Quality Assessment of User-generated Video Using Camera Motion

Jinlin Guo[1], Cathal Gurrin[1], Frank Hopfgartner[1],
Zhenxing Zhang[1], and Songyang Lao[2]

[1]CLARITY and School of Computing, Dublin City University
Glasnevin, Dublin 9, Dublin, Ireland
{jinlin.guo,cgurrin,frank.hopfgartner,zzhang}@computing.dcu.ie
[2]School of Information System and Management
National University of Defence Technology, Changsha, Hunan, China
laosongyang@vip.sina.com

Abstract. With user-generated video (UGV) becoming so popular on the Web, the availability of a reliable quality assessment (QA) measure of UGV is necessary for improving the users' quality of experience in video-based application. In this paper, we explore QA of UGV based on how much irregular camera motion it contains with low-cost manner. A block-match based optical flow approach has been employed to extract camera motion features in UGV, based on which, irregular camera motion is calculated and automatic QA scores are given. Using a set of UGV clips from benchmarking datasets as a showcase, we observe that QA scores from the proposed automatic method and subjective method fit well. Further, the automatic method reports much better performance than the random run. These confirm the satisfaction of the automatic QA scores indicating the quality of the UGV when only considering visual camera motion. Furthermore, it also shows that the UGV quality can be assessed automatically for improving the end users quality of experience in video-based applications.

Keywords: Quality Assessment, User-generated Video, Irregular Camera Motion, Optical Flow

1 Introduction

As the proliferation of Web 2.0 applications, user-generated video (UGV) [1] is poised to inundate the Internet. Recent statistics show that, on the primary video sharing website, YouTube[1], 48 hours of video are uploaded every minute by users, resulting in nearly 8 years of content uploaded every day. Furthermore, over 800 million unique users visit YouTube each month, and more than 3 billion hours of video are watched on YouTube[2]. This has increased the requirement on video

[1] http://www.youtube.com
[2] According to http://www.youtube.com/t/press_statistics

websites to match the video quality expectation of the end users and viewers, such as users always prefer to viewing the best-quality of video among many clips captured in a same concert using personal capturing devices. Therefore, reliable quality assessment (QA) of UGV plays an important role in providing good quality of service (QoS), in improving the end users' quality of experience (QoE) and in managing such a large amount of video data.

Reliable QA of video has attracted a lot of research interest, and numerous of video QA methods and measurements have been proposed over the past years with varying focuses on objective or subjective QA of video. Previous approaches to video QA have the following characteristics.

(1) Many aspects may affect the quality of video including, but not limited to, acquisition, process, compression, transmission, display and reproduction systems. Most of existing approaches to QA of video focus on the distortion caused by compression [2, 3] and transmission [4]. The quality of video itself is not assessed when it's captured. However, capturing conditions such as irregular camera motion (IRRCM) can degrade the perceived video quality.
(2) The commonly-used video QA measurements such as signal-to-noise ratio (SNR), peak-signal-to-noise ratio (PSNR) and mean squared error (MSE) [5], are computationally simply, however, they disregard the characteristics of human visual perception.
(3) A lot of research interest has been focused on objective video QA [2, 6], however, methods to assess the visual quality of digital video as perceived by human observer are becoming increasingly important, due to the large number of applications that target humans as the end users of video. The only reliable method to assess the video quality is to ask human subjects for their own opinions, which is termed subjective video QA. The subjective methods are based on groups of trained or untrained users viewing the video content, and rating for quality [5, 7]. It is impractical for most applications, and also time consuming, laborious and expensive, due to the human involvement in the process. However, subjective QA studies provide the means to evaluate the performance of objective or automatic technologies of QA. Combination of objective and subjective QA, which means objective QA methods should produce video QA scores that highly correlate with the subjective assessments provided by human evaluators, will likely be a trend of future research.

Moreover, the traditional QA methods are usually performed on broadcast video which has been professionally preproduced [8]. Compared to broadcast video such as news and sports video, UGV is usually of lower quality, due to the uncontrolled capturing conditions and various types of capture devices. For example IRRCM and fuzzy backgrounds are very common in UGV [9].

It should be noted that in [10], Wu et al. analyzed the IRRCM in home videos, and proposed a segmentation algorithm for home videos based on the categorization of camera motion. By support vector machines (SVMs), the effects caused by the camera motion were classified into four types: Blurred, Shaky, Inconsistent and Stable according to the changes of camera motion in speed,

direction and acceleration. Finally, video sequence were segmented, and each segment was labeled as as one of the four camera motion effects. However, in this paper, we employ IRRCM to assess the visual quality of UGV. The rationale is that IRRCM or camera shaking is so commonplace in UGV and it causes the degradation of perceived quality of UGV. Our contributions in this work are that 1)firstly, we propose a automatic approach to perform QA of UGV using IRRCM, and 2) subjective QA of UGV is conducted and compared with the proposed automatic QA method.

The remaining parts of this paper are organized as follows. In Section 2, we analyze the IRRCM feature in UGV and describe the approach to UGV IRRCM extraction and scoring proposed in this work. In Section 3, experimental results based on a set of UGV clips from benchmarking datasets are represented, and also compared with the results from the subjective assessment and random run. Finally, we give our conclusions and outline future work in Section 4.

2 IRRCM Extracting and Scoring

It is hard to define the relationship between the video visual quality and camera motion. However, if one video clip contains more IRRCM, the visual quality is generally perceived as of being lower in subjective assessment.

2.1 Camera Motion Analysis in UGV

Camera motion is an important factor affecting visual quality of video. The visual quality of UGV is highly relevant to three properties of camera motion [10], that is *speed*, *direction* (*orientation*) and *acceleration*. These three properties affect UGV quality in different ways. As shown in Fig. 1, the classification of effects caused by the change of three properties of camera motion can be represented as a decision tree [10].

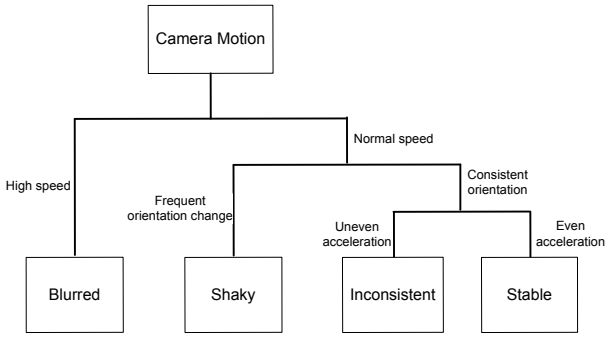


Fig. 1: Classification of camera motion effects in UGV

If the speed of camera motion is high, the captured frames will be blurred. When the speed is normal, but the orientation of camera motion changes frequently, namely, the camera moves back and forth repeatedly, the captured videos are regarded as shaky. When speed is normal and orientation is consistent, but the accelerations of camera motion in consecutive-extracted frames are uneven, that is, the variance of acceleration is large, the captured videos are inconsistent. The normal camera motion with rare orientation changes and even accelerations lead to stable motion.

In this paper, we name IRRCM to be the effects caused by the changes of acceleration and orientation. In the quality assessment, we jointly weight the acceleration and orientation changes. The acceleration change is measured by the magnitude change of two consecutive-extracted motion vectors, whereas the orientation change is calculated by angle between these two vectors. In the following parts of this section, we will describe our QA of UGV method in detail.

2.2 Extraction of Background Camera Motion

We use a two-parameter motion model to deal with camera motion (X and Y). X depicts horizontal movement to the left and right, commonly referred to as X transition and pan. Y depicts vertical movements (up and down), referred to as Y transition and tilt. This provides relative computational efficiency.

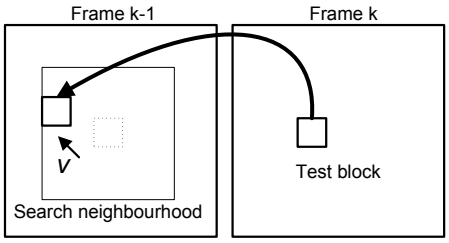


Fig. 2: Neighborhood search for similar blocks (Origin is at bottom left corner)

We adopt a block-match based optical flow approach for extracting camera motion. The rationale behind this approach is that camera movement can be detected by comparing neighboring regions of consecutive frames. Each video clip has a minimum of 24 frames per seconds. Given this high frequency, we do not expect large differences between neighboring frames. Therefore, we limit our approach by extracting five frames per second, which also improves the efficiency. As illustrated in Fig. 2, given a test block with the size of S in the current frame, a search neighborhood of $3S$-size, centered around the test block in preceding frame is defined. We then search for the similar block within the search neighborhood using a sliding window approach. In order to find the most similar block in the search neighborhood, the *Maximum Matching Pixel Count*

(MMPC) is determined as follows:

$$D\left(x_t, y_t; x_p, y_p\right) = \begin{cases} 1 & if \sum_{c\in\{R,G,B\}} |P_c(x+i, y+j) - \\ & Q_c(x+d_x+i, y+d_y+j)| \leq T \\ 0 & else \end{cases} \tag{1}$$

$$(x_p^{'}, y_p^{'}) = \underset{(x_p, y_p)}{argmax} \sum_{i=1}^{S}\sum_{j=1}^{S} D(x_t, y_t; x_p, y_p) \tag{2}$$

Where (x_t, y_t) is the center of a test block. (x_p, y_p) defines the center of searched block located in the search neighborhood. P_c, is the color value of the pixel in the current frame, and Q_c is the color value of the pixel in the previous frame, (x, y) is the coordinate of the bottom left corner of the test block. The displacement between the two centers is defined as $d_x = x_p - x_t$ and $d_y = y_p - y_t$. T is the threshold. S is the size of the test block. Therefore, the displacement vector $\mathbf{v}$ (optical flow) is given by $v_x = x_p^{'} - x_t$, $v_y = y_p^{'} - y_t$. v_x and v_y are the X and Y motion component, respectively.

However, there are many uniform-texture areas in the video frame images that make the detected motion vector $\mathbf{v}$ unreliable. We checked the number of similar blocks within the neighborhood, if there are more than N blocks, this indicates a uniform-texture area, and the motion vector $\mathbf{v}$ is unreliable. In this case, we set $\mathbf{v} = 0$..

This process is repeated over the whole image to obtain a optical flow for each block. Large homogeneous regions, typically half of the image, are considered to be the background. The camera motion between frame $k-1$ and frame k is determined by computing the average motion vector $\overline{\mathbf{v}}_{k-1,k}$ of the blocks within this background region. Specifically, all the detected motion vectors are grouped into a histogram with eight bins by *orientation assignment*, each of which represents one orientation as shown in Fig. 3.

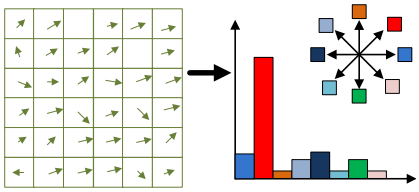


Fig. 3: Orientation assignment of optical flows for finding the camera motion

2.3 Scoring Irregular Camera Motion

After extracting all the camera motions between each consecutive pairwise-frame, we can get the changes of a camera motion by considering both acceleration and orientation change. For two consecutive camera motion vectors

(corresponding to three consecutive-extracted frames $k-1$, k, $k+1$), the acceleration a, and the orientation change $\theta_{k-1,k,k+1}$ both describe the IRRCM. They are computed as follows:

$$a = \|\overline{\mathbf{v}}_{k-1,k} - \overline{\mathbf{v}}_{k,k+1}\| / \Delta t = m_{k-1,k,k+1}/\Delta t \doteq m_{k-1,k,k+1} \tag{3}$$

$$\theta_{k-1,k,k+1} = arccos\left(\frac{\overline{\mathbf{v}}_{k-1,k} \cdot \overline{\mathbf{v}}_{k,k+1}}{\|\overline{\mathbf{v}}_{k-1,k}\| \, \|\overline{\mathbf{v}}_{k,k+1}\|}\right) \tag{4}$$

where Δt is time interval between two consecutive-extracted frames. Since we sample the frames uniformly (five frames per second), that is, Δt is a constant, the acceleration measurement is equal to the magnitude of the difference of two consecutive motion vectors, $m_{k-1,k,k+1}$. For a whole video clip, the final IRRCM is represented by the average acceleration (AA) $\overline{m}$ and average orientation change (AOC) $\overline{\theta}$, which are calculated by the average of all magnitude changes and orientation changes.

In order to give a QA score to a video clip based on how much IRRCM it contains, we firstly consider to assess the AA $\overline{m}$ and AOC $\overline{\theta}$, respectively. A quality grade system is firstly built on a training set by quantifying all the AAs into five levels. Given a test video clip and its $\overline{m}$, its AA rank r_m can be obtained by:

$$r_m = \left\lfloor 5 * \frac{\overline{m}}{m_{max}} \right\rfloor \tag{5}$$

where m_{max} is the maximum AA extracted in the training set (it keeps $\frac{\overline{m}}{m_{max}} < 1$). $\lfloor \cdot \rfloor$ is the *Floor Function*. The AOC rank r_θ can be obtained by the same way as the AA rank. The final rank r indicating the IRRCM in this clip can be described as:

$$r = \lfloor \omega_m * r_m + \omega_\theta * r_\theta \rfloor \tag{6}$$

where ω_m and ω_θ are the weights for the AA rank r_m and AOC rank r_θ respectively, which show the importance attached to the AA and AOC respectively by the observer, and $\omega_m + \omega_\theta = 1$. Here, $r = 0$ means least IRRCM, that is best quality, whereas, $r = 4$ is the worst quality. In order to compare the QA score obtained here with that from user subjective assessment, we set the final QA score of the video clip to $5 - r$.

3 Experiments

3.1 Experimental Setup

We conduct our experiments using the Internet video collection of the NIST TRECVid [11] 2011 Multimedia Event Detection (MED) task[3]. This dataset consists of publicly available UGV posted to various Internet video hosting sites. For this evaluation, we randomly select a subset of 1000 video clips (88 hours playing time) from the dataset and split it into a training (700 video clips)

[3] http://www.nist.gov/itl/iad/mig/med11.cfm

and a smaller-size testing set (300 videos clips) since subjective assessment is very time and labor consuming. The training set is used for training the related thresholds and aforementioned parameters. Based on the preliminary analysis of a preceding experiment, we chose the following settings: test block size $S = 10$, similarity threshold $T = 10$, threshold for number of similar blocks $N = 4$, and $\omega_m = \omega_\theta = 0.5$

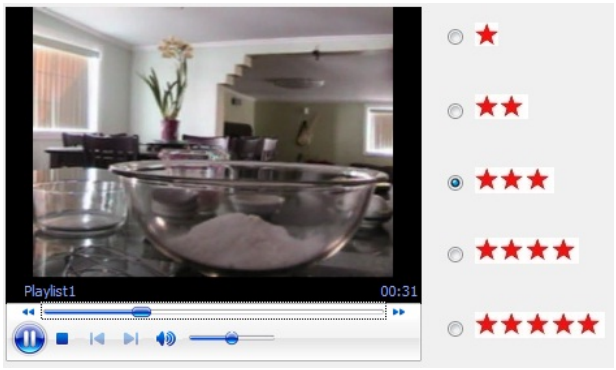


Fig. 4: User assessment interface. A video is displayed on the left, Users were asked to score the video quality by choosing 1-5 stars shown on the right.

We conducted user subjective QA for evaluating the performance of the proposed method in this paper. In total, ten human evaluators, all postgraduate students from our research center, were asked to assess the quality of all video clips in the testing set. Fig. 4 displays the user-assessment interface. This interface allows the users to play the video and to assess its quality on a Five Point Likert scale, ranging from very serious IRRCM (1) to few IRRCM (5). Before the evaluation, the users were given example video clips for each category that allowed them to familiarize themselves with this task and the expected quality of the video clips. For each video, we receive ten subjective scores from ten independent assessors, which we average to receive the final user assessment score s_u:

$$s_u = \begin{cases} \lfloor \overline{s} \rfloor & if\ \overline{s} - \lfloor \overline{s} \rfloor < 0.5 \\ \lfloor \overline{s} \rfloor + 1 & else \end{cases} \tag{7}$$

where $\overline{s}$ is the average of all the independent scores from the subjects. Furthermore, results from a random run are also reported.

3.2 Results and Analysis

We firstly analyze the factors that affect the camera motion detection. There are commonly three different identifiable categories of motion in video sequences, namely, background or camera motion, the foreground object motion and shot

or scene change. The shot or scene change is of different origin, namely external manual editing influences. The motion caused by the shot or scene change is easily removed since there are less matched pairwise-blocks in two consecutive frames that separately belong to two shots or scenes. Video clips showing foreground object motion under static camera are generally assessed as the best quality, with score 5. In this case, the detection method is especially effective if the object (or person) occupies a small part of the background. However, if an object is very close to a camera and moving irregularly, it displays the same visual effect as that caused by an IRRCM. In the case that both the object (person) and camera move, which is more complicated hinder the effectiveness of the detection method, we choose to process more frame images to overcome it compromisingly.

Now, we summarize the QA results from three methods. Table 1 compares the subjective scores, the determined scores using our proposed method and the random scores, respectively. As shown by the distribution of the scores, IRRCM is common in UGV.

Table 1: Number of UGV clips for each score grade from three methods

	Score				
Method	1	2	3	4	5
User #	10	31	62	79	118
Proposed #	16	49	59	55	121
Random #	53	64	59	68	56

A common challenge in user-based evaluation is the subjective nature of the assessment. Given a video clip, different subjects may give varying scores, which may be attributed to the subjective reasons, such as underestimating or overestimating. Fig. 5 depicts the Top 10 video clips with the highest standard deviation of user assessment scores. Moreover, the figure depicts the final subjective score, our automatic score and random score (shown in brackets above the box plot for each video clip). As can be seen, the subjective scores cover a wide range (from score 1 to 5) for these video clips, which most likely is due to subjective reasons or misuse. Nevertheless, the triplewise-score listed in the brackets show that our automatic method reports nearly the same scores as the subjective scores, and outperforms the random run.

In order to compare the differences between the proposed method or random run and the subjective assessment in detail, here, we take the scores from subjective assessment as the ground truth of QA results. We define three match levels between the automatic or random score and the subjective score for each

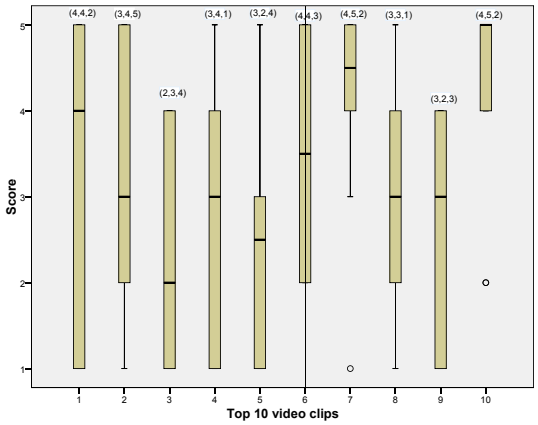
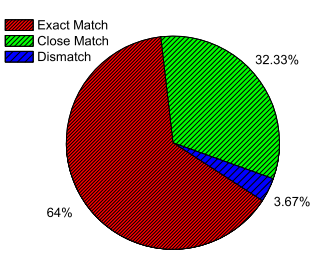


Fig. 5: Top 10 video clips with maximum standard deviation of user scores.

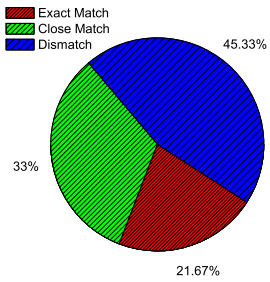
video clip:

$$\begin{array}{c} Exact\ Match,\ if\ |s_{pg} - s_u| == 0\ \ or \\ Close\ Match,\ if\ |s_{pg} - s_u| == 1\ \ or \\ Dismatch,\quad if\ |s_{pg} - s_u| >= 2 \end{array} \tag{8}$$

where, s_{pg} is the score from our proposed method or random run.



(a) Proposed *vs* Subjective



(b) Random *vs* Subjective

Fig. 6: Proportions of three match levels for proposed method *vs* subjective assessment, and random run *vs* subjective assessment

The pie graph in Fig. 6a shows the proportions of the three match levels. Overall, the results are very encouraging, more than 3/5 of the video clips with their two scores matched exactly, and about 32% of them achieved a close match.

However, match results shown in Fig. 6b indicate that random run reports much worse results. Table 2 shows the differences between the automatic and subjective score pairs and the random-subjective score pairs in more detail. The results suggest that the proposed method performs well on the video clips without or with very few camera motion since both methods scored nearly the same numbers of video clips with score 5 (118 *vs* 121). In total, 192 video clips receive the same scores from the proposed and subjective methods, 73 clips got lower scores from our method, whilst 35 clips were overestimated by our method. In contrast, random method only achieves 21.76% exact match and reports larger deviation to the subjective scores. This may be explained by the following facts: 1) Given a video clip, only several short parts of it contain IRRCM. The users easily overestimate the video quality; 2) We also speculate that this may be introduced by the imperfect camera motion detection and the weighting framework in Eq. 6. As future work, we aim to address this problem by comparing different weighting frameworks to compensate for this effect.

Table 2: Numbers of video clips for each score difference

	Score Difference ($s_{pg} - s_u$)							
	-4	-3	-2	-1	0	1	2	3
Proposed #	–	1	9	63	192	34	1	–
random #	23	44	38	62	65	37	22	9

4 Conclusions

In this paper, we have conducted an initial study towards QA of UGV by analyzing IRRCM. We adopt a block-match based optical flow approach to detect the IRRCM, based on which, a QA score is determined. In order to evaluate this quality score, we conducted a user subjective assessment of UGV quality. Using a set of UGV clips from the TRECVid MED task as a showcase, our results suggest that QA scores from proposed and subjective methods fit well. And the proposed method reports much better performance than the random run. Differing from previous work, our main contribution of this work is that UGV quality can be assessed automatically, hence improving the end users' QoE in a low-cost manner. There are still many ways to improve the current QA system. Significant ones include the utilization of better IRRCM detection methods, the utilization of other visual features such as the camera motion speed, and the adoption of UGV-based applications.

Acknowledgments

Thanks to the Information Access Disruptions (iAD) Project (Norwegian Research Council), Science Foundation Ireland under grant 07/CE/I1147 and the China Scholarship Council for funding. And also many thanks to the HMA group[4] for their help in the assessment effort.

References

1. Guo, J., Gurrin, C.: Short user-generated videos classification using accompanied audio categories. In: The First ACM International Workshop on Audio and Multimedia Methods for Large-Scale Video Analysis (AMVA), Nara, Japan (2012)
2. Olsson, S., Stroppiana, M., Baina, J.: Objective methods for assessment of video quality: state of the art. Broadcasting, IEEE Transactions on (1997)
3. Farias, Q., Carli, M., Neri, A., Mitra, S.K., Barbara, C.S., Barbara, S., Tre, R., Navale, V.: Video quality assessment based on data hiding driven by optical flow information. Proceedings of SPIE **5294** (2004) 190–200
4. Van der Auwera, G., Reisslein, M.: Implications of smoothing on statistical multiplexing of h.264/avc and svc video streams. Broadcasting, IEEE Transactions on (2009)
5. Seshadrinathan, K., Soundararajan, R., Bovik, A., Cormack, L.: Study of subjective and objective quality assessment of video. Image Processing, IEEE Transactions on (2010)
6. Chikkerur, S., Sundaram, V., Reisslein, M., Karam, L.: Objective video quality assessment methods: A classification, review, and performance comparison. Broadcasting, IEEE Transactions on (2011)
7. Internatinal Telecom Union: Rec. ITU-R BT.500-11. Technical report
8. Staelens, N., Moens, S., Van den Broeck, W., Marie andn, I., Vermeulen, B., Lambert, P., Van de Walle, R., Demeester, P.: Assessing quality of experience of iptv and video on demand services in real-life environments. Broadcasting, IEEE Transactions on (2010)
9. Guo, J., Scott, D., Hopfgartner, F., Gurrin, C.: Detecting complex events in user-generated video using concept classifiers. In: The 10th Workshop on Content-Based Multimedia Indexing (CBMI). (2012) 1–6
10. Wu, S., Ma, Y.F., Zhang, H.J.: Video quality classification based home video segmentation. In: ICME. (2005) 217–220
11. Smeaton, A.F., Over, P., Kraaij, W.: Evaluation campaigns and trecvid. In: MIR'06: 8th ACM Int. Workshop on Multimedia Information Retrieval, New York, NY, USA, ACM Press (2006) 321–330

[4] http://hma.dcu.ie/HMA/Home.html