Impact of Morphing-based Frame Synthesis on Bandwidth Optimization for 3DTI Video

Shannon Chen
Department of Computer Science
University of Illinois at Urbana-Champaign
cchen116@illinois.edu

Klara Nahrstedt
Department of Computer Science
University of Illinois at Urbana-Champaign
klara@illinois.edu

***Abstract*— In view of the resource demanding nature of 3D Tele-immersion (3DTI), we apply Morphing-based Frame Synthesis (MBFS) on delivery of both online and offline 3DTI visual content to decrease the resource consumption without degrading the perceptual quality. We further investigate the relationship between the level-of-motion of the content and the effectiveness of MBFS. In light of the results, we propose an on-the-fly resource adaptor for 3DTI video transmission which utilizes a perceptual model built from data compiled by a series of subjective experiments. Results show that our adaptor achieves 43% to 87% compression ratio for offline compression on 3DTI videos of different baseline user activities; and a 10% on-the-fly bandwidth saving on complex user activity without perceptible degradation.**

Keywords - 3D Tele-Immersion, QoE, Compression, Morphing, Frame Rate, Adaptation

I. INTRODUCTION

During the past decade, the potential of 3D Tele-immersive (3DTI) applications has gained its attention from both academia and industry. While most commercial 3D systems are specialized for sole purpose (e.g., Xbox for gaming), the development of 3DTI platforms [10] is currently aiming towards multi-purpose [7], multi-sites [8], and multi-modal [9] platform in order to enable a variety of user activities including e-learning, remote therapy, and interactive gaming [11-13].

Inevitably, with its great potential, the resource demand of 3DTI rises due to its interactive characteristic, complexity of 3D rendering, and delivery of media-rich content [10]. These challenges urge researchers to investigate deeper into the relationship between resource usage and the quality of experience (QoE) to find the efficient adaptation that minimizes the resource consumption without perceptible degradation of the service quality.

Echoing the trend of perception-centric quality evaluation, we investigate the viability of Morphing-based Frame Synthesis (MBFS) [7] in both online and offline compression of 3DTI video. MBFS exploits the unique properties of a 3DTI scene to enable auto feature matching and graphical morphing between frames. Synthesized morphing frames are added to the lossy-compressed video stream to restore its frame rate. Based on our observation, participants can have very different tolerance to quality degradation introduced by compression when viewing different types of user activities. Therefore, we investigate the effect of MBFS on perceptual quality of 3DTI videos with different level-of-motions (LoM) of the video content. The results help to build up models in order to quantify the motion level of a user activity and to map between parameters of MBFS compression and the resulting perceptual quality.

In light of the results, we apply the models in offline video compression and on-the-fly rate adaptation of 3DTI content. For offline compression, the LoM of a 3DTI video is analyzed and mapped to parameters of MBFS-based compression. The chosen parameters set the compression scheme at the equilibrium between reduction of bitrate and degradation of the perceptual quality. In result, our scheme achieves 43~87% compression ratio at the Just Noticeable Degradation (JNDG) perception threshold [20] which marks the imperceptible level of quality degradation. As for on-the-fly rate adaptation, we implemented a rate adaptor for 3DTI video delivery which fine-tunes the compression ratio dynamically according to the LoM of current user activity. Objective evaluation shows a 10% bandwidth saving without compromising the satisfaction of observers. We have further evaluated the adaptor via crowdsourcing user study. 72% of the 92 participants could not sense any degradation introduced by the adaptor.

In summary, the contribution of this work is four-folds. 1) Viability of adopting MBFS in the compression of 3DTI content, which renders efficient content delivery and alleviates its critical resource demand. 2) Analysis and quantification of user activities in 3DTI environment. 3) Investigation of the effect of LoM to the mapping between data rate and user satisfaction. 4) Proposition of an offline compression scheme and an on-the-fly data rate adaptor for 3DTI that brings substantial resource saving without perceptible degradation of the service.

The remainder of this paper is organized as follows. In the next section, an overview of previous works is provided. In Section III, we introduce the quantification of user activities. In Section IV, offline and online MBFS-based compression schemes are proposed. In Section V, we introduce the subjective experiments of which the results are used to optimize the effectiveness of MBFS. In Section VI we introduce the resource adaptation scheme for 3DTI video delivery and evaluate its performance via crowdsourcing. Finally, in Section VII we conclude.

II. Related Works

A. Morphing-based Fame Synthesis

The Morphing-based Frame Synthesis (MBFS) was first proposed by Chen and Nahrstedt in [7]. Morphing [3] is a special effect in motion picture that transit one image to another based on predefined feature pairs. Fig. 1 shows an example of applying MBFS on frame rate boosting. The first and the last frames in the figure are the only real frames captured by a camera, while the frames in between are all generated by morphing technique and hence boost up the original frame rate. The marking of the morphing line pairs [3] is automatically done by feature detection and feature matching. Due to the limited number of objects in the scene of 3DTI videos and the fair size of the human subjects, a meaningful number of matching features can be automatically provided for frame morphing. However, the application of MBFS in [7] uses only a discrete activity classifier based on Support Vector Machine (SVM) and body sensor to aid its video compression. This restricts the scheme from being a feasible adaptor for multi-purpose 3DTI platform that could support complex user activities during which user might combine and shift between various monotonic moves. In this work, we propose a continuous numerical indicator (Section III.A) based on image analysis to aid the dynamically fine-tuned compression schemes of both online and offline 3DTI video delivery without the requirement of wearable body sensors.

B. Resource Adaptors in 3DTI

1) Adaptors Based on Application Layer Semantic

Before 3DTI, adaptation schemes of multimedia services focused on individual streams. In other words, correlation and differentiation among streams were overseen. This implied waste of resource on delivering less important data to the application layer. In view of the problem, [10] and [6] introduced application layer semantics to the design of adaptation scheme in 3DTI in order to achieve efficient content dissemination. However, the schemes introduced extra complexity to data sharing in the overlay network. Because every link in the overlay network was transmitting different parts of the stream bundle [2] (a combination of streams containing different sensing data that are highly correlated), the topology of the network became a crucial factor that decided the efficiency of content delivery. A heuristic solution based on genetic algorithm was recently proposed in [14]. However, the computational complexity was inevitably raised with the adoption of the scheme.

2) Adaptor Based on Psychophysics

The effective end-to-end transport of delay-sensitive data has long been a subject of study in interactive multimedia services. The goal of [5] was to provide human-centric adaptation on media playout scheduling in 3DTI. The authors investigate the mappings between Quality of Experience and Quality of Service (QoS) metrics (e.g., end-to-end delay, PESQ, FPS) to find the suitable adaptation for gross-motor and fine-motion user activities. Different from our purpose, the work focused on resource allocation among streams rather than overall reduction of resource consumption. In [20], the authors exploited the limits upon human visual system to balance between spatial resolution (the color-plus-depth level-of-details) and frame rate. As a human nature, users were not able to tell the differences between certain levels of graphical degradation. In light of this nature, two QoE thresholds: Just Noticeable Degradation (JNDG) and Just Unacceptable Degradation (JUADG) were identified. With the two thresholds, the authors were able to adapt the resource consumption without degrading the service quality.

Figure 1. An example of MBFS.

III. Quantifying User Activities

In this section, we introduce the Level-of-Motion (LoM) metric to describe a user activity in a continuous space. The prominence of synthesized frames depends heavily on the motion of the video content. For some relatively static activities (e.g., storytelling, conversation), the difference between frames is small, which makes it more suitable to inject more synthesized frames without being noticed by the observer. On the other hand, for activities that contain more intense body movements (e.g., exercise learning, dancing), the portion of synthesized frames will have higher effect to the perceptual service quality. Our 3D video model is composed of frames in which each pixel contains its RGB-D (color plus depth) information. Thus, by analyzing the difference of RGB-D data of adjacent 3DTI frames, we define the LoM of a user activity in the next section. Different from previous user activity classifications that can only discern different activities in a predefined discrete space [7], the LoM is a continuous numerical metric that can be used to compare the motion intensity between two user activities. In the following, we introduce the definition of LoM and the four user activities which will be used intensively as baselines in the later sections.

A. Level-of-Motion

We use Root Mean Square Deviation (RMSD) to describe the difference between two 3DTI video frames, the RMDS of frame f_s and frame f_t is defined as

$$\mathrm{RMSD}(f_s, f_t) = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}\sqrt{(r_{ij}^s - r_{ij}^t)^2 + (g_{ij}^s - g_{ij}^t)^2 + (b_{ij}^s - b_{ij}^t)^2 + (d_{ij}^s - d_{ij}^t)^2}}{4WH} \quad (1)$$

where W (width), H (height) represent the video resolution; $r_{ij}^x, g_{ij}^x, b_{ij}^x$ represent the color of the pixel at coordination *(i, j)* of frame x, and d_{ij}^x is the depth of the pixel. The LoM metric can be applied under two scenarios: on-the-fly analysis of the motion intensity of a video; or offline analysis of the whole activity session. In the first scenario, the video frames are acquired one after another over time. Thus, we have to

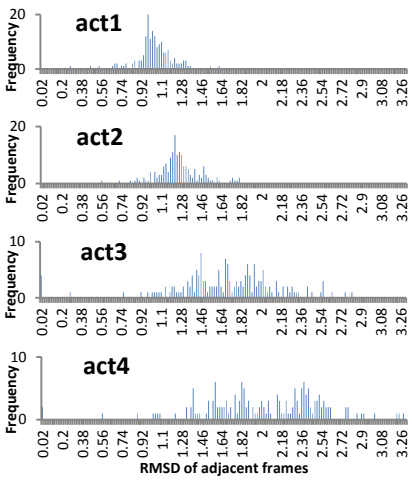


Figure 2. RMSD of baseline activities

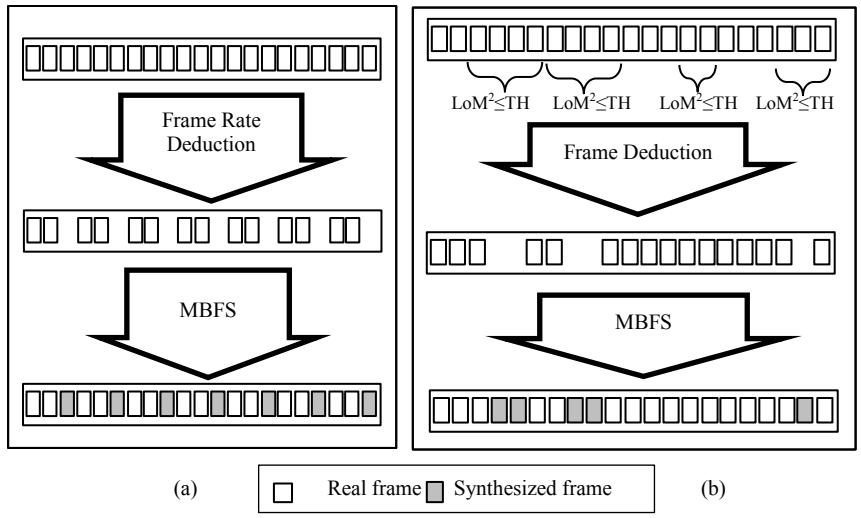


Figure 3. Compression schemes based on MBFS

predict the motion intensity of the activity at the next time instance based on the knowledge about previously acquired frames. Here we define the LoM for on-the-fly analysis as

$$LoM^1(f_k, \mathrm{N}) \equiv \frac{\sum_{i=0}^{N-2} RMSD(f_{k-i-1}, f_{k-i})}{N-1} \qquad (2)$$

where f_k is the newest frame acquired and N is an adjustable value indicating the number of past frames we want to include in the estimation of LoM. By definition, LoM^1 is the average RMSD of adjacent frames in the past N frames.

As for offline motion analysis, since we have every 3DTI video frames of the whole activity session, LoM can give more detailed description of the motion intensity during an arbitrary time interval. For an interval during which the motion intensity is low enough, the group of frames within the interval except the first and the last ones can be replaced by synthesized ones due to their high similarity (low RMSD). Thus, to enable this feature, we define the LoM over a group of frames (from f_i to f_k, $i<k$) for offline motion analysis as

$$LoM^2(f_i, f_k) \equiv \max_{i<j\leq k}(RMSD(f_i, f_j)) \qquad (3)$$

By definition, LoM^2 is the tight upper bound of RMSDs between the first frame in the group and all the other frames.

B. Baseline User Activities

In this section, we have four baselines for user activities to test the feasibility of RMSD as an indicator of motion intensity. The baseline user activities are some common activities in 3DTI environment with monotonic movements. Each of the baseline activities has its own motional and postural uniqueness. Thus, as a feasible indicator, distinctive RMSD values should be acquired with different baselines.

The baseline activities and their motional/postural characteristics are listed as follows. Ordered by their motion intensity from low to high, we denote them as act1 to act4.

- Act1 (Storyteller): The character is sitting in the center of the 3DTI environment where most of his action concentrates on facial area. Occasional gesture changing is expected during the activity.
- Act2 (Lecturer): The character is standing in the center of the 3DTI environment. Frequent facial movement is expected along with occasional gesture and body movement.
- Act3 (Trainer): The character is demonstrating a sequence of body movements. Slow and gross-motor movements of all body parts are expected at all times during the session.
- Act4 (Performer): The character is acting out a fluent sequence of body movements. Both posture and position of the character in the 3DTI environment are changing in a rapid but rhythmic fashion. Fast and gross-motor movements are expected at all times.

For each of the four baseline activities, a 15 seconds 3DTI clip is recorded. In Fig. 2, we plot the distributions of RMSD of adjacent frame pairs in each baseline activity. From the four distinctive histograms we can see that RMSD reflects the motion intensity of each baseline activity. For an activity of high motion intensity, its RMSD tends to be higher; for a low motion activity, the RMSD is lower.

IV. MBFS-Based Compression Schemes

A. Frame Deduction

The frame deduction is targeting an on-the-fly adjustment of the delivery of 3DTI videos (Fig. 3a). Under this scenario, our goal is to deduct real frames on the video producer site to alleviate the bandwidth consumption of 3DTI. The deducted frames can be later recovered by MBFS in the receiver site. This lossy compression of the stream will inevitably degrade the spatial quality of the video due to possible artifacts in the synthesized frames. Thus, in later investigation on the tradeoff between resource saving and quality preservation, we use an integer R to denote the ratio between real frames and synthesized frames:

$$R \equiv \frac{FPS}{SFPS} \qquad (4)$$

For example, when R=3 (Fig. 3a), one captured frame in every three frames on the producer site is omitted. The omitted frame is later replaced by a synthesized frame via MBFS on the receiver site. For on-the-fly compression, the ratio R is dynamically adjusted according to the current LoM. The mapping from LoM to suitable R will be investigated via subjective experiment in Section V.

B. *Offline Video Compression*

The offline video compression is used when the whole activity session ends and is recorded in the memory buffer of the hard drive. With the whole footage, the analysis of LoM can be more fine-grained in the sense that a group of continuous frames can be analyzed to see whether they can all be replaced by synthesized frames (Fig. 3b). If the LoM over an frame group is smaller than a predefined threshold (TH), then the group is defined as replaceable. Algorithm 1 details the procedure of offline compression with MBFS.

Algorithm 1: Offline 3DTI video compression with MBFS

```
start=0
end=start+1
repeat
    if LoM^2(f_start, f_end) > TH do
        for i=start+1..end-2 do
            discard frame_i
        end for
        start=end
        end=start+1
    else do
        end=end+1
    end if
until end == the last frame
```

The objective of Algorithm 1 is to find the largest group under the constraint of $LoM^2 < TH$. When the motion intensity (LoM^2) of a group of frames is lower than this given threshold (TH), these frames are identified as replaceable by synthesized ones. Again, TH controls the tradeoff between resource consumption and quality. For a large TH value, more frames will be discarded, which saves more bandwidth but also lowers the perceptual quality. We investigate this complication in the next section.

V. Parameter Space Analysis

In this section, we introduce a series of subjective experiments to help us find the suitable parameter settings (R and TH) that balance between resource saving and the preservation of perceptual quality.

A. *Experiment Settings*

1) *Test Cases*

The clips of the four baseline activities (Section III.B) were used as the test cases. These original clips will be referred to as the reference clips in the following discussion.

The first phase of the experiment is to investigate the relationship between perceptual quality and the proportion of synthesized frames (the ratio R). Therefore, the test cases of this phase are generated by uniformly replacing different numbers of real frames in the reference clips by synthesized ones. In total, 44 clips were generated for the four baseline activities with nine different R values ranging from two to twelve.

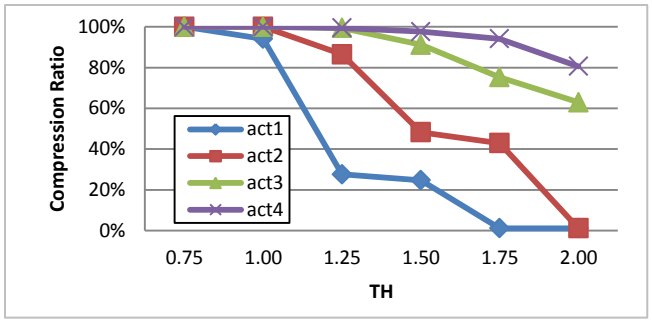


Figure 4. Compression ratios of different TH settings

The second phase of the experiment investigates the relationship between perceptual quality and the TH threshold related to the offline compression of baseline videos with different LoM. To generate the test cases, we compress four videos with TH being set at 0.75, 1.00, 1.25, 1.50, 1.75, and 2.00. The resulting compression ratio of each test case is plotted in Fig. 4. In total, 24 videos were generated for this phase.

2) *Participants and Experiment Procedures*

29 volunteers (four females and 25 males) were recruited in our subjective experiment. Their maximum, average, and minimum ages were 50, 26.3, and 20, respectively. There were two phases in the experiment and each phase was divided into four parts, with each part showing videos of only one of the baseline activities. In the beginning of each part, the reference video of the baseline activity was shown to the participant. After that, we adopt the Ascending Method of Limits [19] and shown the test cases in the order of increasing quality degradation.

After viewing each video, the participants were asked to answer the following questions:

Q1. Can you perceive any difference between the video you just saw and the reference video?

Q2. On a scale of one to five, providing the reference video being the score five, what score would you give to this video on its visual quality?

Q3. Do you consider this visual quality acceptable?

During the test, the participant could rest or request to be shown the reference video again at any time.

B. *QoE Metric*

Among the various QoE metrics that are used to quantify the satisfaction of users, the Mean Opinion Score (MOS [18]) is the most popular one. However, more works [15-17, 21] starting to point out the insufficiency of using MOS as an indicator for fine-grained experience quality. Perhaps the most criticized flaw of the MOS system is that it is an ordinal scale [1]. According to the original definition of MOS in [18], its numbering from one to five is merely a mapping to ‘bad’, ‘poor’, ‘fair’, ‘good’, and ‘excellent’, respectively. In an ordinal scale system, numbers only describe the ordering of

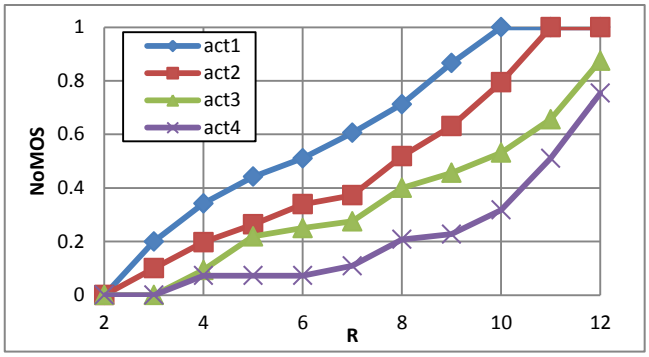
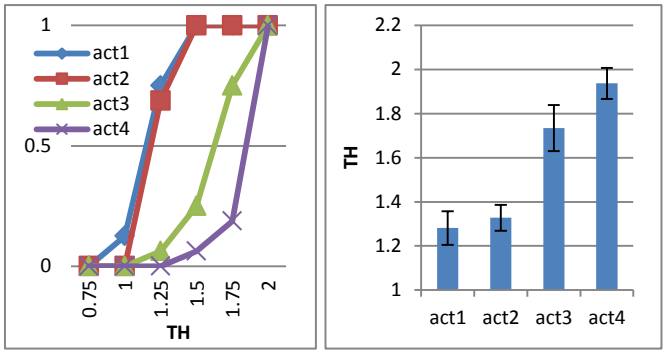


Figure 5. Relationship between R, NoMOS, and LoM



(a) Figure 6. The JNDG of TH (b)

labels but they do not stand for exact difference between the observer's perceptual magnitudes. For instance, the difference between MOS=1 and MOS=2 is not universal for every observer. Moreover, for the same observer, the difference between 'bad' and 'poor' may not necessarily equal to the difference between 'good' and 'excellent' [15]. In view of the insufficiency, recent psychophysics studies started to propose new QoE metrics that reflect the relation between two quality settings while preserving unification of perceptual magnitude represented by the same indicating value. Towards this objective, previously proposed solutions can be classified into two main categories: numerical scores [15-17] and binary thresholds [20-21]. Obtained with statistical scaling and calibration, the former can describe the relative perceptual magnitude perceived by an observer towards different quality settings. The latter, on the other hand, request the observer to answer simple yes/no questions rather than to give subjective numerical scores. Thus, the results serve as a universal boundary of quality adjustment.

In our experiment, we adopt the merits from both sides of the proposed solutions. The first and the third question we asked in the experiment are inquiries about the Just Noticeable Degradation (JNDG) threshold and the Just Unacceptable Degradation (JUADG) threshold [20]; while the second question provides the relative score of the video to the two thresholds.

After we compile the scores of a set of experiment, we normalized them into an interval between zero and one, for one being the score given to the test case at the JNDG threshold and zero being the score at JUADG. For example, if the original score given to the test cases at JNDG and JUADG are 4 and 1, respectively, then for a test case with original score equal to 3, it will be normalized to (3-1)/(4-1)=0.67.

This calibration inherits the relative characteristic from the numerical score and the universal characteristic from the binary thresholds. Thus, the calibrated score given by different users can be placed on the same ground for statistical operations. For the ease of discussion, we will refer to this calibrated score as NoMOS (not only MOS) in the following context.

C. *Results and Analysis*

1) *Frame Rate Deduction*

In Fig.5, we plot the NoMOS of 3DTI videos of the four baseline activities with gradually increasing proportion of real frames (R). We can see from the figure that for the same frame replacement setting (the same R), the NoMOS value is negatively correlated with the LoM of the video. The phenomenon is well expected. Naturally, for user activity with low motion intensity, the difference between adjacent frames will be less then high motion activities. This lower difference between frames makes the synthesis of morphing frame easier and hence near-perfect morphing frames can replace real frames without being noticed by the observers. On the other hand, for high motion videos, the higher difference between frames increases the chance of feature mismatch in morphing. In addition, the high motion content makes the observer concentrate more on the moving parts of the character's body, which makes the artifacts of morphing more detectable.

On a resource saving point of view, when targeting at the same perceptual quality (the same NoMOS), the R values increase along with the LoM. By the definition of R, the compression ratio of a stream equals to 1-1/R. As an example, for a fixed targeted perceptual quality, say NoMOS=0.5 (the middle point between JNDG and JUADG), the compression ratios are 83%, 88%, 90%, and 91% for baseline activity 1 through 4, respectively.

2) *Offline Compression*

In Fig. 6 we show the JNDG threshold against the TH compression threshold. In psychophysics, a perception threshold is defined as the intensity of stimulus (in our case, the parameter setting) which incurs positive/negative answers (in our case, do/do not perceive difference) at 50% of the times [19]. In Fig. 6a, the x-axis stands for the TH value and the y-axis stands for the proportion of observers that can perceive difference between the compressed 3DTI video and the reference. Again, the LoM plays a decisive role in parameter setting. For activities with higher LoM, the TH grows higher before the degradation becomes perceptible by the observers. This is due to the fact that, for high LoM activities, the RMSDs between frames are already high. Thus, a low TH threshold does not include many frames to be discarded by the compression and hence the degradation of quality is mild. From Fig. 6b, in which the JNDG threshold of TH values along with the 95% confidence intervals are plotted, we can obtain the suitable TH values in offline

compression for different user activities. The compression ratios obtained by applying these TH values are 43%, 52%, 82%, and 87%, respectively for the four baseline activities.

VI. ADAPTATION SCHEME

In this section, we combine frame rate deduction with a control unit based on the perceptual model constructed by the data compiled from our user study and propose an on-the-fly resource adaptor for 3DTI delivery.

A. Scenario and Overall Design

The objective of our adaptor is to balance between resource consumption on content delivery of 3DTI and the degradation on users' satisfaction caused by resource saving, in other words, an efficient content delivery. The overall design of the adaptation scheme is illustrated in Fig. 7. In the scenario we are targeting, one producer 3DTI site is delivering live visual content to an observer site. In light of the frame rate deduction via MBFS (Section IV.A), the producer site does not have to transmit each and every frame. Instead, a control unit, residing in the producer site, will grant the omission of transmission of a frame based on motion intensity of the content (LoM[1]), and the targeted perceptual quality (a given NoMOS value). The idea is to grant the omission of more frames when the content has a low motion or when we only targets low perceptual quality due to resource limitations. At the observer site, the omitted frames are recovered via MBFS technique and the frame rate of the 3DTI video is restored. Due to the fact that the movement of the user in the producer site can be constantly changing, the control unit has to adjust the omission rate in a dynamic manner. In the next section, we introduce the design of the control unit and the algorithms run on the producer and the observer sites.

B. Control Unit

The control unit in the producer site[1] is the core of the whole mechanism. It bridges the perceptual quality of the final video and the consumption of resource. In order to exploit the perceptual limitations of the human visual system, we analyze the data compiled in the investigation of frame rate deduction (Phase II, Section V.A.1). With the user data, we construct a general model that accommodates complex user activities in which the users may combine and shift between different baseline movements. In the following, we introduce the construction and the specifications of the model.

To construct our model, we first interpret the baseline activities in Fig. 5 into LoM by using the on-the-fly function of LoM (formula 2) with N equals to the total number of frames in a clip. Next, we apply polynomial regression on the data set depicted in Fig. 5 to obtain the mathematical formula of our model. This moves the three dimensional (<R, NoMOS, LoM>) data set from discrete (44 data points compiled from Phase II, Section V.A.1) to continuous space so that it can take an arbitrary pair of *<R, NoMOS>* as input. In Fig. 8, we show the result of the regression in 3D surface as well as in contour plot. The polynomial we obtained is shown as follows, with its coefficients (p_{ij}) listed in Table I:

[1] The control unit resides at the producer site because it needs to take the LoM of produced frames as one input. The second input, which is the targeted NoMOS, can be decided by the request from the receiver and the throughput between the sites. At this point we assume a fixed targeted NoMOS value for simplicity.

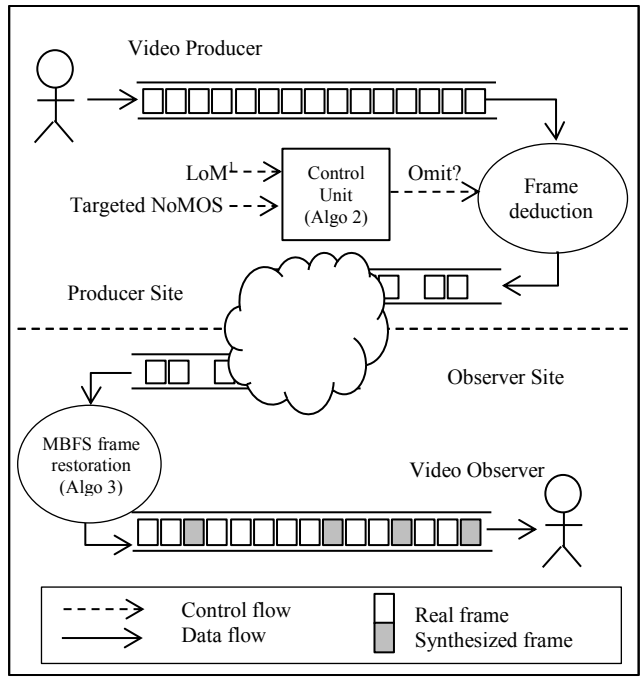


Figure 7. Overall design of rate adaptor.

$$F(R, NoMOS) = \sum_{i=0}^{4} \sum_{j=0}^{4-i} (p_{ij} \times R^i \times NoMOS^j) = LoM_{TH} \quad (5)$$

TABLE I. COEFFICIENTS (p_{ij}) OF THE POLYNOMIAL MODEL

i \ j	*0*	*1*	*2*	*3*	*4*
0	0.7063	4.203	30.38	62.33	-10.73
1	-0.06109	-6.6	-15.93	-2.948	
2	0.2512	1.603	0.9359		
3	-0.04896	-0.08435			
4	0.002387				

Goodness-of-fit: R^2=0.8767 and RMSE=0.1625

Since the model covers a continuous space (formula 5), it can accept a *<R, NoMOS >=<r,q>* pair as input, and output a LoM threshold (LoM_{TH}). This signifies that if the motion intensity of the current content is lower than this threshold, the frame rate can be reduced by replacing the captured frames with synthesized ones by the ratio of $R=r$ with the perceptual quality preserved at $NoMOS=q$.

In practice, the usage of the model is described as follows. First, the adaptor will target on a fixed NoMOS value q and guarantees that the perceptual quality will not drop below this value along its adaptation. When NoMOS is assigned a fixed constant q, formula 5 becomes two-dimensional and maps from R to LoM_{TH}:

$$F(R, q) = f(R) = LoM_{TH} \quad (6)$$

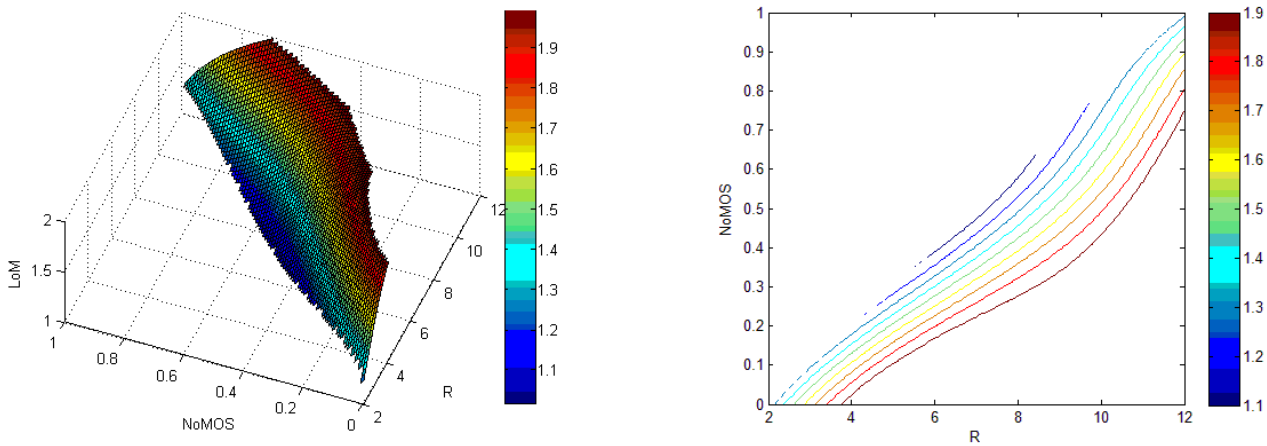


Figure 8. Polynomial regression of data compiled

Since R is a discrete value ($R \in [2,12] \cap \mathbb{Z}$), the control unit can build up a table from formula 6 which records the respective LoM_{TH} for each R. An example for q=0.60 is listed in Table II. The result shows that if one wants to apply frame rate deduction with R = 10 to a 3DTI video while preserving the perceptual quality to be no less than 0.60 NoMOS then the LoM of the current video must be less than 1.569. On the other hand, if one wants to further reduce the bandwidth by applying R=8, unless the motion intensity is lower than 1.0699, the targeted perceptual quality (0.60 NoMOS) can be compromised.

TABLE II. LoM_{TH} AND R MAPPING UNDER NoMOS = 0.75

R	8	9	10	11	12
f(R)	1.0699	1.2996	1.5769	1.8485	2.1184

C. Algorithms

1) Producer Site

The algorithm of adaptation on the producer site is shown in Algorithm 2. The input is the latest frame f_k and the targeted *NoMOS*. The output is either true or false, indicating omission or transmission of the frame. In the producer site, the control unit first constructs the LoM_{TH} table we mentioned in the previous section according to the given *NoMOS* (line 3-5). Next, to compute the current LoM, the adaptor keeps a buffer that can store up to 12 most recent frames (line 6-13). It then computes the LoM by the on-the-fly LoM function (LoM^1) in Section III.A on the acquired frames in the buffer. If the resulting LoM exceeds the LoM threshold, the frame cannot be replaced by a synthesized frame or else the *NoMOS* target cannot be preserved. Hence, the frame is sent out and a copy is kept. Otherwise, the frame is omitted and is not sent out, and the buffer is reset (line 14-19).

Algorithm 2: Adaptation on the producer site

```
1:  buffer= Ø
2:  Omit(f_k, NoMOS){
3:      for i=2..12 do
4:          LoM_TH[i]=F(i, NoMOS)
5:      end for
6:      if buffer== Ø do
7:          put f_k into buffer
8:          return false
9:      end if
10:     if |buffer|==12 do
11:         discard the oldest frame in buffer
12:     end if
13:     put f_k into buffer
14:     if LoM^1(f, |buffer|)>LoM_TH[|buffer|] do
15:         return false
16:     else
17:         buffer= Ø
18:         return true
19:     endif
    }
```

2) Observer Site

The algorithm of MBFS frame restoration on the observer site is shown in Algorithm 3. It keeps track of the latest real frame it encountered (line 7-8). When it notices a frame is skipped, it applies the MBFS technique to create the missing frame by morphing with the latest real frame and the frame it just received (line 3-5).

Algorithm 3: MBFS frame restoration on the observer site

```
1:  last= -1
2:  Receive(f_i){
3:      if last+1≠i do
4:          display morph(f,f_i)
5:      end if
6:      display(f_i)
7:      last=i
8:      f=f_i
9:  }
```

D. Crowdsourcing Performance Evaluation

In this section, we evaluate the performance of the adaptor by asking users to compare the adapted 3DTI video and the original one. We recorded a test video that contains complex user activity which combines the previously mentioned baselines (Section III.B) in order to test the performance of the LoM metric and the control unit. For evaluation on perceptual quality, we adopt the crowdsourcing methodology for user study. This way, we

can aggregate an enough amount of user data to support the result.

1) Test 3DTI Video

A 60 second 3DTI video is made as the test case for evaluation. During the 60 second footage, about 22% of the time, the patient's movement can be classified as baseline activity one; 12% can be classified as baseline activity two; and 67% is a mixture of baseline activity three and four. After processed by our adaptation scheme, the video stream is compressed by 90% of its original size. The perceptual quality we target is NoMOS=0.75.

2) Crowdsourcing Methodology

Due to the fact that our adaptation scheme aims for a straight forward goal: to cut back resource consumption without being noticed by the observer, the procedure of our user study becomes fairly simple and hence is suitable for adopting the crowdsourcing methodology [4]. First we merge the adapted 3DTI video and the original video into the same scene and upload the video to web server. Next, we set up a website that contains some brief greeting sentences and a simple instruction:

"This video consists of two clips playing side-by-side. Please watch the video and then answer the following question: ***Which of the clips has a better visual quality?****"*

Beneath the instruction, there are three buttons denoted "The one on the LEFT", "The one on the RIGHT", and "None of them is better than the other". When a viewer hits the button, she will be redirected to a thank you page; and the result will be recorded in our webserver. The link to the webpage is propagated through social networks as well as advertised via two mailing lists of University of Illinois at Urbana-Champaign and National Taiwan University.

3) Results

By the time this paper is being written, 147 users have visited the website and 92 of them have given feedbacks. 72% of the viewers failed to perceive the degradation in the adapted version of the 3DTI video.

Comparing the adapted 3DTI video and the original video, 69 of the 706 frames are chosen by our control unit to be omitted and hence need not to be transmitted, which contributes to a 10% reduction on bandwidth consumption. The omitted frames are replaced by synthesized frames as we designed in the receiver site, which brings no perceptible quality degradation for the 54% of the viewers who voted 'no difference' and the 18% of the viewers who prefer the adapted video over the original version. This shows that our adaptor can save substantial networking resource without introducing any negative effect on the experience quality.

VII. CONCULSION

In this work, we investigate the viability of applying Morphing-based Frame Synthesis on resource adaptation of 3D Tele-immersion content delivery. In view of the effect of motion of a user activity on the prominence of the artifacts introduced by MBFS, we conducted a series of subjective experiments to investigate the relationship between the level-of-motion, the quality of experience, and the parameters of compression schemes based on MBFS. In addition, we propose of a novel QoE metric: NoMOS, which combines the merits of numerical and binary metrics of QoE. The results shows that MBFS is able to achieve 43~87% offline compression ratio for baseline user activities; and a 10% on-the-fly resource saving on complex user activity with negligible quality degradation which is imperceptible by 72% of the 92 human viewers.

REFERENCES

[1] G. A. Gescheider. Psychophysics: the fundamentals. Lawrence Erlbaum Associates, Inc. 1997.

[2] P. Agarwal et al. Bundle of streams: concept and evaluation in distributed interactive multimedia environments. ISM, 2010.

[3] T. Beier et al. Feature-based image metamorphosis. In Proc. of SIGGRAPH 1992.

[4] A. Kittur et al. Crowdsourcing user studies with mechanical turk. In Proc. of CHI, 2008.

[5] Z. Huang et al. Perception-based playout scheduling for high-quality real-time interactive multimedia. In Proc. of INFOCOM, 2012.

[6] Z. Yang et al. A multi-stream adaptation framework for bandwidth management in 3D tele-immersion. In Proc. of NOSSDAV, 2006.

[7] S. Chen et al. Activity-based synthesized frame generation in 3DTI Video. ICME, 2013.

[8] W. Wu et al. Towards multi-site collaboration in 3D tele-immersive environments. In Proc. of ICDCS, 2008.

[9] Z. Huang et al. SyncCast: synchronized dissemination in multi-site interactive 3D tele-immersion. MMSys, 2011.

[10] Z. Yang et al. Enabling multi-party 3D tele-immersive environments with ViewCast. TOMCCAP, 2009.

[11] R. Vasudevan et al. High quality visualization for geographically distributed 3DTI applications. TMM 2011.

[12] K. Nahrstedt. 3D Tele-immersion for remote injury assessment. International Workshop on Socially-Aware Multimedia, 2012.

[13] W Wu et al. I'm the Jedi! - A case study of user experience in 3D tele-immersive gaming. In Proc. of ISM, 2010.

[14] A. Arefin et al. Prioritized evolutionary optimization in open session management for 3D tele-immersion. In Proc. of MMSys, 2013.

[15] S. Tasaka et al. Psychometric analysis of the mutually compensatory property of multimedia QoS. ICC, 2003.

[16] S. Egger et al. Quality-of-experience beyond MOS: experiences with a holistic user test methodology for interactive video services. In Proc. of 21st ITC specialist seminar on multimedia applications, 2010.

[17] Y.-C. Chang et al. Radar Chart: scanning for satisfactory QoE in QoS dimensions. IEEE Network, 2012.

[18] ITU. Methods for subjective determination of transmission quality. P.800. 1996.

[19] E. B. Goldstein. Sensation and perception. Cengage, 2010.

[20] W. Wu et al. Color-plus-depth level-of-detail in 3D tele-immersive video: a psychophysical approach. MM, 2011.

[21] W. Song et al. Saving bitrate vs. pleasing users: where is the break-even point in mobile video quality. MM 2011.