Classification of Video Media: The Aesthetics Way

Pritesh S Patel[1], Madhura V Phatak[2], Ruhi A Patankar[3]

Abstract— With the popularization of digital cameras and the rapid development of the Internet, number of images and videos taken and its broadcasting is growing explosively. The aesthetics evaluation of video media can be used as useful clue to improve user satisfaction in many applications like search; broadcasting and recommendation. To automatically assess the quality of videos with a strong correlation with human perception is a challenging task. There are two different challenges that exist in this task: First one is, there is no any publically benchmarked datasets available till date and second one is lack of standard video shooting techniques.

Index Terms— Aesthetics quality assessment, computational aesthetics, visual aesthetics, image aesthetics quality assessment, motion features.

I. INTRODUCTION

Aesthetics in the world of photography refers to the appreciation of beauty or art of nature [1,3]. To judge the beauty or aesthetics of images or videos is highly subjective task because people have different preference over it. Hence there is no standard for measuring the aesthetics value. For example, the average individual sees an image in different scenario while a professional photographer may look aesthetics of image in different perspective. So, this subjective rating is difficult in assessment.

However, some work came up with computational aesthetics [2,3,4] to solve this problem. They observed that there are many factors/features affecting the aesthetics of video media. So researchers tried to extract such features from images and videos.

Inspired by this [2,3,5] began to extract low-level features from images for visual aesthetic impression including color, shape, depth of human body, texture, brightness, hue etc and used it for aesthetics assessment. As opposed to this, the previous work have made number of attempts to computationally model the assessment of natural images in terms of aesthetic quality. These approaches have been dealing with the high-level semantic qualities of such photographic images considered as high quality photographs and, from their visual content, researchers tried to assess the aesthetic value. To design a predictor or classifier using machine learning techniques has gained much interest to automatically classify videos or images based on their aesthetics appeal [11].

Manuscript received December, 2014.
***Pritesh S Patel**, Computer Engineering Department. MAEER'S MIT, Pune,*
***Madhura V Phatak**, Computer Engineering Department, MAEER'S MIT, Pune,*
***Ruhi A Patankar**, Computer Engineering Department, MAEER'S MIT, Pune,*

Previous work [2,3,6,7,9,10] focuses on 2-class photo classification which dealt with well defined classes, such as photo/graphics, indoors/outdoors, city/landscape, or photo/painting, but the concept of aesthetic quality is largely subjective. Along with it, the unavailability of adequate data-sets and assessment metrics for evaluation of the developed solutions makes it difficult to compare the results achieved by different authors.

II. LITERATURE SURVEY

A. Image Aesthetics Quality Assessment (IAQ)

Photographic aesthetics evaluation began [6] with extracting simple low-level features such as color; texture and shape, these feature are extracted from image and used to build a classifier that could sort photographs according to low or high aesthetic appeal. Inspired by this [5] designed some low-level features for visual aesthetic impression including color, shape, motion, spatial layout, depth, and human body. They observed that these features affect the aesthetics of videos. So, extracting these features from images and used them to build a classifier can help to select high appealing photo from unappealing.

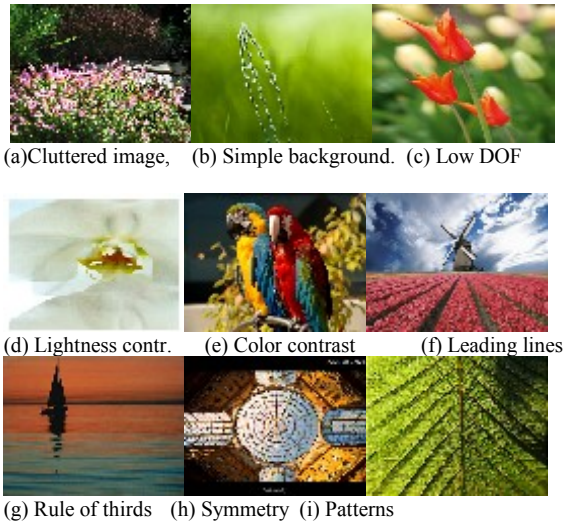
. As opposed to this, guidance from professionals helped to design more semantic or high level features. Professional photographers adopt some special techniques to make their photo more appealing. Using this as base study researchers proposed high –level features. These high-level features are the base of the photography compositional rules. On the basis of this [3] designed some aesthetics features e.g. color, light, exposure, and composition. Compositional rules including Rule of Thirds, which says that, divide the image into 9 equal parts using two equally spaced horizontal and vertical lines and photographers are enforced to place their subject of interest in one of the intersecting point. According to this rule, if the pictures follow this rules that would be appealing to eyes. (As shown in figure1)

Depth of Field i.e. difference between foreground and background, shape convexity i.e. some shapes are more appealing to eyes, region composition and light is also a good way to distinguish between low and high quality images . They formulated all these features mathematically using machine learning techniques and attempted to classify images and video media. [8,9,10] focused more on some other important high-level features i.e. Simplicity, Diagonal Dominance and Visual Balance. They observed that most of the high quality photo obeys simplicity [9].

According to this photographer keeps the subject on the focus and make the background blur or simple i.e. homogenous because in a photo or image subject attracts most of the visual attention.

Some work also says that apart from the thirds lines in Rule of Thirds, Diagonal Dominance [10] of the image also reflects the aesthetic impression, so putting subject of interest on the diagonal lines also makes the images appealing to the eyes. Visual Balance [10] states that in a visually balanced image, the visually salient objects are evenly distributed around center. However [9] observed that the entire previous work extracting feature from whole images, but what author observed that, in a high quality image subject gathers most of the visual attention. So, author attempted to extract features from subject only and they formulated all the aesthetics features on the subject only.

This work gave researcher a new direction to focus on another aspect of image from where they can extract features for aesthetics assessment. But their results severely depends on the method that extract subject of interest from image. Inspired by this work [10] attempted to extracted features from salient region of images instead of the subject. It is more convenient to extract features from salient region because it gives better scenario of the images. Again, the results of their work highly depend on the method that extracts salient regions from images. However, their work not focused to aesthetics classification but more on enhance of image i.e. converting a low-quality image into high quality. Their work was a revolution for many application and tools of image enhancement.



(a)Cluttered image, (b) Simple background. (c) Low DOF

(d) Lightness contr. (e) Color contrast (f) Leading lines

(g) Rule of thirds (h) Symmetry (i) Patterns

Figure1: Examples of some photographic principles

B. *Video Aesthetics Quality Assessment (VAQ)*

Unlike photo aesthetics there are number of rules exists from professionals for amateurs to follow, but for video there is no such rules induced from professionals. Learning from professional is a typical approach for aesthetics assessment [11]. However work from [11,12] says that their dataset can serve as a publically benchmarked dataset but there is some ground truth is needed.

As discussed above that there is a lot work [2,3,5,7,8] has been focused on image aesthetics assessment. However video is collection of frames in a time period. So, all the aesthetics features which performed well for images can be used for video aesthetics assessment. However, there is one important feature a video have i.e. motion. When these aesthetics features are computed for video, these features does not perform well as does with images. So, it is a challenging task for researchers of the video aesthetics assessment. Some authors ([9], [11]-[15]) has started work on it, to describe such aesthetics features for videos. However, they are not able to show how the presentation of image based features in the temporal domain affects video aesthetics [11].

The motion features extracted or designed by researcher are based on the some basic film guidelines. Each feature shows characteristics of the professional videos or techniques of experienced photographer. In video aesthetics assessment authors have extracted features from different levels of videos i.e. cell-level, frame-level and shot-level.

A. *Frame-level Feature Extraction*

In previous studies it is shown that subject of interest attracts most of the visual attention of human eyes, using this [9] attempted to extract features from subject for video aesthetics assessment. As a result, these subjects based features serve as criteria for aesthetics quality assessment for their work. But work [9,11,12] assume that subject based feature extraction has limited applicability as consumer videos has many subject in their videos.

Inspired from this work [11,12,13] focused on feature extraction from salient regions of images and videos. Recently work from [17] achieved significant performance in this task. By assuming this, they designed some motion based features: *length of the subject region motion* [9,12] and *motion stability* [9,11,13].

These features are designed based on the study that experienced videographer/photographers change their shooting angle and focus continually to show the audience, subjects expressions and to tell story more effectively. Again, they see that camera shaking is much less in high quality videos than in low quality videos taken by amateurs. They think that these features are more effective in distinguishing between high and low-quality videos. These features are basic motion features. Feature Extraction from salient regions has got much attention in recent work.

To extend above work [11,13] proposed new motion features: they observed that while shooting videos, space must be reserved in moving direction of videos for better aesthetics, they designed a feature based on this study and called it *Motion Space* . This feature has been tested in many (i.e. categories of) videos and result show that it is quite important in video aesthetics assessment. They also considered *Motion Direction Entropy* as a motion feature which measures the amount of motion of every pixel, because amount of motion is different for different category of videos.

Apart from the features specified above one more basic guideline from film has used which proved to be very effective for aesthetics assessment i.e. *shooting Type*. There are three different shooting techniques are used by photographer i.e. static shot, which defines the large camera motion with respect to subject motion, another one is panorama shooting type, which defines equivalent motion of both the camera and subject and third guideline is the large subject motion with respect to camera motion. Apart from these motion features some work [16] shows the use of dense trajectories for aesthetics assessment. These trajectories generally show the motion direction of subject. However, these trajectories can be used for both foreground and background motion estimation and somehow related to shooting type.

B. Shot-level Feature Extraction

The previous work was more focused in extracting motion features from frames. Then authors thought of taking shot as a unit of assessment of videos because video is generally composed of shots and shots actually gives meaningful information. Work from [14,15] attempted to design motion based features for shots also and followed guidelines from film makers. They think that frame level assessment only give independent measurement or some value for features but shot level assessment is more focused on identifying characteristics of motion also. Inspired from this idea authors focused on video shooting type i.e. static or panorama. They took *foreground and background motion*, *characteristic of motion* and *locality of motion* i.e. motion in foreground as a feature. First two features are more concerned about the characteristics of camera and subject motion while third one is measure of motion.

The work above was focusing on extracting features from all frames in video. However, key-frames in video contain the rich information of video. Work from [15] extracts *texture dynamics* as a feature from key-frame for aesthetics assessment. However their results show that these features are not much helpful for aesthetics assessment of videos.

C. Cell-level Feature Extraction

As observation from above work, some authors think that as frame-level features perform well because all the frames are considered for aesthetics assessment. However [16] think that features can be extracted from cells of the frames by dividing the frames into blocks for low-level aesthetics assessment. Considering this author divides the key-frame into cells as a rectangular block of m×m grid and then design features which are equivalent to the low-level features like color, shape i.e. *Dark-channel*, which reflect the clarity, saturation and hue in image.

Every motion features are induced by some photography rules, so all the features are important but for different categories of videos.

III. DATASET

In absence of a common ground truth, one of the main issues that researchers face is the lack of an adequate dataset annotated in terms of aesthetic score over a reasonable number of voting observers. The alternative of creating it by means of controlled experimental studies has not been explored so far; only video sharing website like youtube.com videos can be taken for reference study, share the common drawback of their noisy data, but have the advantage of counting with a large set of videos.

The paper [12] includes an analysis of the goodness of random subsets of various hundred videos (i.e. 160 videos) from youtube.com databases, they took controlled user study and rated their dataset for two class labels i.e. positive and negative, with a constant ratio of fifty (50-50) and make it publically available as a benchmarked dataset. However, work from [11] find some cross validation errors and improve this dataset by adding 40 professional movie clips and said it publically benchmarked. So, this dataset can be considered for video aesthetics assessment or may improve it.

IV. DISCUSSION

There is a clear need for a stable dataset of rated videos available for researchers. In addition to the online community-based databases, counting with experimental dataset obtained under a controlled environment would probably significantly increase the performance of the current solutions. Studies have already been conducted for the evaluation of image and videos aesthetics appeal in consumer and professional photography using a number of controlled observers and for different sets of videos in different categories [11,12,13].

Computational image aesthetics has proven to be an increasingly emerging field of interest. Building upon the basis of the presented research, we can trace some of the major paths to be explored by the upcoming attempts to solve any of the problems in this field. The category of feature extraction unit i.e. frames, shot and cell, is also important to say as it can clearly define that which study is most helpful for learning and assessing videos.

So, this study may or may not be much helpful for aesthetics assessment but work from [11,12,13,14,16] taking frame as a unit for aesthetics assessment. However, this is most preferable unit to extract features. Apart from these some work [14,15] focusing on shot level aesthetics assessment and it is proved to be good because shot actually gives meaningful information of video. Research in this direction might be more helpful and also important for learning shot characteristics. Features from shots as *shooting type* and *length of subject motion* performs well, however because of benchmarked dataset results cannot be compared but we can compare performance on the same database.

V. CONCLUSION

Computational aesthetics is an achievement for the researchers of the domain, which is generally based on guidelines from the experience people. Unlike image aesthetics, for videos there is no publically benchmarked dataset available and also there is lack of knowledge from professionals. But there is some work [11,12,13,16] which seems to be promising for the video media aesthetics assessment. Their results show that frame-level feature assessment is well for video and shot-level features also help to understand the characteristics of motion.

REFERENCES

[1] M. Saini, R. Gadde, S. Yan, and W. T. Ooi, "Movimash: Online mobile video mashup," in Proc. 20th ACM Int. Conf. Multimedia (ACMMM'12), 2012, pp. 139–148.

[2] Y. Ke, X. Tang, F. Jing, "The Design of High-Level Features for Photo Quality Assessment," CVPR '06 Proceedings June, 17 2006

[3] R. Datta, D. Joshi, J. James and Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach," Computer Vision – ECCV 2006

[4] Y. Niu and F. Liu, "What Makes a Professional Video? A Computational Aesthetics Approach," IEEE Transactions on Circuits and systems for Video Technology, VOL. 22, NO. 7, JULY 2012

[5] G. Peters, "Aesthetics primitives of images for visualization," in Proc. 11th International Conference Info. Visualization (IV' 07), 2007, pp. 316-325

[6] H. Tong, M. Li, H. Zhang J. He, C. Zhang, "Classification of Digital Photos Taken by Photographers or Home Users," Advances in Multimedia Information Processing - PCM 2004, Volume 3331, 2005, pp 198-205

[7] Dhār, V. Ordonez, T.L. Berg, "High level describable attributes for predicting aesthetics and interestingness," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011
[8] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato, "Aesthetic quality classification of photographs based on color harmony," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'11), 2011, pp. 33-40
[9] Y. Luo and X. Tang, "Photo and Video Quality Evaluation: Focusing on the Subject," ECCV 2008, Part III, LNCS 5304, pp. 386–399, 2008. Springer-Verlag Berlin Heidelberg 2008
[10] L. Liu, R. Chen, L. Wolf, D. Cohen-Or, "Optimizing Photo Composition," EUROGRAPHICS 2010
[11] C. Yang, H. Yeh and C. Chen, "Video Aesthetic Quality Assessment by combining semantically independent and dependent features," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011
[12] A. K. Moorthy, P. Obrador, and N. Oliver, "Towards computational models of visual aesthetic appeal of consumer videos," in Proc. Eur. Conf. Computer Vision (ECCV'06), 2010, pp. 1–14.
[13] H. Yeh, C. Y. Yang, M. Lee, and C. Chen,"Video Aesthetic Quality Assessment by Temporal Integration of Photo- and Motion-Based Features," IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 8, DECEMBER, 2013
[14] S. Chung, J. Sammartino, J. Bai, B. A. Barsky, **"**Can Motion Features Inform Video Aesthetic Preferences?," University of California at Berkeley Technical Report No. UCB/EECS-2012-172June 29, 2012
[15] S. Bhattacharya, B. Nojavanasghari, T. Chen, **"**Towards a comprehensive computational model for aesthetic assessment of videos**,**" ACM 978-1-4503-2404-5/13/10 October 21–25, 2013
[16] Y. Wang, Qi Dai, R. Feng, Y. Jiang, "Beauty is Here: Evaluating Aesthetics in Videos Using Multimodal Features and Free Training Data" ACM 978-1-4503-2404-5/13/10 October, 2013
[17] Esa Rahtu, J. Kannala, M. Salo, J. Heikkila, "Segmenting Salient Object from Images and Videos," Computer Vision – ECCV 2010, Volume 6315, 2010, pp 366-379, Dec 2010

