Assessing the Quality of User–Generated Content

Stefan Winkler
(Advanced Digital Sciences Center (ADSC), Singapore 138632)

Abstract

With the widespread use of digital cameras, imaging software, photo–sharing sites, social networks, and other related technologies, media production and consumption patterns have become much more multifaceted and complex than they used to be. User–generated content in particular has grown tremendously. As a result, quality of experience (QoE) and related quality assessment (QA) methods must also be looked at from a different angle. This paper contrasts some of the traditional quality assessment approaches with newer approaches designed for user–generated content. It also describes some sample applications we have developed.

Keywords

image quality; photo collections; social media; photowork; summarization

1 Introduction

Quality assessment (QA) for images dates back to the 1970s, when the first studies were done on the visual cortex, vision modeling, and digital imaging. Algorithms based on models of the human visual system now compete with more pragmatic, image– or feature–based methods [1]. Video QA has a similar, albeit shorter, history [2].

Ubiquitous and affordable digital cameras (also in the form of embedded image capture devices) now enable users to take pictures and videos almost anywhere, anytime. This has led to an explosion in the amount of picture material produced by both amateurs and professionals.

However, traditional media and user–generated content are fundamentally different on many levels, especially from a quality assessment perspective (Table 1).

1.1 QA for Traditional Media

Traditional QA methods have focused mainly on the processing and distribution chains of broadcast media, that is, the compression, transmission, and enhancement of images and video. In many cases, there is an explicit reference (e.g. the source image or video), that passes through the system and undergoes certain changes (e.g. loss of fidelity, compression artifacts, packet loss, noise removal).

For such high–value, professional content, QA is typically done manually during the production process. Afterwards, the content is prepared for and distributed to many paying consumers via channels such as cinema and broadcast TV. It is very much a linear process in which a single high–quality source passes through various processing steps that may change or affect the quality of the content. The entities concerned with quality throughout this process are typically encoder manufacturers, content providers, service providers, and operators. Furthermore, traditional media is designed for a wide audience, and as a consequence, the average user and mean opinion score (MOS) are the quality unit and benchmark of choice.

Most QA algorithms [1], [3], [4]; databases [5]; standards [6]; and products have so far have focused on this stage, where fidelity (i.e. how closely the processed image/video resembles the original source content) is of primary importance.

1.2 QA for User–Generated Content

With user–generated content, fidelity is secondary. The criteria for QA and enhancement are not only image– or content–specific (e.g. impairments or scene composition) but also user–centric (i.e. what is most relevant to the user in a collection). This is contrary to traditional QA approaches.

Automated QA for user–generated content is useful primarily in the production process for a number of reasons:

- Even if QA could be done manually by the user, it would be too time–consuming for most. Besides, the average user needs guidance to produce good–quality content.
- Processing and distribution are simple and are largely hidden from or opaque to the user (e.g. compression in the camera or uploading content to a website).
- Quality becomes a much more personal concept because it is mainly the user and the circle of people they may share

▼Table 1. Traditional media vs. user–generated content

Stages	Traditional Media	User–generated Content
Production	Professional quality, premium content	Amateur content/quality
Processing	Encoding, transcoding, multiplexing, etc.	Minimal user intervention, or hidden from user
Distribution	Real–time streaming, many users, high network demands	Sharing with friends, typically downloads

the content with who matter the most. Indeed, personalization has not received much attention so far despite its importance for user−generated content.

We discuss these aspects in more detail, using the example of photo collections. However, they also apply to other types of media.

2 Photo Collections

The most common type of user−generated content today is digital photos. Collections typically comprise pictures taken during an event or trip, possibly by multiple users using different devices. They may also comprise images shared in social networks, images on websites, and images stored in third−party repositories. Devices may include single−lens reflex (SLR) cameras, point−and−shoot cameras, and camera phones.

It has become so easy to take lots of pictures that users regularly have to deal with large photo collections. The role of QA here is primarily the selection of the best and most representative pictures from a collection. This task can be broken down into two basic steps: screening and summarization. In screening, the best photo is selected from a group of similar photos (typically multiple shots of the same scene) and enhancements are applied if necessary. In summarization, a subset of pictures is chosen for an album. Often, the purpose is to tell a story or share an experience.

Intelligent user interface design and personalization are essential in both steps because of the importance of user−specific criteria, tastes, and preferences. It is also difficult to fully automate these processes to the satisfaction of users.

2.1 Screening

With digital cameras, it has become common practice to take multiple pictures of the same scene. Users typically take two or three shots per scene on average, and for certain scenes and situations, between eight and ten shots [7]. Selecting the best picture from a group of pictures typically involves evaluating lighting, exposure, and white balance; framing and perspective; postures, actions and faces of people in the scene; and basic image quality [8].

Typically, the quality of a picture is assessed by comparing it to a reference with the same content but without impairments (full−reference comparison). It may also be assessed on its own (no−reference comparison). When there is no reference image, traditional full−reference methods do not apply. No−reference methods can be used in principle, but they generally work best along a single impairment dimension, for example, quantization or blur, of the same image.

The problem here is more general and revolves around comparing pictures that have similar, related (but not identical) content and different quality/impairment dimensions and levels (one image may be blurred while another may be underexposed). We need to choose the best of the pictures, and this requires a good understanding of the effects that different impairment dimensions have on perception.

Much of the existing work done in this area has been focused on the aesthetic aspect of quality. Features used to estimate the aesthetic value or classify aesthetic categories of consumer photographs are color and illumination, composition, depth of field and perspective, and subject matter [9]. Such features have even been used to provide automatic feedback to photographers when composing a shot [10].

Although aesthetic aspects are no doubt important, they become secondary for the rather large class of "family photos" that most amateur photographers are concerned with. For family photos, human factors, such as facial expressions, pose, activity, and interaction, are by far the most important factors that determine the value of an image. If there are people in a scene, a human observer will immediately focus their attention on them and their faces and largely ignore the other characteristics of the image [11]. Consequently, assessing human factors is of paramount importance to intelligently process family photo collections.

Unfortunately, human factors are much harder to measure than aesthetic or other low−level factors. Problems such as face detection or recognizing people, poses, activities, and expressions are still some of the most challenging problems in computer vision, especially for images captured in uncontrolled conditions.

2.2 Summarization

Selecting the most representative pictures from a set is similar to storytelling or summarization; the key is to identify which scenes the user considers to be important in the story. There may not be a unique set of pictures that can fully represent a collection because of the large number of possible subsets and different possible themes.

An effective summary should have certain properties: quality, the selected photos have to be interesting and attractive; diversity, there should be no duplication or redundancy; and coverage, important people or events should appear in the summary [12].

Criteria that people use to choose pictures from a collection have been studied previously. Such criteria include specific people, variety of places, and general image quality [8]. These can be used to guide the (semi)automatic selection process. Furthermore, it can be helpful that some events, such as weddings, in certain cultures follow a specific sequence of events. There may be a number of important milestones that need to be included.

It may also be desirable to find pictures that are not be part of the initial set but that are nevertheless relevant to the story and can be sourced from external collections. Examples of this are a map of places visited or a better picture of a popular sight if the ones present in the collection are not satisfactory.

Finally, the purpose of summarization is not necessarily to produce a static album or set of photos; instead, it can be used for dynamic browsing of photo collections. Of particular interest is "associative" browsing, which refers to any method that assists users to discover, browse, or navigate large data libraries in a more intuitive way. With associative browsing, a user is guided towards other similar data that is relevant to the data currently being viewed. Naturally, the human factor plays a big role in the associations a person has when looking at photos. This makes it difficult not only to develop approaches for meaningful summarization but also to evaluate their effectiveness. Furthermore, associations are highly subjective, which brings us to the topic of personalization.

2.3 Personalization

Personal and social factors are much more important for user–generated content because such content is often only meaningful to the user and their family and friends. Consequently, generic models for appeal may be even more short–lived than those for aesthetics [13]. For personalization to be effective, it must be carefully tailored toward learning personal or situational preferences. Personalization implies that the criteria for selecting images are not those of the average user (as typified by the traditional MOS) but of the specific user. This approach is quite different to the way the topic is usually approached.

Pictures selected from a personal collection by a random person are unlikely to be the meaningful or relevant pictures for the owner. A given person may have certain preferences in terms of perspective, lighting, color, enhancements, subjects, expressions, and poses. Any QA system for such content should be able to offer personalized suggestions according to the user's individual taste and preferences. Image content and visual characteristics alone are likely insufficient, and image metadata such as tags, geographical information, time, and date can greatly help with personalization tasks.

3 Examples

3.1 Interactive Photo Screening

People often take multiple shots of the same scene and then select the best picture(s) from the set afterwards. This is especially common for photos that involve people, for example, family photos with babies and kids or photos of certain important events, such as weddings or graduations. In these cases, we want to capture the best moments when the subjects of the photos have the most memorable poses. Then, we want to share our favorite photos with family and friends.

Photo screening (triaging) is one of the most common photowork activities. Existing photo software provides very limited computational or interface support for photo triaging; in many cases, this basic task still relies on flipping through the photos and viewing them one by one, which is a primitive interaction method. Using thumbnail images as an alternative does not work well either because the relevant image features to be compared are often too small in a thumbnail view. Thus, details such as facial expressions are not easily recognizable. This is especially problematic on mobile devices with limited screen space and resolution.

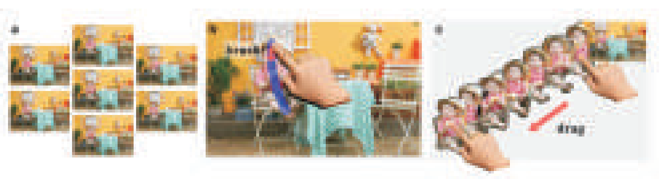
We therefore propose an effective and easy–to–use brush–and–drag interface that allows the user to interactively explore and compare photos within a broader scene context (Fig. 1) [7]. First, the user brushes an area of interest on a photo. Our tailored segmentation engine automatically determines corresponding image elements among the photos. Then, the user can drag the segmented elements from different photos across the screen to explore them simultaneously and further use simple finger gestures to interactively rank photos, select favorites to sharing, or remove unwanted photos. This focus + context design allows the user to choose any area or object of interest by brushing (focus) while retaining the overview photo (context). The photo triaging process becomes more flexible and user–centric.

We implemented our interface on an Apple iPad 2 and evaluated it with a number of users. According to both objective and subjective measurements, our brush–and–drag interface is better than the conventional method of browsing photos by flipping. The participants preferred our interface in terms of ease of use and were able to select favorite photographs from groups of similar images more quickly [7].

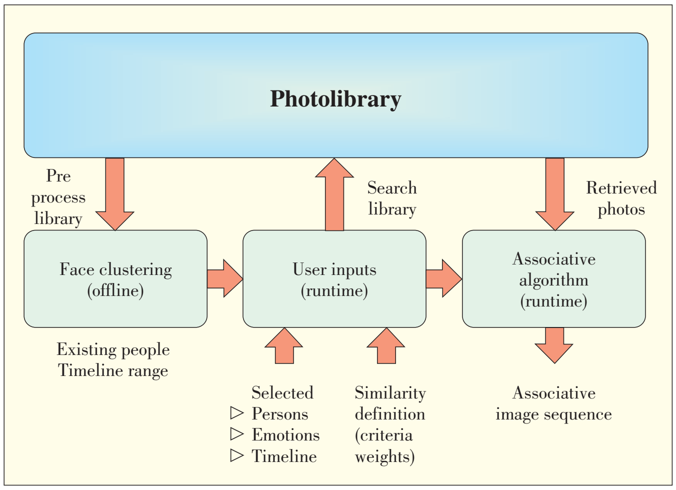
3.2 People–Centric Summarization

There is a considerable body of research on slideshows and even some commercial products (e.g. Apple iPhoto) for automatic face annotation in personal photo–albums. However, there are no existing systems that can identify people and their emotions in photo libraries and use this information, along with other similarity features, to form an associative chain of image transitions or browsing suggestions. The majority of existing techniques that estimate human emotions do not take into account the human factor; they mainly focus on other global or local image features.

We have developed a method of creating people–centric slideshows that takes into account people and their emotions [14]. Fig. 2 shows how this system operates. The user specifies the person(s) that they wish to include along with the importance that they assigns to different similarity criteria. The system automatically scans the photolibrary for photos of the requested person(s) and performs face recognition and emotion



▲Figure 1. The brush–and–drag interface for photo screening.


Photolibrary
Pre process library
Search library
Retrieved photos
Face clustering (offline)
User inputs (runtime)
Associative algorithm (runtime)
Existing people Timeline range
Selected
▷ Persons
▷ Emotions
▷ Timeline
Similarity definition (criteria weights)
Associative image sequence

▲**Figure 2. People–centric slideshow creation.**

estimation [15]. The retrieved images are then automatically arranged into a meaningful sequence, taking into consideration the importance values assigned by the user. The resulting image sequence can be displayed as a slideshow or to give browsing suggestions to the user. The current similarity criteria include facial emotions/expressions, timeline, color, and scene characteristics.

The system has a flexible design, and new similarity attributes can be easily added. It is also adaptable to the user's preferences. Different degrees of importance can be defined for the similarity attributes, making it a useful tool for personalized associative browsing or slideshow creation. The proposed system also takes into account emotions, which makes it useful for filtering out undesirable expressions such as angry faces.

4 Conclusions

We have contrasted traditional media and user–generated content in terms of their requirements for quality assessment. Using the example of photo collections, we have outlined a number of relevant research areas for novel quality assessment approaches. We have also highlighted two sample applications for photo screening and summarization that address issues such as user interface design and personalization that are important for user–generated content.

Acknowledgements

This research is supported by the Advanced Digital Sciences Center (ADSC) under a grant from the Agency for Science, Technology and Research of Singapore (A*STAR).

References

[1] W. Lin, C.–C. J. Kuo: "Perceptual visual quality metrics: A survey." *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, May 2011.

[2] S. Winkler, P. Mohandas: "The evolution of video quality measurement: From PSNR to hybrid metrics." *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, Sept. 2008.

[3] S. Chikkerur, V. Sundaram, M. Reisslein, L. J. Karam: "Objective video quality assessment methods: A classification, review, and performance comparison." *IEEE Transactions on Broadcasting*, vól. 57, no. 2, pp. 165–182, June 2011.

[4] M. Vranješ, S. Rimac–Drlje, K. Grgic: "Review of objective video quality metrics and performance comparison using different databases." *Signal Processing: Image Communication*, vol. 28, no. 1, pp. 1–19, Jan. 2013.

[5] S. Winkler. "Analysis of public image and video databases for quality assessment." *IEEE Journal on Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 616–625, Oct. 2012.

[6] S. Winkler: "Video quality measurement standards–current status and trends." In *Proc. 7th International Conference on Information, Communications and Signal Processing (ICICS)*, Macau, Dec. 7–10, 2009.

[7] S. J. Kim, H. Ng, S. Winkler, P. Song, C.–W. Fu. "Brush–and–Drag: A multi–touch interface for photo triaging." In *Proc. ACM SIGCHI International Conference on Human–Computer Interaction with Mobile Devices and Services (Mobile HCI)*, San Francisco, Sept. 21–24, 2012.

[8] A. E. Savakis, S. C. Etz, A. Loui: "Evaluation of image appeal in consumer photography." In *Proc. SPIE Human Vision and Electronic Imaging*, vol. 3959, San Jose, CA, Jan. 2000.

[9] C. Li, T. Chen, "Visual aesthetic quality assessment of digital images." In R. Lukac (ed.), *Perceptual Digital Imaging: Methods and Applications*. Chap. 4, pp. 91–122, CRC Press, 2012.

[10] L Yao, M. Qiao, P. Suryanarayan, J. Z. Wang, J. Li: "OSCAR: On–site composition and aesthetics feedback through exemplars for photographers." *International Journal of Computer Vision*, vol. 96, no. 3, pp. 353–383, 2012.

[11] E. Birmingham, W. F. Bischof, A. Kingstone: "Saliency does not account for fixations to eyes within social scenes." *Vision Research*, vol. 49, pp. 2992–3000, 2009.

[12] P. Sinha, S. Mehrotra, R. Jain: "Summarization of personal photologs using multidimensional content and context." In *Proc. ACM International Conference on Multimedia Retrieval (ICMR)*, Trento, Italy, April 17–20, 2011.

[13] D. Joshi, R. Datta, Q.–T. Luong, E. Fedorovskaya, J. Z. Wang, J. Li, J. Luo, " Aesthetics and emotions in images: A computational perspective." *IEEE Signal Processing Magazine*, vol. 28, no. 5, pp. 94–115, Sept. 2011.

[14] V. Vonikakis, S. Winkler. "Emotion–based sequence of family photos." In *Proc. ACM Multimedia Conference*, Nara, Japan, Oct. 29–Nov. 2, 2012.

[15] V. Vonikakis, S. Winkler. "System for creating slideshows based on people and their emotions." In *Proc. ACM Multimedia Conference*, Nara, Japan, Oct. 29–Nov. 2, 2012.

Manuscript received: January 16, 2013

Biography

Stefan Winkler

Stefan Winkler (stefan.winkler@adsc.com.sg) is principal scientist and director of the Interactive Digital Media Program at the University of Illinois Advanced Digital Sciences Center (ADSC) in Singapore. He has previously co–founded a start–up, worked in several large corporations, and held faculty positions at two universities. Dr. Winkler received his PhD degree from the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, and his MEng. degree from the Technische Universität Wien, Austria. He has published more than 80 papers and authored a book called *Digital Video Quality*. He is an associate editor of *IEEE Transactions on Image Processing* and *IEEE Signal Processing Magazine* (standards column). He has also contributed to video quality standards in VQEG, ITU, ATIS, VSF, and SCTE. His research interests include video processing, computer vision, perception, and human–computer interaction.