# Exploiting Contextual Information to Enable Efficient Content Delivery for 3D Tele-Immersion Applications

Shannon Chen
University of Illinois at Urbana-Champaign
cchen116@illinois.edu
Advised by Klara Nahrstedt

## ABSTRACT
The tradeoff relationship between resource requirement, content complexity, and user satisfaction is magnified when more and more modern 3D Tele-immersive (3DTI) applications with higher quality demands and/or scalability requirements come into the picture. These demanding applications introduce challenges in different phases throughout the delivery chain of 3DTI systems. To tackle them, we propose to exploit contextual information of 3DTI systems, such as contextual information on resource, content, and satisfaction aspects. Understanding contextual information will improve the utilization of different computing environments to fulfill the objectives of targeted application.

## Categories and Subject Descriptors
H.4.3 [**Information Systems**]: Information Systems Applications – *Communication Applications*

## Keywords
3D Tele-immersion, Contextual information.

## 1. INTRODUCTION
3D Tele-immersion (3DTI) technology allows full-body, multimodal interaction among geographically dispersed users, which opens a variety of possibilities in cyber collaboration. Users of 3DTI systems are captured by 3D camera array surrounding their user spaces along with other application-specific sensors. The captured 3D models with 360° coverage of users' bodies are put into a shared virtual space where activities such as art performance, exergaming, and physical rehabilitation happen.

However, with its great potential, the resource and quality demands of 3DTI rise inevitably, especially when some advanced applications target resource-limited computing environments (e.g., mobile or home environments) with stringent scalability requirements (e.g., multi-site interaction or large-scale broadcasting). Under these circumstances, the tradeoffs between (1) resource requirements, (2) content complexity, and (3) user satisfaction in delivery of 3DTI are magnified. Intuitively, when resource budget is low and content complexity is high, user's perceptual quality will be sacrificed. For example, watching live broadcasting 3DTI sport events on limited bandwidth budget incurs bad viewing experience. On the other hand, when the computing resource is scarce and the service quality is demanded, the system can only support low complexity activities as its content. For example, a 3DTI client running on a smart phone over 3G network will not support telehealth applications.


MM'15, October 26-30, 2015, Brisbane, Australia
ACM 978-1-4503-3459-4/15/10.
http://dx.doi.org/10.1145/2733373.2807994

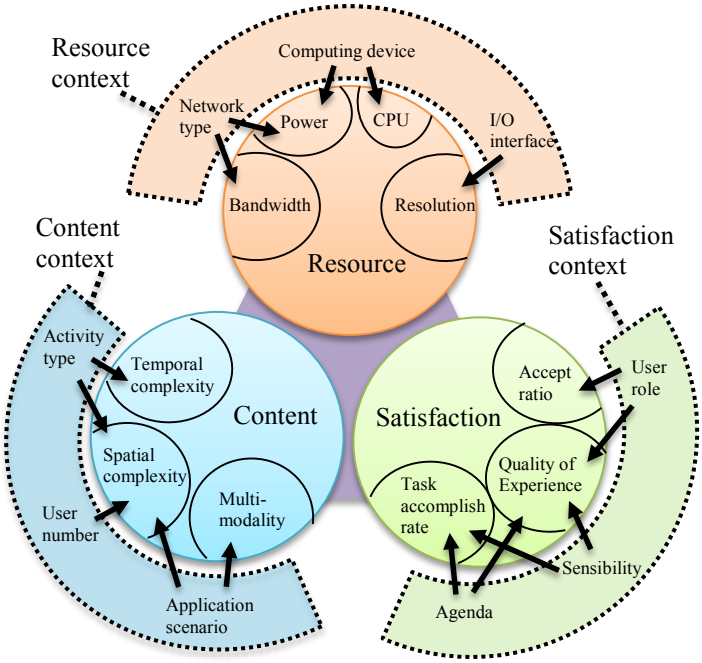


**Figure 1. Contextual information.**

In this proposal, we argue that these tradeoffs of 3DTI applications are actually avoidable when the underlying delivery chain takes contextual information into consideration. We propose the concept of *contextual information* in 3DTI, which spans across the three factors: resource, content, and satisfaction in 3DTI systems. With contextual information, we expect 3DTI systems to be able to (1) identify the characteristics of its computing environment to allocate computing power and bandwidth to delivery of prioritized contents, (2) pinpoint and discard the dispensable, unnoticeable details in content capturing according to properties of target application, and (3) differentiate contents by their contributions on fulfilling user's expectation so that the adaptation module can allocate resource budget accordingly. With these capabilities we can change the tradeoffs into synergy between resource, content, and user satisfaction. For example:

- **In 3DTI capturing**, when delivering activity involving only little user movement (e.g., storytelling), the system could tune down the temporal resolution of video capturing to save extra bandwidth without degrading perceptual quality and vice versa.
- **In 3DTI dissemination**, when the preferences of viewer towards each 3DTI performer are acquired by the system, streams can be prioritized on their delivery to reach maximum service satisfaction under bounded bandwidth.
- **In 3DTI receiving**, when targeting on mobile devices with limited power and display resolution, computation-intensive render functions should be automatically offloaded.

We implement contextual-information-aware 3DTI systems to verify the performance gain on the three phases in 3DTI systems' delivery chain: capturing phase, dissemination phase, and receiving phase. The implemented systems are tested with various applications including gross/fine motor activities (e.g., physical rehabilitation/storytelling); high/low motion user movements (e.g., exergaming/lecturing); and small/large scale applications (e.g., telehealth/broadcasting).

To sum up, in this proposal, we aim to change the tradeoff between resource, content, and satisfaction in 3DTI by exploiting the contextual information about the three factors. We expect the proposed mechanisms will enhance the efficiency of 3DTI systems targeting on serving different user activities, and applications with different computational requirements and scales.

# 2. CHALLENGES & PREVIOUS WORKS

A delivery chain of 3DTI lies behind the media-enriched cyber collaborations to cope with real-time constraint of interaction, processing complexity of graphic rendering, and heavy load of data transmission. The delivery chain can be broken down into three phases: capturing, dissemination, and receiving.

**Capturing phase and its challenges.** Equipped with 3D camera array, microphone, and other application-specific sensors, a 3DTI site captures the activity of its user and digitize it to enable full-body, free viewpoint experience. However, as pointed out by many previous works [1][2][3] on 3DTI capturing interfaces, bitrate of the captured content bundle (i.e., the aggregation of heterogeneous content streams) is oftentimes too high to be supported by common networking environments. An empirical calculation provided by Yang et al. [1] reports a 300 Mbps bitrate for visual streams in 3DTI with minimum quality setting (320x240 resolution, 10 fps). A more modern hardware setting in interface proposed by Mekuria et al. [2] introduces an even higher 1,032 Mbps bitrate with Kinect camera (assuming 640x480 resolution, 30 fps) without compression. Yet, advanced applications of 3DTI such as event broadcasting [4], remote healthcare [5][6], and mobile communication [7] all picture modest CPU/GPU and low bandwidth budget in their computing environments in order to enable them on home or mobile devices.

Thus, in the capturing phase of 3DTI, the tradeoffs happen between outbound bandwidth of the capturing site (resource); temporal/spatial resolutions of the captured stream (content); and the perceptual quality of streams (satisfaction). It is necessary to identify the dispensable details in the content bundle to reduce the total bitrate to a manageable level.

**Dissemination phase and its challenges.** On dissemination of the captured content, stream bundles captured by multiple performer sites (3DTI sites that capture their users' activity) are exchanged and shared between viewer sites (all participating site regardless of whether its user is been captured) via P2P network. With the received content, a synchronous, shared virtual space is rendered locally in each site. In this process, the first challenge comes from 3DTI's multi-view characteristic. The multi-source (multi-performer), multi-content (multi-camera) dissemination becomes a content distribution forest construction problem in the P2P network formed by all participating sites. Multiple viewers subscribing to multiple streams from multiple performers introduces massive bandwidth consumption, especially when the user number scales up. Platform for 3DTI broadcasting envisioned by Ahsan et al. [4] aims to serve 1,000 concurrent viewers. The system is estimated to consume up to 6 Gbps out-bound bandwidth of the content sources. The second challenge is efficient delivery of the multi-view content. Streams produced by all performer sites are not equally important in a viewer's rendering. Yang et al. [1] first identifies the relationship between a camera stream's shooting angle and its contribution to the field of view (FOV) of a viewer. The work identifies the importance of content differentiation in the dissemination phase. However, in multi-site settings [4][8], Yang's dissemination only brings marginal (<5%) improvement on scalability.

Thus, in the dissemination phase of 3DTI, the tradeoffs happen between bandwidth/delay restrictions of the dissemination network (resource); size of the stream bundle and user group (content); and the acceptance ratio for user subscriptions (satisfaction). We need an efficient prioritization scheme to enable efficient utilization of the overlay network to create a manageable environment for large-scale applications.

**Receiving phase and its challenges.** Depending on the characteristic of application, a 3DTI site may receive the content bundle for immediate display (for live interactive applications) or for storage and later retrieval (for asynchronous on-demand viewing). For these different purposes, the challenges of receiving phase lie in (1) efficient storage, (2) review summary generation, and (3) adaptive rendering. On efficient storage, Mekuria et al. [2] proposed a mesh-based compression scheme for 3DTI content which reaches 1:10 compression ratio. However, this is still far from being comparable to video codecs for conventional 2D content (e.g., 1:100 for MPEG). On review summary generation, Jain et al. [9] propose a metadata analysis module in their 3DTI system for activity recognition. Yet, training and tuning phase of the module make it impractical for most applications in home environment. On adaptive rending, previous 3DTI clients proposed are all designed for specific computing environments (mobile: [7], PC: [3]) and platforms (Windows: [3], Linux: [1], Android [7]). However, in modern use cases, many applications require the flexibility to run under different resource limitations and the capability of crossing platforms.

Thus, in the receiving phase of 3DTI, the tradeoff happen between power/bandwidth limits of the client (resource); total size of the streamed data (content); and controllability/accessibility of service (satisfaction). We need an adaptive, cross-platform client and a database entity with efficient storage and retrieval features.

# 3. PROPOSED APPROACHES

To tackle these identified challenges, we propose the concept of contextual information, which spans across resource, content, and satisfaction. We argue that, a 3DTI system should reference the contextual information of these three factors throughout its delivery chain in order to avoid the tradeoff situation between them. In the following sections, we formally define the scope of contextual information and then we discuss where and how different contextual information will help improving each phase in the delivery chain of various 3DTI systems of different scales and purposes.

## 3.1 Contextual Information

As illustrated in Figure 1, contextual information spans across resource, content, and satisfaction. Thus, we start from formalizing the scope of these three factors.

- Resource: The computing resources that can be utilized for delivering 3DTI contents. This covers the capabilities of computing device and I/O devices in each 3DTI site as well as the network connecting them. For example: network bandwidth, display resolution, CPU/GPU capability, and power limitation.

- Content: The complexity of the captured content bundle throughout an application session. This covers the complexity of the bundle itself as well as the complexity of each data stream inside the bundle. For example: number of heterogeneous streams inside the bundle (i.e., multimodality), temporal resolution of bio-sensing streams, and spatial variance (i.e., frame difference) of 3D video stream.
- Satisfaction: The achievement level of application's objectives and user's expectation. The scope of satisfaction is application-dependent but it can cover both subjective opinions and objective metrics. For example: quality of experience, subscription acceptance ratio, and task accomplishment rate.

With the scopes formalized, we define contextual information as:

*"The aggregation of information about the three factors - resource, content, and satisfaction - of a 3DTI system and its application."*

For example, contextual information of resource includes the targeting environment of the 3DTI application (e.g., home, mobile, or dedicated facilities). From the environment, we can infer the connection type (e.g., 3G, home cable network, or dedicated optical connection), computation budgets (e.g., CPU rate, RAM size) and I/O interfaces (e.g., desktop screen or phone screen). This help the underlying system adapts its content complexity to avoid wasting unnecessary resource, or offloads/un-offloads computation intensive functions to more powerful entities in the delivery chain. Similarly, contextual information of content and satisfaction (illustrated in Figure 1) helps the underlying system to enhance the efficiency on utilizing the computing environment to fulfill the objectives of users. In the following sections, we will introduce our design which utilizes different contextual information in each phase of the delivery chain of 3DTI systems. In the rest of this proposal, we use the terms resource context, content context, and satisfaction context for contextual information of resource, content, and satisfaction, respectively.

## 3.2 Content Context in Capturing Phase

**Preface.** From user studies of Wu et al. [10] and Schulte et al. [11], we see that 3DTI users can have very different quality preferences when participating in different user activities. Thus, in the capturing phase of 3DTI system, we propose the Activity-Aware Adaptive Compression (A3C) [12], which is a real-time compression scheme that utilizes content context to dynamically adjust its compression ratio and quality.

**Solution.** A3C adopts the Morphing-based Frame Synthesis (MBFS) [13] as its core function. Taking advantage of the unique properties of 3DTI scenes, MBFS allows content receiver to boost up video framerate by injecting synthesized frames. Thus, the capturing frame rate of the performer site can be reduced to save substantial bandwidth for content delivery.

To maximize the effectiveness of MBFS compression, we apply activity recognition to identify content context: the motion characteristics of current user activity. We assign compression settings according to the activity type to adjust the compression ratio and spatial quality of captured content. The quality of synthesized frames depends heavily on the motion level of the activity. For high motion activities, the graphical difference between frames is large and the sharpness degrades due to motion blur. These effects increase the artifacts introduced by feature-based morphing, making the synthesis unnatural and noticeable to viewers. Therefore, we propose to combine the compression scheme with user activity recognition to dynamically tune the compression and achieve quality preservation and resource saving.

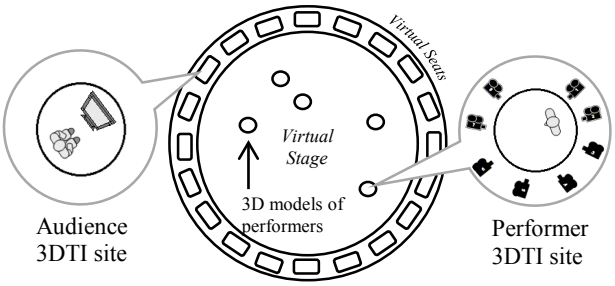


**Figure 2. 3DTI Amphitheater.**

**Verification.** To verify our design, we implement the A3C scheme on our 3DTI platform: the TEEVE Endpoint [3], which includes a general purpose capturing interface that does not target specific application. We will evaluate A3C via objective compression ratio as well as subjective user study. Objective evaluation will focus on the bandwidth saving (resource factor) with A3C under different application scenarios, e.g., gaming, exercise, lecture, and storytelling. Subjective evaluation will be conducted via interviewing player of our 3DTI virtual fencing game built upon the TEEVE Endpoint with the A3C module embedded (satisfaction factor).

## 3.3 Satisfaction context in Dissemination

**Preface.** Due to the high bandwidth demand in dissemination phase, previous 3DTI applications are restricted with a small user group (fewer than four [1]). In order to promote 3DTI to applications with more flexible scalability, we propose the 3DTI Amphitheater, a broadcasting platform which renders a shared virtual space (Figure 2) that accommodates hundred-scale users.

**Solution.** Users in the 3DTI Amphitheater are divided into two groups: performers and the audience. A user is a performer when she is captured by a camera array. The 3D models of performers interact on the virtual stage while actual users can be physically dispersed. A user is in the audience when she passively view the 3D streams created by the performers without involvement.

We propose to tackle the challenge of high bandwidth demand by content differentiation. We argue that not all streams in the amphitheater are equally important. Content's importance to a viewer can be affected by two types of satisfaction context: view-based priority and role-based priority. For example, when the shooting angle of the visual content complies with the chosen view of a user, the stream should have higher view-based priority to her. We propose to adopt the contribution factor (CF) [1] to quantify the view-based importance of streams. As for role-based priority, we argue that: to a particular viewer, not all performers are equally important. The importance of a particular performer depends on her role in the performance. For example the audience may be more interested in the vocalist of a rock band. Also note that the importance of a performer is not universal. A viewer who is an audience and a viewer who is a performer have different viewing preferences. The amphitheater constructs its content dissemination forest based on these satisfaction context to allocate bandwidth to the prioritized streams.

**Verification.** We propose to verify the performance of 3DTI Amphitheater by emulation with real-world network topology and the configurations of our 3DTI platform. The verification is two-fold. First, we evaluate the effectiveness of our semantic stream

prioritization by examining the application quality of service (AQoS) of user sites. The AQoS is defined as a weighted acceptance ratio of stream subscription request, where the weight of a request is the semantic priority. Second, we investigate the bandwidth saving brought by content dissemination scheme in the amphitheater. We compare its bandwidth usage (resource factor) and its sustainable performer crew size (content factor) with other multi-site/broadcasting platforms [4][8].

## 3.4 Resource context in Receiving Phase

**Preface.** Challenges in the receiving phase are (1) efficient storage, (2) review summary generation, and (3) adaptive rendering. To approach them, we focus on 3DTI application with asynchronous service model (i.e., on-demand service for offline reviewing) because its storage and retrieval features make these challenges inevitable in the receiving phase. As a platform for investigating resource context in the receiving phase, we implement an asynchronous 3DTI system for physiotherapy session reviewing: the CyPhy (Cyber-Physiotherapy) system.

**Solution.** The envisioned use case model of CyPhy is as follows. As part of the prescription given in the end of at-clinic face-to-face meeting between therapist and patient, a "CyPhy kit" will be provided to the patient. The kit includes required devices (e.g., Kinect cameras) for the patient to set up a light-weighted 3DTI recording studio at home. On a daily basis, CyPhy will stream to the patient a pre-recorded exercise demonstration prescribed by the therapist. Patient will follow the video to conduct correct therapeutic exercises and have this rehabilitation session recorded with the CyPhy kit. The recorded session is upload to an electronic health record (EHR) cloud to be archived. Recorded sessions will be played out by the therapist whenever and wherever she is available on mobile device or PC. Therapist can supervise the correctness of patient's moves by viewing the streamed content bundle and provide professional feedbacks.

On efficient storage in the EHR cloud, we propose a new compression scheme for 3D videos that reaches a compression ratio close to 2D video codecs. We identify the likeliness of everyday rehabilitation video and exploit this property for inter-3D-video coding. On review summary generation, we propose to analyze the metadata of inter-video coded frame sizes to detect anomaly events (i.e., patient fall or injured). This non-intrusive analysis shortens the detection time, which is a crucial improvement due to the large size and quantity of 3DTI recordings in the EHR cloud. On adaptive rendering, CyPhy references the resource context for bitrate adjustment and offloaded rendering. First, we propose to adopt the DASH [14] standard in CyPhy client to accommodate with different network capabilities in heterogeneous environments. Second, to enable different computing environments (e.g., PC, mobile) as CyPhy clients, we propose an offloading mechanism to offload the 3D rendering to DASH server. These features allow the CyPhy system to adaptively adjust its streaming content complexity and provide different levels of user control.

**Verification.** We generate a series of rehabilitation session recordings in our lab as our dataset to test CyPhy's archiving feature. Preliminary results show our storage compression scheme achieves 1:1255 compression ratio. The same dataset will be used in the verification of metadata-based anomaly detection. For adaptive rendering, the DASH-based design will be verified against bandwidth fluctuations under different network environments. The offloading feature will be verified on mobile devices with power limitation. We expect CyPhy will adjust its streaming quality and user controllability according to bandwidth dimension and power dimension of resource context.

# 4. CONCLUSION

The core idea of this proposal is to leverage the contextual information in resource, content, and satisfaction throughout the three phases of delivery chain of 3DTI applications. In summary our expected contributions are:

1-1. Identifying the dispensable details in content capturing phase according to content context.
1-2. Validating the utilization of content context in capturing phase by implementation of an adaptive compression module.

2-1. Enabling efficient utilization of the overlay network in dissemination phase by differentiating streams according to satisfaction context.
2-2. Recognizing the satisfaction context in dissemination phase by validating a manageable large-scale 3DTI platform.

3-1. Proposing 3DTI client which adapts with the resource context in the receiving phase.
3-2. Devising an asynchronous 3DTI system which realizes the adaptation to resource context in bandwidth and power.

The expected contributions combined will lead to 3DTI frameworks that are capable of sustaining advance applications with higher quality and scalability requirements. Preliminary results of our ongoing researches indicate optimistic expectations on various aspects (e.g., scalability [15][16], compression [13][17], activity recognition [9][12]). The outcomes thus far persuade us to continue on extending the utilization of contextual information and on verifying them in actual 3DTI prototypes.

# 5. REFERENCES


[1] Z. Yang, W. Wu, and K. Nahrstedt et al., "Enabling multi-party 3D tele-immersive environments with ViewCast," ACM TOMM, 2010
[2] R. Mekuria, M. Sanna, S. Asioli et al., A 3D Tele-Immersion System Based on Live Captured Mesh Geometry, ACM MMSys 2013
[3] P. Xia and K. Nahrstedt, "TEEVE endpoint: towards the ease of 3D tele-immersive application development", ACM MM, 2013
[4] A. Arefin, Z. Huang, K. Nahrstedt et al., 4D TeleCast: towards large scale multi-site and multi-view dissemination of 3DTI contents, IEEE ICDCS, 2012
[5] J. J. Han, G. Kurillo, R. T. Abresch, E. de Bie, A. Nicorici, R. Bajcsy, Upper extremity 3D reachable workspace analysis in dystrophinopathy using Kinect, Muscle Nerve. 2015.
[6] D. Sonnenwald, H. Söderholm, G. Welch, Illuminating collaboration in emergency healthcare: paramedic-physician collaboration and 3D telepresence technology. Information Research, 2014
[7] S. Shi, K. Nahrstedt, R. Campbell, A Real-Time Remote Rendering System for Interactive Mobile Graphics, ACM TOMCCAP, 2012.
[8] W. Wu, Z. Yang, and I. Gupta et al., "Towards multi-site collaboration in 3D tele-immersive environments," ICDCS, 2008
[9] A. Jain, A. Arefin, R. Rivas, et al. 3DTI Activity Classification Based on Application-System Metadata, ACM Multimedia, 2013.
[10] W. Wu, A. Ahsan, G. Kurillo et al., Color-plus-depth level-of-detail in 3D tele-immersive video: a psychophysical approach", MM 2011
[11] S. Schulte, S. Chen, and K. Nahrstedt, "Stevens' Power Law in 3D tele-immersion: towards subjective modeling of multimodal cyber interaction", Proc. ACM MM, 2014
[12] S. Chen, P. Xia, and K. Nahrstedt, "Activity-aware adaptive compression: a morphing-based frame synthesis application in 3DTI", Proc. ACM MM, 2013
[13] S. Chen and K. Nahrstedt, "Activity-based Synthesized Frame Generation in 3DTI Video", IEEE ICME, 2013.
[14] ISO/IEC 23009-1, Information technology -- Dynamic adaptive streaming over HTTP (DASH), 2014.
[15] S. Chen, K. Nahrstedt, and I. Gupta, "3DTI Amphitheater: a manageable 3DTI environment with hierarchical stream prioritization", Proc. ACM MMSys, 2014
[16] S. Chen, Z. Gao, K. Nahrstedt, and I. Gupta, "3DTI Amphitheater: Towards 3DTI Broadcasting", ACM TOMM, 2015
[17] S. Chen, K. Nahrstedt, "Impact of morphing-based frame synthesis on bandwidth optimization for 3DTI video", Proc. IEEE ISM, 2013