Learning to Wirelessly Deliver Consistent and High-Quality Interactive Panoramic Scenes

Juaren Steiger Xiaoyi Wu Bin Li
Department of Electrical Engineering, The Pennsylvania State University, University Park, PA, USA
Email: {juaren.steiger, xfw5193, binli}@psu.edu

***Abstract*—Wireless interactive panoramic scene delivery imposes unique challenges compared to its wired or non-interactive counterparts. The wireless channel is throughput-constrained, which limits the ability to deliver large and high-quality panoramic images. On the other hand, the interactivity imposes a real-time constraint on the system and limits the use of a playback buffer. Also, wireless inputs are not delivered instantaneously like wired interrupt-based inputs, so the system must predict the user's head pose and the portion of the scene visible to them, called the viewport. This reveals a tradeoff: delivering a portion too small may not cover the viewport if the prediction error is too large, while delivering a portion too large may result in a failed wireless transmission. Likewise, delivering the portion at too high a quality may result in a failed transmission. Despite these challenges, we would like to guarantee an immersive experience for the user by delivering a high-quality and visually consistent panoramic scene. To that end, we aim to maximize the user's quality of experience, which we define as the combination of (1) cumulative quality, (2) long-term consistency, which quantifies the overall variance in perceived quality, and (3) short-term consistency, which quantifies abrupt quality changes. We formulate this problem as a risk-averse multi-armed bandit problem with reward-dependent switching costs, and develop a novel block-based UCB algorithm with opportunistic switching. We derive its theoretical regret upper bound, which matches results in prior work, and corroborate this result in trace-based simulations using a panoramic video streaming data trace.**

***Index Terms*—Interactive panoramic scene delivery, risk-averse multi-armed bandit, bandits with switching costs.**

I. Introduction

Interactive panoramic scene delivery, which encompasses transformative technologies such as virtual reality (VR), mixed reality, and volumetric videos, enables immersive experiences across a wide range of applications, from entertainment to remote collaboration and education. Many applications of interactive panoramic scenes are best experienced wirelessly, with an edge server streaming panoramic content to the user and the user sending their *head pose* (position and orientation) and other inputs to the edge server over a wireless channel. However, delivering high-quality panoramic content wirelessly poses significant technical challenges compared to wired or non-panoramic digital content. To maintain real-time interactivity, the system needs to operate at a high framerate, typically 60 frames-per-second, with hard packet deadlines. Also, the user's device (e.g. VR headset) cannot use a playback buffer, as the interactive scene may need to be updated in real time according to input received from the user.

To avoid motion sickness when using a VR headset, there must be zero mismatch between the user's actual head motion and their perceived virtual head motion. This means that the edge server cannot use the user's head pose received from the previous frame to calculate the portion visible to them, called the *viewport*, for the current frame. Instead, the edge server must deliver a portion of the panoramic scene large enough to cover the user's viewport for the current frame.

However, successfully delivering the viewport is not as simple as sending the entire 360° scene, which requires a bandwidth around 4 to 6 times greater than that of traditional video streaming [1], and may result in a failed viewport delivery due to packet loss. For the same reason, sending the frame at the highest visual quality level, i.e. sending the uncompressed frame, may result in a failed viewport delivery. Instead, the edge server should intelligently select a delivery portion sufficient to account for the user's head motion and encode at a quality level that both enhances the visual quality and results in a successful transmission.

Despite these challenges, it is essential to maximize the visual quality perceived by the user in order to enhance their immersive experience. On the other hand, maintaining visual consistency is also crucial for sustaining immersion [2]. Visual inconsistency, especially with abrupt changes, is easily noticeable and distracting to the user [3]. For example, a consistent drop in quality may be tolerable if gradual, but abrupt quality shifts are disorienting and immersion-breaking. Therefore, both the visual quality and visual quality consistency must be carefully managed to guarantee the best user experience. While several studies have examined visual variability in both traditional and panoramic video streaming [4]–[6], they have largely overlooked the impact of abrupt visual quality changes.

To address this gap, our work aims to optimize the user's *Quality of Experience* (QoE), quantified as the combination of cumulative visual quality, *long-term* visual consistency, and *short-term* visual consistency. Here, long-term visual inconsistency refers to the variance of perceived visual quality, while short-term visual inconsistency refers to the instantaneous change in quality between consecutively delivered frames. The main challenges in maximizing QoE arise from two key uncertainties: (1) the probability of successfully transmitting a portion encoded at a particular quality level is unknown due to

This work has been supported in part by NSF under the grants CNS-2152610, CNS-2152658, and the ARO Grant W911NF-24-1-0103.

ISBN 978-3-903176-73-7 © 2025 IFIP

the time-varying wireless channel, and (2) the probability that a transmitted portion can fully cover the user's viewport is also unknown, as predicting head motion is inherently inaccurate. To address these challenges, we need to adaptively select both the portion to be delivered and its quality level to maximize the user's QoE. Existing studies (e.g., [5], [7]–[9]) tend to rely on heuristic-based algorithms to select the portion and quality rather than providing rigorous theoretical analysis.

To overcome these limitations, we formulate the problem of delivery portion and quality selection as a *risk-averse multi-armed bandit* (MAB) problem [10], [11] with a new type of *reward-dependent switching cost*. In this formulation, each pair of delivery portion and quality level corresponds to an arm. Maximizing the user's QoE can also be viewed as minimizing the *QoE regret*, which quantifies the gap in QoE compared to the optimal policy. To that end, we develop an algorithm that uses mean-minus-variance UCB estimates [11] in a block-based arm selection regime with opportunistic arm switching. This approach allows us to simultaneously identify the arm with the best quality and long-term consistency tradeoff with the UCB estimates, while controlling short-term consistency with the block structure. Before listing our contributions in detail, we review prior work on interactive panoramic scene delivery, risk-averse bandits, and bandits with switching costs.

A. Related Work

1) Interactive Panoramic Scene Delivery: Panoramic scene delivery requires significantly higher network bandwidth than traditional video streaming. To enhance user experience, researchers have explored various strategies for optimizing panoramic scene transmission (e.g., [8], [9]). In [9], the authors proposed an adaptive algorithm that selectively delivers only the portion of the panoramic scene corresponding to the user's predicted viewport, effectively reducing bandwidth consumption. On the other hand, various studies (e.g., [12]) explored how the compression level affects the visual quality. However, all of them utilized heuristic algorithms without theoretical analysis. Subsequently, existing studies attempted to apply the MAB framework to the panoramic scene delivery problem to give more theoretical insights. Some recent work has attempted to intelligently leverage the two separate feedback signals of motion prediction and wireless transmission. For example, [13] proposed Two-Level Thompson Sampling to select the portion to maximize the system throughput, and [14] proposed Two-Level KL-UCB to maximize the system throughput for panoramic scene delivery. [15] consider multi-user panoramic scene delivery as an application of their regular and fair bandit learning framework. However, none of the aforementioned works considered the quality variance in the application of panoramic scene delivery, which is important to guarantee an immersive experience to users.

2) Risk-Averse Bandits: Unlike the traditional MAB problem, which focuses solely on maximizing the expected cumulative reward over a finite horizon, risk-averse MAB (e.g., [10], [11], [16]–[18]) aims to mitigate reward uncertainty (i.e., risk). A widely used risk measure is the mean-minus-variance criterion [19]. The study in [10] was among the first to incorporate the mean-minus-variance metric into the MAB framework, introducing the MV-UCB algorithm. Building on this, Vakili and Zhao [11] demonstrated that MV-UCB achieves order-optimality with respect to the instance-dependent lower bound in terms of mean-minus-variance. More recently, [16] applied Thompson Sampling to risk-averse MAB and established its asymptotic tightness under specific regimes. Subsequently, researchers have increasingly explored the application of risk-averse bandit frameworks in practical scenarios. For example, [20] introduced a risk-aware contextual MAB approach to optimize beam selection at base stations to enhance vehicular communication reliability in dynamic, non-stationary conditions. However, the mean-minus-variance metric only measures reward fluctuations over a given time horizon and does not capture sudden, instantaneous changes in rewards as a component of risk. This limitation is particularly critical in panoramic scene delivery, where abrupt variations can significantly disrupt the immersive experience. Despite its relevance, existing studies have largely overlooked the impact of such abrupt reward fluctuations within the mean-minus-variance framework.

3) Bandits with Switching Costs: The bandits with switching costs problem, where the player suffers a penalty for switching arms between rounds, was first introduced by Agrawal et al. [21] who developed a block-based UCB algorithm that achieves optimal instance-dependent logarithmic regret. Block-based algorithm design has since proven very popular in bandit problems with switching costs. For example, the authors in [22]–[24] all use block-based algorithms for their wireless networking and mobile computing applications. A classic result [25] shows a $\widetilde{\Omega}(T^{2/3})$ instance-independent lower bound in the adversarial reward regime. Recently, [26] developed a block-based algorithm for the best-of-both worlds (adversarial and stochastic) regime that achieves optimal instance-independent regret in the adversarial setting. The work [27] improved the block-based algorithm design of [26] and extended the lower bound from [25] to the case of stochastically-constrained rewards, which is more general than i.i.d. rewards, but more narrow than adversarial rewards. The work [28] characterize the $\widetilde{\Omega}(T^{2/3})$ to $\widetilde{\Omega}(T^{1/2})$ tradeoff from bandit feedback to full feedback in this problem. While to our knowledge, no work has yet considered switching costs and variance control simultaneously, some authors have simultaneously considered switching costs and other performance metrics, e.g. arm-selection constraints [29], [30].

B. Our Contributions

Our main contributions in this paper are as follows:

1) We formalize the problem of maximizing the QoE of a user experiencing interactive panoramic content wirelessly as a risk-averse MAB problem with reward-dependent switching costs, aiming to maximize cumulative visual quality while ensuring both long-term and short-term visual consistency.

2) To solve this problem, we propose the ***L**earning **O**ptimistically with **B**lock-based **O**pportunistic **S**witching* (LOBOS) algorithm, a block-based UCB algorithm with *opportunistic switching* that simultaneously maximizes the mean-minus-variance metric and minimizes the reward-dependent switching cost.
3) We establish a theoretical upper bound on the QoE regret (Theorem 1) by adapting the analysis of MV-UCB [11] to the block-based design, with novel analytical techniques to account for the opportunistic switching. Our bound shows that LOBOS accurately generalizes the MV-UCB algorithm in that it matches the optimal instance-dependent bound derived by Vakili and Zhao [11] as the switching cost diminishes, and matches bounds from prior work on risk-averse bandits and bandits with switching costs otherwise (see Remark 1).
4) We validate our theoretical insights by conducting experiments on a real-world dataset of user head motion and panoramic video. These experiments show that our algorithm surpasses state-of-the-art algorithms from the literature when the three-way tradeoff in the QoE metric is tuned in various ways.

The rest of this paper is organized as follows. In Section II, we introduce the high-quality and consistent viewport delivery problem. In Section III, we present the LOBOS algorithm, and in Section IV, we give its QoE regret upper bound. In Section V, we present our trace-based simulations comparing LOBOS against other state-of-the-art algorithms from the literature. Finally, in Section VI, we conclude the paper and discuss future research directions.

II. Problem Setting

Consider an edge server delivering an interactive panoramic scene over a wireless channel to a user equipped with a VR headset. The panoramic scene can be thought of as video content being projected inside a sphere surrounding the user's head, with their viewport occupying a portion of the surface. When the user moves through the scene, the content projected inside the sphere changes, and when the user rotates their head, the viewport's location on the surface of the sphere changes.

Using a map projection technique, such as *equirectangular projection*, we can transform the sphere representing the panoramic scene to a 2D grid of *tiles*. Then the user's viewport will be covered by some subset of these tiles. The edge server needs to deliver the content corresponding to this subset of tiles, called the *delivery portion*, to the user for the frame corresponding to each timeslot $t = 1, 2, \ldots$ To do so, the edge server needs to encode the video content in the delivery portion and send it to the user as a sequence of UDP packets.

The interactivity of the user adds a real-time constraint to the system, and therefore we assume that all packets sent from the server to the user that are not received within the timeslot are permanently lost and never retransmitted. However, we assume the small packets sent from the user to the server containing transmission acknowledgments, head poses, and other inputs add negligible congestion and are never lost.

A. Motion Prediction and Delivery Portion Selection

The user sends their head pose to the edge server in each timeslot t. However, due to motion sickness, the user cannot tolerate a mismatch between their actual head motion and their perceived virtual head motion, meaning that the edge server cannot use the head pose feedback from timeslot t to compute the viewport content for timeslot $t + 1$. Instead, the edge server uses the viewport location from timeslot t to *predict* the viewport location for timeslot $t + 1$.

In order to account for the motion prediction error, the edge server should choose a delivery portion that is strictly larger than the minimal subset of tiles that covers the predicted viewport. We assume there are N possible delivery portions to choose from and let $\widehat{n}(t) \in [N]$ denote the delivery portion chosen by the edge server in timeslot t. Let $X_n(t) \in \{0, 1\}$ indicate that delivery portion n covers the user's viewport, which we assume to be i.i.d. over time [1] An example is illustrated in Fig. 1, where two portions both cover the predicted viewport. However, only the larger portion fully covers the user's actual viewport, i.e. $X_1(t) = 0$ and $X_2(t) = 1$.

B. Quality Selection and Wireless Transmission

In addition to selecting a delivery portion, the edge server also selects one of M possible quality levels, which correspond to different levels of video compression. For example, we could use the *constant rate factor* (CRF) in H.264 encoding, where a higher quality level m corresponds to a lower CRF value and vice versa. Let $q_m > 0$ denote the utility to the user for receiving a video frame at quality level m, with $q_1 < q_2 < \cdots < q_M$. For example, in Fig. 1, the left image depicts portions with lower quality (quality level 1) due to higher compression, whereas the right image shows portions with higher quality (quality level 2) resulting from lower compression. These two quality levels have utilities q_1 and q_2 respectively, with $q_1 < q_2$. We use $\widehat{m}(t)$ to denote the quality level selected by the edge server for frame t.

Note that the content of a fixed delivery portion n varies over time because the content in each tile varies according to the changing scene and the user's changing position. Also, the particular tiles that constitute the delivery portion vary over time due to the user's changing head orientation. Furthermore, the size (in bytes) of the delivery portion encoded at a fixed quality level m varies over time because the compression efficiency depends on the content. For example, quickly changing content or highly detailed content will result in a lower compression ratio compared to static or low detail content. We assume that the size (in bytes) of a fixed delivery portion n encoded at a fixed quality level m is i.i.d. over time. We also assume that the wireless channel state is i.i.d. over time. If delivery portion n and quality level m are selected in timeslot t, then a successful transmission occurs when all packets constituting the chosen portion compressed at the chosen quality level are received by the user, which we indicate by $Y_{mn}(t) \in \{0, 1\}$, and which is therefore i.i.d. over time.

[1] A user's motion is typically not i.i.d. in practice, but trace-based simulations still demonstrate the efficacy of our proposed algorithm.

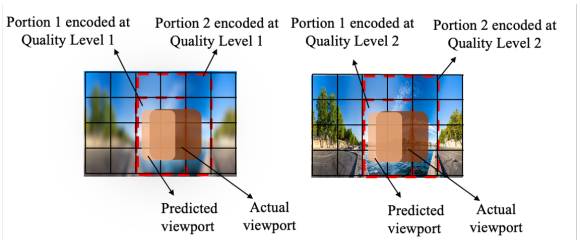


Fig. 1: A panoramic scene encoded at two different quality levels with two different portions (red dashed lines).

C. High-Quality and Consistent Viewport Delivery

For the user's viewport to be successfully delivered when delivery portion n and quality level m are chosen, delivery portion n needs to cover the viewport, indicated by $X_n(t)$, and all packets constituting the encoded delivery portion need to be delivered to the user, indicated by $Y_{mn}(t)$. Then $Z_{mn}(t) \triangleq X_n(t)Y_{mn}(t)$ indicates the successful viewport delivery. Since both $X_n(t)$ and $Y_{mn}(t)$ are i.i.d. over time, so is $Z_{mn}(t)$. We denote its mean by $\mu_{mn} \triangleq \mathbf{E}[Z_{mn}(t)]$, and its variance by $\sigma^2_{mn} \triangleq \mathrm{Var}[Z_{mn}(t)]$, which are unknown a priori due to the unknown behavior of the user and the wireless channel.

We would like to maximize the *cumulative quality* of the user's viewport up to a time horizon T, which is given by

$$R(T) \triangleq \sum_{t=1}^{T} \mathbf{E}\big[q_{\widehat{m}(t)}\, Z_{\widehat{m}(t)\,\widehat{n}(t)}(t)\big]. \tag{1}$$

In addition to maximizing the quality, we would like to maintain visual consistency to avoid breaking the immersion of the user. Therefore, in order to deliver a consistent experience to the user, we would like to minimize the *cumulative quality variance*, which is given by

$$V(T) \triangleq \sum_{t=1}^{T} \mathbf{E}\left[\left(q_{\widehat{m}(t)}\, Z_{\widehat{m}(t)\,\widehat{n}(t)}(t) - \frac{1}{T}\sum_{\tau=1}^{T} q_{\widehat{m}(\tau)}\, Z_{\widehat{m}(\tau)\,\widehat{n}(\tau)}(\tau)\right)^2\right]. \tag{2}$$

While the cumulative quality variance quantifies the long-term visual inconsistency, we would also like to maintain short-term visual consistency by minimizing the number of times the quality of *consecutively delivered* frames is switched. Then the *cumulative quality switching cost* is given by

$$C(T) \triangleq \sum_{t=2}^{T} \mathbf{E}\big[\mathbb{1}\{\widehat{m}(t) \neq \widehat{m}(t-1)\} Z_{\widehat{m}(t)\widehat{n}(t)}(t) Z_{\widehat{m}(t-1)\widehat{n}(t-1)}(t-1)\big]. \tag{3}$$

In summary, maximizing $R(T)$ helps deliver a *high-quality* experience to the user, minimizing $V(T)$ helps deliver a *long-term consistent* experience to the user, and minimizing $C(T)$ helps deliver a *short-term consistent* experience to the user. Therefore, we would like to maximize the *quality of experience* given by

$$\mathrm{QoE}(T) \triangleq \alpha_R\, R(T) - \alpha_V\, V(T) - \alpha_C\, C(T). \tag{4}$$

where $\alpha_R, \alpha_V, \alpha_C > 0$ tune the three-way tradeoff. Increasing one of the three tuning parameters increases the impact of its corresponding metric on the overall QoE.

In the language of multi-armed bandits, we call each pair $(m, n) \in [M] \times [N]$ of delivery portion and quality an *arm*. Denote the optimal delivery portion and quality selection policy by $*$, and its quality of experience by

$$\mathrm{QoE}^*(T) \triangleq \alpha_R\, R^*(T) - \alpha_V\, V^*(T) - \alpha_C\, C^*(T). \tag{5}$$

Note that differently from the classic MAB setup, the optimal policy is not generally given by playing a single fixed arm. Then, the regret of a policy is defined as

$$\mathrm{Reg}(T) \triangleq \mathrm{QoE}^*(T) - \mathrm{QoE}(T). \tag{6}$$

The ratio $\rho \triangleq \alpha_R/\alpha_V$ of the tuning parameters α_R and α_V can be thought of as the *risk-tolerance* and $\rho R(T) - V(T)$ is the *mean-minus-variance* in the risk-averse learning literature [11]. Consider the policy $\dagger$ with optimal mean-minus-variance $\rho R^\dagger(T) - V^\dagger(T)$. Also, note that $-\alpha_C\, C^*(T) \leq 0$. Then we can decompose

$$\begin{aligned}
&\mathrm{Reg}(T) \\
&\leq \alpha_V\, [\rho R^*(T) - V^*(T) - (\rho R(T) - V(T))] + \alpha_C\, C(T) \\
&\leq \alpha_V\, \underbrace{\big[\rho R^\dagger(T) - V^\dagger(T) - (\rho R(T) - V(T))\big]}_{\text{mean-minus-variance regret}} + \alpha_C\, C(T).
\end{aligned} \tag{7}$$

This form of the regret further motivates our algorithm design, which we present in the next section.

III. Algorithm Design

In this section, we introduce our algorithm called ***L**earning **O**ptimistically with **B**lock-based **O**pportunistic **S**witching (LOBOS)*. "Learning optimistically" refers to the UCB estimates that use the principle of *optimism under uncertainty* and which we discuss in Section III-A. We then discuss the "block-based opportunistic switching" in Section III-B.

A. UCB Estimates for Mean-Minus-Variance

We indicate that the edge server plays arm (m, n) in timeslot t by $I_{mn}(t) \triangleq \mathbb{1}\{\widehat{m}(t) = m, \widehat{n}(t) = n\}$. Let $H_{mn}(t)$ denote the number of times arm (m, n) has been played by the beginning of timeslot t. Then $H_{mn}(1) = 0$ and for $t > 1$, we update $H_{mn}(t+1) = H_{mn}(t) + I_{mn}(t)$.

Let $\overline{\mu}_{mn}(t)$ denote the estimate of μ_{mn} at the beginning of timeslot t. Then $\overline{\mu}_{mn}(1) = 0$ and for $t > 1$, we update

$$\overline{\mu}_{mn}(t+1) = \frac{H_{mn}(t)\,\overline{\mu}_{mn}(t) \;+\; I_{mn}(t)\, Z_{mn}(t)}{\max\{1, H_{mn}(t+1)\}}. \tag{8}$$

Let $\overline{\sigma}^2_{mn}(t)$ denote the estimate of σ^2_{mn} at the beginning of timeslot t. Then $\overline{\sigma}^2_{mn}(1) = 0$ and for $t > 1$, we update

$$\overline{\sigma}^2_{mn}(t+1) = \frac{H_{mn}(t)\,\overline{\sigma}^2_{mn}(t) \;+\; (I_{mn}(t)\,Z_{mn}(t) \;-\; \overline{\mu}_{mn}(t+1))^2}{\max\{1, H_{mn}(t+1)\}}. \tag{9}$$

Define the empirical mean-minus-variance for arm (m,n) in timeslot t as

$$\overline{w}_{mn}(t) \triangleq \rho\, q_m\, \overline{\mu}_{mn}(t) - q_m^2\, \overline{\sigma}^2_{mn}(t) \tag{10}$$

where we recall that $\rho \triangleq \alpha_R/\alpha_V$ tunes the tradeoff between mean and variance according to the weights in (4). Finally, for $H_{mn}(t) > 0$, the UCB estimate for arm (m,n) at the beginning of timeslot t is given by

$$U_{mn}(t) \triangleq \overline{w}_{mn}(t) + \varphi_m \sqrt{\frac{\log t}{H_{mn}(t)}} \tag{11}$$

where $\varphi_m \triangleq \sqrt{\frac{3}{2}\max\{q_m^2, q_m^4\}}(2+\rho)$ is a scaling factor on the confidence radius derived using Hoeffding's lemma and according to the mean-variance concentration bounds derived by Vakili and Zhao [11] (their Lemma 1). [2]

B. Block-Based Arm Selection with Opportunistic Switching

We use block-based arm selection to control the short-term consistency. We group time slots into windows indexed by $f = 0, 1, 2, \ldots$ and divide each window f into blocks indexed by $b = 1, 2, \ldots, b_f$, where b_f denotes the last block in window f. We use $t_s(f,b)$ and $t_e(f,b)$ to denote the starting and ending timeslots for block b of window f.

Window $f = 0$ is the *initial exploration window* and contains MN blocks, each containing a single timeslot. Here, we use each block to explore each arm once, i.e., we choose arm (m,n) in block $b = (m-1)N + n$ of window $f = 0$.

Let $\mathcal{T}_s$ denote the set of *block-starting timeslots*, i.e., a timeslot $t \in \mathcal{T}_s$ if there exists a window f and a block b such that $t_s(f,b) = t$. Then, the edge server is allowed to switch arms in timeslots $\mathcal{T}_s$. However, note that we only consider switching to be costly in consecutively delivered frames. Then if $Z_{\widehat{m}(t-1)\widehat{n}(t-1)}(t-1) = 0$, we know no switching cost will be suffered and therefore we can *opportunistically switch* in timeslot t even if it's in the middle of a block. Then in each timeslot $t \geq t_s(1,1)$, the edge server selects the arm

$$(\widehat{m}(t), \widehat{n}(t)) \in \begin{cases} \underset{(m,n)\in[M]\times[N]}{\operatorname{argmax}} U_{mn}(t) & \text{if } t \in \mathcal{T}_s \text{ or} \\ & Z_{\widehat{m}(t-1)\widehat{n}(t-1)}(t-1) = 0\text{'} \\ (\widehat{m}(t-1)\,,\, \widehat{n}(t-1)) & \text{otherwise.} \end{cases} \tag{12}$$

Note that the definition of quality switching cost (3) implies that we should allow portion switching in every timeslot. However, we noticed in experiments that this does not give a noticeable regret improvement under opportunistic switching.

[2]Updating the empirical estimates of each arm takes $O(MN)$ time and dominates the per-timeslot running time of POLAR.

The block and window structure for windows $f > 0$ is the following: each block b in window f contains f timeslots and each window f contains $\left\lceil \frac{f^\beta - (f-1)^\beta}{\alpha_C\, f} \right\rceil$ blocks for some tuning parameter $\beta \geq 2$. α_C appears in the denominator so that when the switching cost has a smaller weight, the windows are longer and therefore the block lengths increase slower, allowing for more switching overall. Increasing β has a similar effect. Fig. 2 shows the progression of windows and blocks.

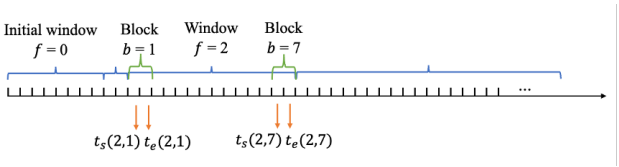


Fig. 2: A demonstration of windows and block for $M = 2$, $N = 4$, $\beta = 3$, and $\alpha_C = 0.5$.

IV. Quality of Experience Regret Guarantee

In this section, we derive a theoretical regret upper bound for the LOBOS algorithm (Theorem 1) and give its interpretation in Remark 1. Before doing so, we derive a preliminary upper bound that decomposes the regret into the sum of suboptimal arm plays and switching costs (see (15)).

In general, the optimal policy † in the mean-minus-variance problem (without switching) is not given by playing a fixed arm. However, Vakili and Zhao [11] bound the difference in mean-minus-variance between the optimal policy † and the optimal fixed-arm policy. The stationary mean-minus-variance for each arm (m,n) is given by $w_{mn} \triangleq \rho q_m \mu_{mn} - q_m^2 \sigma^2_{mn}$. Let $(\widetilde{m}, \widetilde{n}) \in \operatorname{argmax}_{(m,n)} w_{mn}$ denote the optimal fixed arm and assume it is unique [3]. Then the optimal fixed-arm policy is given by always playing $(\widetilde{m}, \widetilde{n})$. Let $\widetilde{R}(T)$ and $\widetilde{V}(T)$ denote the average cumulative quality and average cumulative quality variance under the optimal fixed arm policy. By Theorem 1 of Vakili and Zhao [11], we have

$$\rho R^\dagger(T) - V^\dagger(T) - \left(\rho \widetilde{R}(T) - \widetilde{V}(T)\right) \leq \min\left\{ \sigma^2_{\max} \sum_{\substack{(m,n) \\ \neq(\widetilde{m},\widetilde{n})}} \left(\frac{\Gamma^2_{mn}}{\Delta_{mn}} + 1 \right), \frac{MN}{a} \log T \right\} \triangleq \mathtt{gap}(T) \tag{13}$$

where $\sigma^2_{\max} \triangleq \max_{(m,n)} \sigma^2_{mn}$, $\Gamma_{mn} \triangleq q_{\widetilde{m}}\mu_{\widetilde{m}\widetilde{n}} - q_m \mu_{mn}$, $\Delta_{mn} \triangleq w_{\widetilde{m}\widetilde{n}} - w_{mn}$, and $a \triangleq \min_m \frac{2}{\max\{q_m^4, q_m^2\}}$ is derived using Hoeffding's lemma to satisfy the conditions of Lemma 1 from Vakili and Zhao [11]. By their Lemma 2, we have that under any policy,

$$\rho \widetilde{R}(T) - \widetilde{V}(T) - (\rho R(T) - V(T)) \leq \sum_{m=1}^{M} \sum_{n=1}^{N} (\Delta_{mn} + \Gamma^2_{mn})\, \mathbf{E}[H_{mn}(T+1)] + \sigma^2_{\widetilde{m}\widetilde{n}}. \tag{14}$$

[3]This assumption is necessary in analysis following Vakili and Zhao [11]. However, the assumption is mild since it is highly unlikely that two arms share the exact same real-valued mean-minus-variance.

Then combining (7), (13), and (14) gives that under any policy,

$$\mathrm{Reg}(T) \le \alpha_V \left(\mathtt{gap}(T) + \sigma^2_{\widetilde{m}\widetilde{n}}\right) + \alpha_C\, C(T) + \alpha_V \sum_{m=1}^{M}\sum_{n=1}^{N} [\Delta_{mn} + \Gamma^2_{mn}]\, \mathbf{E}[H_{mn}(T+1)]. \tag{15}$$

Then to derive a regret bound for our algorithm, it suffices to bound the expected number of plays $\mathbf{E}[H_{mn}(T+1)]$ of each "suboptimal" (with respect to the best fixed-arm policy) arm $(m,n) \neq (\widetilde{m},\widetilde{n})$ and the switching cost $C(T)$.

We begin by bounding the expected number of suboptimal arm plays in the following lemma.

Lemma 1. *Under the LOBOS algorithm, the expected number of plays of any arm $(m,n) \neq (\widetilde{m},\widetilde{n})$ is bounded by*

$$\mathbf{E}[H_{mn}(T+1)] \le \frac{8\varphi^2_m \left((\alpha_C\, T)^{1/\beta}+1\right)\log T}{\min\{\Delta^2_{mn}\,,\, 4(2+\rho)^2\}} + 5 + 4(\beta+2)\mathfrak{D}(\beta) \tag{16}$$

where $\mathfrak{D}(\beta) \triangleq \sum_{k=1}^{\infty}(1+\frac{1}{k})^{\beta}k^{-\beta} < \infty$ is a Dirichlet series.

Proof. The proof uses techniques from Vakili and Zhao's regret bound for their mean-variance UCB algorithm [11], but with many modifications to accommodate the block-based opportunistic switching, including partitioning the blocks into "sub-blocks" where opportunistic switches occur, and bounding the blocks and windows by a series. It can be found in the appendix of the full technical report [31]. □

Next, we bound the quality switching cost.

Lemma 2. *Under the LOBOS algorithm, the quality switching cost is bounded by*

$$\alpha_C\, C(T) \le \left((MN+2) + \tfrac{\beta}{\beta-1}2^{\beta-1}\right)(\alpha_C\, T)^{1-1/\beta}. \tag{17}$$

Proof. The proof uses a similar technique to Steiger et al. [29] to bound the quality switching cost, which takes a broad approach and essentially counts the total number of blocks up to the time horizon. It can be found in the appendix of the full technical report [31]. □

Our final regret bound is an *instance-dependent* bound, meaning it depends on $(\mu_{mn}, \sigma^2_{mn})_{(m,n)}$. It is given as follows.

Theorem 1. *Under the LOBOS algorithm, the QoE regret is bounded by*

$$\begin{aligned}\mathrm{Reg}(T) &\le \alpha_V \left(\mathtt{gap}(T) + \sigma^2_{\widetilde{m}\widetilde{n}} + (5+4(\beta+2)\,\mathfrak{D}(\beta))\,\|\boldsymbol{\zeta}\|_{1,1}\right) \\ &\quad + \xi\left((\alpha_C\, T)^{1/\beta}+1\right)\log T \sum_{\substack{(m,n)\\ \neq(\widetilde{m},\widetilde{n})}} \frac{\max\{q^2_m, q^4_m\}\zeta_{mn}}{\min\{\Delta^2_{mn}\,,\,4(2+\rho)^2\}} \\ &\quad + \left((MN+2) + \frac{\beta\, 2^{\beta-1}}{\beta-1}\right)(\alpha_C\, T)^{1-1/\beta}\end{aligned} \tag{18}$$

where we define the constants $\zeta_{mn} \triangleq \Delta_{mn} + \Gamma^2_{mn}$ and $\xi \triangleq 12\left(4(\alpha_R+\alpha_V) + \alpha^2_R/\alpha_V\right)$ for convenience.

Proof. Combine (15), Lemma 1, and Lemma 2, and note that each $8\alpha_V\varphi^2_m \le \max\{q^2_m, q^4_m\}\,\xi$. □

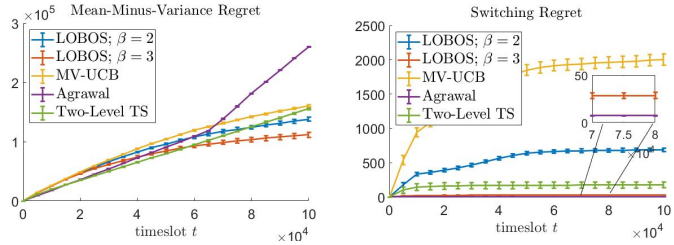
Remark 1 (interpretation of the regret bound)**.** For the bound in Theorem 1, the first line is $O(\log T)$. Notice that if $\alpha_C = O(1/T)$, then $\mathrm{Reg}(T) = O(\log T)$, which matches the instance-dependent bound for MV-UCB from Vakili and Zhao [11]. Otherwise, setting $\beta = 2$ gives $\mathrm{Reg}(T) = \widetilde{O}(T^{1/2})$, which matches the instance-dependent bound for MV-UCB in [32], but falls short of the best instance-dependent bounds for both bandits with switching costs [21] and bandits under mean-variance measure [11]. The second line depends heavily on the problem instance. However, if we restrict to problem instances having $\Delta_{mn} \ge T_0^{-1/(2\beta)}$ for all $(m,n) \neq (\widetilde{m},\widetilde{n})$ for some fixed time T_0, then the second line will be on the order $\widetilde{O}(T^{2/\beta})$. Setting $\beta = 3$ gives $\mathrm{Reg}(T) = \widetilde{O}(T^{2/3})$ overall, which matches (up to log factors) the $\Omega(T^{2/3})$ instance-independent lower bound shared by both the mean-variance problem (without switching) [11] and the switching cost problem (without variance and with stochastically-constrained rewards) [27]. Note that compared to regular stochastic bandits, where an instance is characterized only by its reward gaps Δ_{mn}, we also have Γ_{nm} in our bound due to the variance. Therefore, we cannot simply apply the usual technique of separately considering small and large reward gaps (see the proof of Theorem 7.2 in [33]) to derive a general instance-independent bound from our instance-dependent bound.

V. Panoramic Video Streaming Experiments

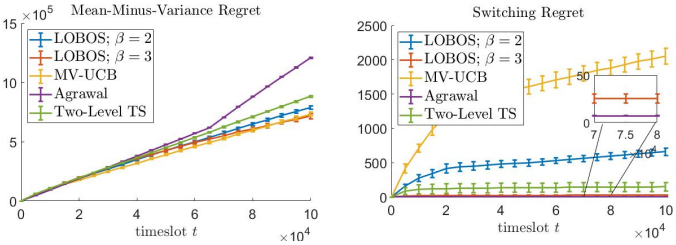
In this section, we validate our theoretical regret bound in a panoramic video streaming application according to a data trace recorded from a user interacting with a panoramic video stream in the real world. We compare our algorithm LOBOS with tuning parameters $\beta = 2$ and $\beta = 3$ (as per Remark 1) against three algorithms from the literature. The first is the original block-based algorithm from Agrawal et al. [21] that achieves $O(\log T)$ instance-dependent regret in the bandits with switching costs problem, and which we call "Agrawal". The second is the MV-UCB algorithm from Vakili and Zhao [11] that achieves $O(\log T)$ instance-dependent regret in the risk-averse bandits problem. The third is the Two-Level Thompson Sampling (TS) algorithm from Chen et al. [13] that leverages the two feedback signals $X_{\widehat{n}(t)}(t)$ and $Y_{\widehat{m}(t)\,\widehat{n}(t)}(t)$ independently for improved bandit learning. All experiments are repeated 10 times to reduce the variance due to the wireless channel, which is synthetically simulated.

A. Trace-Based Simulation Setup

We collect a data trace from a user experiencing a free educational panoramic video (see [34]). Specifically, we record a 3 DoF (orientation only) motion trace, and encode the panoramic content into different quality levels. The dataset includes 3000 timeslots and we iterate over the dataset multiple times to achieve a total of $T = 100000$ timeslots in our experiments with a uniformly random offset over the first half of the dataset each time it is looped. The details for each are as follows:



(a) $(\alpha_R, \alpha_V, \alpha_C) = (1, 1, 1)$ (b) $(\alpha_R, \alpha_V, \alpha_C) = (1, 1, 1)$



(c) $(\alpha_R, \alpha_V, \alpha_C) = (1, 5, 1)$ (d) $(\alpha_R, \alpha_V, \alpha_C) = (1, 5, 1)$

Fig. 3: Regret in the trace-based simulation setups.

1) Quality levels: We project the panoramic scene into a rectangular texture using equirectangular projection and split each texture into 4×6 tiles. We render the equirectangular map of the panoramic content at a 3840×2160 resolution and encode all tiles at $M = 4$ different quality levels with CRF values $(31, 27, 23, 19)$ [4] using FFmpeg [35]. The corresponding utility values are $\boldsymbol{q} = (1, 2, 3, 4)$. We then construct a dataset by recording the number of bytes composing each tile encoded at each quality level for each frame.

2) Delivery portions: We process the motion trace at each time slot t by utilizing a linear regression model to forecast the head pose in the next timeslot. This involves training the model with motion trajectories from the past three time slots. Then we can calculate $X_{\widehat{n}(t)}(t)$ after selecting the delivery portion $\widehat{n}(t)$ by checking if the portion covers the user's actual viewport from timeslot t according to the motion trace. We set $N = 5$ different portions to select from: $100° \times 90°$ (minimum viewport), $102° \times 91°$, $108° \times 94°$, $120° \times 100°$, and $360° \times 180°$ (entire $360°$ scene), where each pair gives the angles corresponding to the yaw and pitch axes respectively.

3) Wireless Transmissions: We define $\texttt{pkt}_{\texttt{net}}(t)$ as the number of packets that can be successfully delivered over the wireless channel from the edge server to the user in timeslot t and assume this follows a Poisson distribution: $\texttt{pkt}_{\texttt{net}}(t) \sim \text{Poisson}(\lambda_{\texttt{net}})$ with $\lambda_{\texttt{net}} = 80$. Let $\texttt{pkt}_{mn}(t)$ denote the number of packets required to transmit delivery portion n encoded at quality level m in timeslot t. In each timeslot t, after selecting the portion $\widehat{n}(t)$ and quality level $\widehat{m}(t)$, we extract the corresponding number of bytes from the dataset and split it up into $\texttt{pkt}_{\widehat{m}(t)\,\widehat{n}(t)}(t)$ UDP packets, where each packet's payload size is 1400 bytes. We observe a successful transmission if the required number of packets does not exceed the channel rate, i.e. $Y_{\widehat{m}(t)\,\widehat{n}(t)}(t) = \mathbb{1}\{\texttt{pkt}_{\texttt{net}}(t) \geq \texttt{pkt}_{\widehat{m}(t)\,\widehat{n}(t)}(t)\}$.

[4]Note that for a higher CRF value, the panoramic scene will be encoded at a lower bitrate, resulting in a lower visual quality.

B. Discussion of Results

In Fig. 3, we show how LOBOS compares to the other algorithms under two different QoE setups. We separately plot the *mean-minus-variance regret* $\alpha_V \left[\rho\widetilde{R}(T) - \widetilde{V}(T) - (\rho R(T) - V(T))\right]$ and the *switching regret* $\alpha_C C(T)$ with respect to the best fixed-arm policy to more clearly visualize the tradeoffs. The best fixed arm $(\widetilde{m}, \widetilde{n})$ is calculated according to the statistics of the trace.

1) First Setup: We set the parameters $(\alpha_R, \alpha_V, \alpha_C) = (1, 1, 1)$, ensuring equal emphasis on total perceived quality, long-term consistency, and short-term consistency. LOBOS outperforms other algorithms in both mean-minus-variance regret and switching regret when $\beta = 3$. As shown in Fig. 3a, LOBOS achieves lower mean-minus-variance regret than MV-UCB, likely because its block-based approach guarantees zero quality variance over intervals within each block in which all frames are delivered, i.e. it reduces the variance incurred by exploration. We also notice that Two-Level Thompson Sampling and Agrawal exhibit linear mean-minus-variance regret, as they only account for mean, and therefore may settle on an arm with higher variance. Notably, Agrawal's regret curve steepens after $t = 60000$ due to its block-based design, which locks the algorithm into selecting a significantly worse arm than before. Additionally, Fig. 3b shows that MV-UCB incurs significantly higher switching costs compared to other algorithms, as it lacks a mechanism to control switching. Interestingly, while Two-Level Thompson Sampling also lacks an explicit switching-cost control, its switching cost remains low because even if it fails to identify the true optimal arm, it quickly commits to a single suboptimal arm.

2) Second Setup: We adjust the parameters to $(\alpha_R, \alpha_V, \alpha_C) = (1, 5, 1)$, increasing the weight of long-term consistency α_V. LOBOS consistently outperforms other algorithms in both mean-minus-variance regret and switching regret when $\beta = 3$. As seen in Fig. 3c, the performance gap between MV-UCB and LOBOS narrows. This is expected, as greater emphasis on long-term consistency reduces the importance of short-term consistency. Furthermore, both MV-UCB and LOBOS outperform Two-Level Thompson Sampling more noticeably in terms of mean-minus-variance regret under this setup. Since Two-Level Thompson Sampling primarily optimizes cumulative perceived quality without considering long-term consistency, its shortcomings become more evident when α_V increases. The switching cost patterns in Fig. 3d remain similar to those in Fig. 3b.

It's interesting to note that throughout the simulations in Fig. 3, LOBOS with $\beta = 3$ outperforms LOBOS with $\beta = 2$. This leads us to believe that for the regret bound (18) in Theorem 1, the $O(T^{1/\beta})$ term scaled by the inverse mean-minus-variance gaps in line 2 dominates the $O(T^{1-1/\beta})$ term in line 3. It also shows that the bound in Lemma 2 is too loose to account for the reduction in overall switching due to increased early exploration allowed by a larger β. However,

there appears to be a sweet spot for a given problem instance, since LOBOS becomes the MV-UCB algorithm as $\beta \to \infty$.

VI. Concluding Remarks and Future Work

In this paper, we studied the problem of delivering a high-quality and consistent experience to a user experiencing interactive panoramic content wirelessly by formulating it as a risk-averse multi-armed bandit problem with reward-dependent switching costs. To solve this problem, we developed a block-based UCB algorithm with opportunistic switching and proved that it achieves a competitive regret upper bound compared to the state-of-the-art. We then validated our theoretical findings in trace-based simulations using real panoramic video and head movement data traces. In future work, we would like to extend our results by leveraging the separate feedback signals of motion prediction and wireless transmission to improve the learning, as done in recent work [13], [14].

Acknowledgment

The authors would like to thank Jiangong Chen from the Smart Networking and Computing (SNeC) lab at Penn State for building the system to collect the data trace [14].

References

[1] Y. Bao, H. Wu, T. Zhang, A. A. Ramli, and X. Liu, "Shooting a moving target: Motion-prediction-based transmission for 360-degree videos," in *2016 IEEE International Conference on Big Data (Big Data)*, 2016, pp. 1161–1170.

[2] C.-F. Hsu, A. Chen, C.-H. Hsu, C.-Y. Huang, C.-L. Lei, and K.-T. Chen, "Is foveated rendering perceivable in virtual reality? exploring the efficiency and consistency of quality assessment methods," in *Proceedings of the 25th ACM international conference on multimedia*, 2017, pp. 55–63.

[3] S. Chen, B. Duinkharjav, X. Sun, L.-Y. Wei, S. Petrangeli, J. Echevarria, C. Silva, and Q. Sun, "Instant reality: Gaze-contingent perceptual optimization for 3d virtual reality streaming," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, pp. 2157–2167, 2022.

[4] V. Joseph and G. de Veciana, "Jointly optimizing multi-user rate adaptation for video transport over wireless systems: Mean-fairness-variability tradeoffs," in *2012 Proceedings IEEE INFOCOM*. IEEE, 2012, pp. 567–575.

[5] J. Chen, F. Qian, and B. Li, "Enhancing quality of experience for collaborative virtual reality with commodity mobile devices," in *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2022, pp. 1018–1028.

[6] V. Joseph and G. de Veciana, "Nova: Qoe-driven optimization of dash-based video delivery in networks," in *IEEE INFOCOM 2014-IEEE conference on computer communications*. IEEE, 2014, pp. 82–90.

[7] Y. Zhang, P. Zhao, K. Bian, Y. Liu, L. Song, and X. Li, "Drl360: 360-degree video streaming with deep reinforcement learning," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 1252–1260.

[8] X. Liu, C. Vlachou, M. Yang, F. Qian, L. Zhou, C. Wang, L. Zhu, K.-H. Kim, G. Parmer, Q. Chen *et al.*, "Firefly: Untethered multi-user {VR} for commodity mobile devices," in *2020 USENIX Annual Technical Conference (USENIX ATC 20)*, 2020, pp. 943–957.

[9] F. Qian, B. Han, Q. Xiao, and V. Gopalakrishnan, "Flare: Practical viewport-adaptive 360-degree video streaming for mobile devices," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, pp. 99–114.

[10] A. Sani, A. Lazaric, and R. Munos, "Risk-aversion in multi-armed bandits," *Advances in neural information processing systems*, vol. 25, 2012.

[11] S. Vakili and Q. Zhao, "Risk-averse multi-armed bandit problems under mean-variance measure," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 6, pp. 1093–1111, 2016.

[12] M. S. Anwar, J. Wang, A. Ullah, W. Khan, S. Ahmad, and Z. Fei, "Measuring quality of experience for 360-degree videos in virtual reality," *Science China Information Sciences*, vol. 63, pp. 1–15, 2020.

[13] J. Chen, B. Li, and R. Srikant, "Thompson-sampling-based wireless transmission for panoramic video streaming," in *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*. IEEE, 2020, pp. 1–3.

[14] H. Gupta, J. Chen, B. Li, and R. Srikant, "Online learning-based rate selection for wireless interactive panoramic scene delivery," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 1799–1808.

[15] X. Wu and B. Li, "Achieving regular and fair learning in combinatorial multi-armed bandit," in *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications*, 2024, pp. 361–370.

[16] Q. Zhu and V. Tan, "Thompson sampling algorithms for mean-variance bandits," in *International Conference on Machine Learning*. PMLR, 2020, pp. 11 599–11 608.

[17] N. Galichet, M. Sebag, and O. Teytaud, "Exploration vs exploitation vs safety: Risk-aware multi-armed bandits," in *Asian conference on machine learning*. PMLR, 2013, pp. 245–260.

[18] J. Y. Yu and E. Nikolova, "Sample complexity of risk-averse bandit-arm selection." in *IJCAI*, 2013, pp. 2576–2582.

[19] H. M. Markowitz, "Foundations of portfolio theory," *The journal of finance*, vol. 46, no. 2, pp. 469–477, 1991.

[20] M. Wirth, A. Klein, and A. Ortiz, "Risk-aware multi-armed bandits for vehicular communications," in *2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring)*. IEEE, 2022, pp. 1–7.

[21] R. Agrawal, M. Hedge, and D. Teneketzis, "Asymptotically efficient adaptive allocation rules for the multiarmed bandit problem with switching cost," *IEEE Transactions on Automatic Control*, vol. 33, no. 10, pp. 899–906, 1988.

[22] Y. Zhou, C. Shen, X. Luo, and M. van der Schaar, "A non-stationary online learning approach to mobility management," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–6.

[23] H. Li, T. Li, F. Li, Y. Wu, and Y. Wang, "Cumulative participant selection with switch costs in large-scale mobile crowd sensing," in *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, 2018, pp. 1–9.

[24] A. Meetoo Appavoo, S. Gilbert, and K.-L. Tan, "Shrewd selection speeds surfing: Use smart exp3!" in *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, 2018, pp. 188–199.

[25] O. Dekel, J. Ding, T. Koren, and Y. Peres, "Bandits with switching costs: T2/3 regret," in *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, ser. STOC '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 459–467. [Online]. Available: https://doi.org/10.1145/2591796.2591868

[26] C. Rouyer, Y. Seldin, and N. Cesa-Bianchi, "An algorithm for stochastic and adversarial bandits with switching costs," in *Proceedings of the 38th International Conference on Machine Learning (ICML)*, ser. Proceedings of Machine Learning Research. PMLR, 2021, pp. 9127–9135.

[27] I. Amir, G. Azov, T. Koren, and R. Livni, "Better best of both worlds bounds for bandits with switching costs," 2022. [Online]. Available: https://arxiv.org/abs/2206.03098

[28] D. Cheng, X. Zhou, and B. Ji, "Understanding the role of feedback in online learning with switching costs," 2023. [Online]. Available: https://arxiv.org/abs/2306.09588

[29] J. Steiger, B. Li, B. Ji, and N. Lu, "Constrained bandit learning with switching costs for wireless networks," in *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications*, 2023, pp. 1–10.

[30] Y. Huang, Q. Liu, and J. Xu, "Adversarial combinatorial bandits with switching cost and arm selection constraints," in *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications*, 2024.

[31] "Learning to wirelessly deliver consistent and high-quality interactive panoramic scenes." [Online]. Available: https://gitlab.com/juarensteiger/interactive-panoramic-scenes

[32] A. Sani, A. Lazaric, and R. Munos, "Risk-aversion in multi-armed bandits," 2013. [Online]. Available: https://arxiv.org/abs/1301.1936

[33] T. Lattimore and C. Szepesvari, "Bandit algorithms," 2017. [Online]. Available: https://tor-lattimore.com/downloads/book/book.pdf

[34] N. Geographic, "First-ever 3d vr filmed in space." [Online]. Available: https://vuze.camera/vr-gallery/3d-360-video/first-ever-3d-vr-filmed-space

[35] "Ffmpeg." [Online]. Available: http://ffmpeg.org/