 Pattern Recognition 71, 158-172.

[17] Sanjay K. Kuanar, Rameswar Panda, Ananda S. Chowdhury, "Video key frame extraction through dynamic Delaunay clustering with a structural constraint", 2013 journal of Visual Communication and Image Representation, Vol. 24, issue: 7, Pages 1212-1227.

[18] Ritendra Datta, James Z. Wang, "ACQUINE: Aesthetic Quality Inference Engine –Real-time Automatic Rating of Photo Aesthetics", Proceedings of the ACM international conference on Multimedia information retrieval Pages 421-424.