Socially-Aware Tile-Based Point Cloud Multicast with Registration

Han-Rong Lai[†], Ru-Jun Wang[†♯], Chih-Hang Wang[♯], De-Nian Yang[♯], Wen-Tsuen Chen[†], and Jang-Ping Sheu[†]
[†]Dept. of Computer Science, National Tsing Hua University, Taiwan
[♯]Institute of Information Science, Academia Sinica, Taiwan
E-mail: s109062706@m109.nthu.edu.tw, allenw84@gapp.nthu.edu.tw,
{superwch7805, dnyang}@iis.sinica.edu.tw, and {wtchen, sheujp}@cs.nthu.edu.tw

Abstract—With the emergence of new applications for holographic-type communication in healthcare, entertainment, and education, point cloud video transmission has become essential. This paper aims to reduce the bandwidth cost by leveraging tile-based video transmission with point cloud registration in a wireless multicast network. A point cloud video is divided into multiple tiles, and each tile contains a portion of point cloud objects and can be registered by adjacent tiles with some similar objects under the registration rotation and registration overlap constraints. We formulate a new optimization problem and prove that it is NP-hard, and then we design an algorithm Multicast Multi-Tile Registration (MMTR) to select multicasting and registered tiles under consideration of socially related users' preferences with the idea of a *tile registration graph*. A more popular tile can be multicasted to more friends to minimize the bandwidth cost. Experimental results with real datasets show that MMTR can reduce bandwidth costs by more than 20% and achieve better video quality compared to state-of-the-art point cloud transmission algorithms.

I. Introduction

Holographic-type communication (HTC) enables a six-degree-of-freedom immersive experience for 3D hologram transmission [1], [2]. A common representation for virtual 3D holograms in HTC is the point cloud object, which uses 3D points with attributes like coordinates, color, and material to capture the appearance of real-world objects [2]. For example, Microsoft's HoloLens headset uses point cloud data to simulate patients' organs for healthcare [3]. A point cloud video can be divided into several *tiles* for reducing the bandwidth cost and enabling adaptive tile selection based on the user's field of view (FoV) [4]. A tile may contain several point cloud objects, and a portion of a point cloud object in a tile is defined as a *scan* [5], [6]. For example, in Fig. 1(a), a point cloud video with a TV (in blue) and a chair (in red) is divided into $5 \times 2 \times 1 = 10$ tiles, denoted by z_i. Scans s_i (in blue) and s_j (in red) are the portions of the TV and chair objects, respectively.

To optimize bandwidth efficiency for transmitting 3D video, conventional multi-view synthesis methods (such as Depth Image-Based Rendering (DIBR) [7], [8] and multi-view stereo [9]) synthesize a view from its neighboring left and right views. In contrast, *point cloud registration* can register[1] (synthesize) any point cloud object under the *registration rotation*
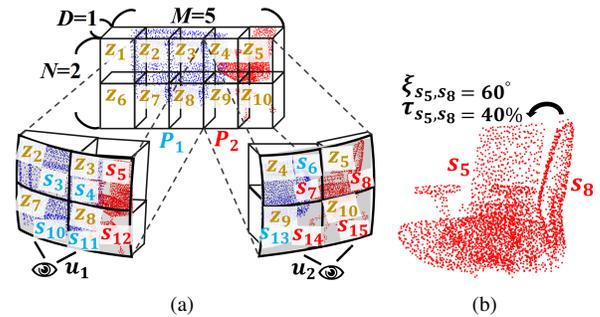

Fig. 1: A tile-based point cloud video example. (a) Objects $\{P_1, P_2\}$, tiles $\{z_1, \ldots, z_{10}\}$, scans $\{s_1, \ldots, s_{15}\}$, and users $\{u_1, u_2\}$. The requested tiles of u_1 and u_2 based on their FoVs are $\{z_2, z_3, z_7, z_8\}$ and $\{z_4, z_5, z_9, z_{10}\}$, respectively. (b) From u_1 and u_2's FoVs, s_5 (in tile z_3) and s_8 (in tile z_5) have overlap ratio $\tau_{s_5,s_8} = 40\%$ and rotation angle $\xi_{s_5,s_8} = 60°$.

and *registration overlap* constraints [10], [11] (detailed in Section II), which ensure the maximum rotation angle and the minimum *overlap ratio*[2] between the pair of scans used for registration [11], [12], respectively. However, multicast with point cloud registration for delivering tiles with shared objects[3] has not been explored.

Socially related users in a virtual space usually watch some identical objects [13]. For example, in holographic telepresence, only the attending friends can see the objects in the scene [14]. To reduce the number of transmitted tiles, we can multicast tiles with shared objects to socially related users, who can register their requested tiles from the multicast if the overlap ratio and rotation angle satisfy the registration constraints [11]. Zhang et al. [15] adopted K-mean++ to group users into multicast groups and used the Lyapunov optimization method to determine tile quality for multicast. Okamoto et al. [16] compressed tiles to reduce bandwidth costs by following the viewport transition of each user. However, they did not employ point cloud registration to prevent multicasting every tile under consideration of friends' preferences for tiles.

This paper aims to leverage point cloud registration to minimize the total bandwidth cost for multicasting the tile-

---

[1]*Register* is a term used for 3D point cloud registration, which means to synthesize the point cloud object.

[2]The overlap ratio between two scans quantifies the number of 3D points sharing identical spatial coordinates for the same object within both scans.

[3]Shared objects represent portions of the same point cloud object within the different tiles requested by the users with similar FoV.

based point cloud video to socially related users (e.g., friends who have similar preferences for watching tiles). However, new research challenges appear as follows. **C1) Minimizing the bandwidth cost with registration overlap.** We can choose tiles with more scans for registration to reduce the number of transmitted tiles to minimize the bandwidth cost because they have more opportunities to contain an identical object. However, each scan's size is limited by the number of scans in a tile. Hence, we may need to select more tiles to serve users if the scans in different tiles do not have enough overlap ratio for point cloud registration. **C2) Quality and rotation angle of tiles for registration.** To register a tile, we need to register every scan in the tile. The quality[4] and rotation angle of tiles affect *point cloud registration* [11], [12]. When higher-quality tiles (scans) are used for registration, the rotation angle between them can be larger to satisfy the *registration rotation constraint* easier. However, choosing lower-quality tiles (scans) can reduce the bandwidth cost, but it will require lower rotation angles of scans for successful registration. This limits the number of scans in the tile that can be used for registration and thus may need to select more tiles, which increases the bandwidth cost. **C3) FoV-aware tile selection and multicast grouping.** The data rate and quality requirement for each tile of each user vary due to the user's FoV. Multicasting a tile with higher quality can simultaneously meet the quality requirements of more users for the tile. However, since the Modulation and Coding Scheme (MCS) employed by the multicast group is determined by the lowest MCS among the group members, it may lead to a failure to meet the data rate requirements of users demanding higher MCS.

To address the above issues, we formulate a new optimization problem, named Tile Registration and Selection with Multicast (TRSM), to minimize the total bandwidth cost. Different from current video multicast that does not employ point cloud registration [15], [16], TRSM considers the registration relations between tiles with different qualities and users' requirements for registration and multicast to avoid delivering every tile. We prove TRSM is NP-hard and design a new algorithm, named Multicast Multi-Tile Registration (MMTR), to select multicasting and registered tiles, considering the preference of socially related users. MMTR first builds a tile registration graph (TRG) by constructing each quality of each tile as a node to represent if the pairs of scans between the nodes can meet the *registration overlap constraint*. MMTR introduces the object contribution indicator (OCI) for each node to examine how many pairs of scans meet the registration overlap constraint and how many users request it to address **C1**. MMTR adjusts the selected nodes for multicasting by examining the registration assistance indicator (RAI), which assesses the ratio of the number of scans that can be registered to the induced bandwidth cost to address **C2**. MMTR evaluates the quality redundancy indicator (QRI) to remove the selected nodes for reducing the bandwidth cost, and groups socially

related users into groups for multicast by considering their rate requirements to deal with **C3**. Experiments with a real dataset show that MMTR can reduce the bandwidth cost by more than 20% and achieve better video quality than baselines.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

Due to the space constraint, we provide the notation table and a comparative example in [17]. We consider tile-based point cloud video transmission and registration for multicast users. Let $\mathbb{E}$ be the set of qualities for point cloud video, where quality $e \in \mathbb{E}$ is defined by the ratio of the number of transmitted 3D points to the total 3D points [5], [6]. Following [18], a point cloud video with quality $e \in \mathbb{E}$ encodes a set of objects $\mathbb{P}$ with quality e. We divide the point cloud video into $M \times N \times D$ tiles for adaptive transmission [4], denoted as $\mathbb{Z} = \{z_1, \ldots, z_{M \times N \times D}\}$, indexed from the upper left to the lower right along the direction of depth. Each tile contains scans of different objects, which may span multiple tiles. Let $\mathbb{S} = \{S_1, \ldots, S_{|\mathbb{P}|}\}$ be the set of collections of scans for all objects, where S_P is the collection of scans of object $P \in \mathbb{P}$.

Let $\mathbb{U}$ be the set of users and α_z be the preference weight of tile z, which can be calculated according to the subscribers' social relations [19], because a user tends to prefer the tile (scene) that is popular among her friends, and a tile will be more popular if it is requested by users with more friends. A lower preference weight is assigned to a more popular tile because our goal is to minimize the bandwidth cost, and a popular tile can be multicasted to more users for registration to minimize the bandwidth cost. For each user $u \in \mathbb{U}$, let $\mathbb{Z}_u \subseteq \mathbb{Z}$ be the tiles requested by u and $Q_{z,u} \in \mathbb{E}$[5] be the minimum quality of each tile $z \in \mathbb{Z}_u$. We multicast tile z with quality e to a group of users with the most robust MCS among them [15]. Let $f(\cdot)$ map the user's channel state to the MCS [20] and γ_u be the channel state of user u. The MCS for user u receiving tile z with quality e in multicast group g is $\mathrm{MCS}_{z,e,u} = \min_{u,v \in g}(f(\gamma_u), f(\gamma_v))$. Users can receive requested tiles either via multicast or by registering them on the device using point cloud registration [21], [22]. Let binary variable $x_{s,s'}$ indicate whether scan s' is used to register scan s and $r_{s,u} \in \mathbb{E}$ be the quality of scan s received by user u. The quality of a scan from point cloud registration is the minimum quality of the scans that are used to register it [23]. That is, for a scan $s \in S_P$ that belongs to an object P, $r_{s,u} = \min_{s' \in S_P} x_{s,s'} r_{s',u}$. For point cloud video display, the quality $q_{z,u} \in \mathbb{E}$ of a registered tile z for user u is the lowest quality of the registered scans in the tile [24], i.e., $q_{z,u} = \min_{s \in K_z} r_{s,u}$, where K_z is the scans in tile z.

B. Problem Formulation

We formulate a new optimization problem, namely Tile Registration and Selection with Multicast (TRSM), to minimize

---

[4]Following [5], [6], we define the quality of a tile (scan) as the ratio of the number of transmitted 3D points in the tile (scan) to its total 3D points.

[5]Quality requirement $Q_{z,u}$ depends on two factors [5]: 1) The distance weight is the inverse of the distance between the user and the tile, which means the closer the tile is to the user, the higher the weight. 2) The proportion weight is the fraction of points in the tile out of the total points in the requested tiles, which means the more points the tile has, the higher the weight.

the total bandwidth cost in a wireless network with multicast and point cloud registration. Let $y_{z,e}$ be the binary variable indicating if tile z with quality e is multicasted and $b(e)$ be the bandwidth cost for multicasting a tile with quality e, where a tile with higher quality requires a larger bandwidth cost for multicasting more 3D points. The objective function is $\sum_{z\in\mathbb{Z}}\sum_{e\in\mathbb{E}}\alpha_z\times y_{z,e}\times b(e)$.[6] TRSM has the following constraints. 1) *Quality of Experience (QoE) constraint*. Each user has a different quality requirement for each tile according to his FoV, i.e., $q_{z,u}\geq Q_{z,u},\forall z\in\mathbb{Z}_u,\forall u\in\mathbb{U}$. 2) *Registration rotation constraint*. To successfully register a scan, the rotation angle $\xi_{s,s'}$ between a pair of scans s and s' for the same object P needs to be smaller than a threshold $\xi_{\text{threshold}}$ [11], i.e., $\xi_{s,s'}<\xi_{\text{threshold}},\forall s,s'\in S_P$. For example, with $\xi_{\text{threshold}}=30°$, scan s_{12} in tile z_8 cannot be registered by s_5 and s_8 in Fig. 1(a), since $\xi_{s_5,s_8}=60°>\xi_{\text{threshold}}$ as shown in Fig. 1(b). 3) *Registration overlap constraint*. To exploit point cloud registration, the overlap ratio $\tau_{s,s'}$ between scans s and s' for the same object P in different tiles needs to be larger than a threshold $\tau_{\text{threshold}}$ [11], i.e., $\tau_{s,s'}\geq\tau_{\text{threshold}},\forall s,s'\in S_P$. 4) *Rate requirement constraint*. Let $h(\cdot)$ be a function mapping MCS to the data rate [25]. The data rate $h(\text{MCS}_{z,e,u})$ of user u when transmitting tile z with quality e must be larger than threshold ψ_e to provide a complete viewing experience for u [1], i.e., $h(\text{MCS}_{z,e,u})>\psi_e,\forall z\in\mathbb{Z},\forall e\in\mathbb{E},\forall u\in\mathbb{U}$. 5) *Quality of tile selection constraint*. When user u receives tile z with different qualities, the higher-quality tile satisfies both the user's requirements and registration constraints, rendering the lower-quality tile redundant. Hence, each z is transmitted to u with one quality e. That is, $\sum_{e\in\mathbb{E}}\rho_{z,e,u}\leq 1,\forall z\in\mathbb{Z},\forall u\in\mathbb{U}$, where $\rho_{z,e,u}$ is a binary variable that indicates tile z with quality e is transmitted to user u.

Definition 1 (TRSM). Given a set of tiles $\mathbb{Z}$ with the qualities $e\in\mathbb{E}$, each tile z has a preference weight α_z, a set of users $\mathbb{U}$, each user u requests a set of tiles $\mathbb{Z}_u$, a set of point cloud objects $\mathbb{P}$, a set of collections of scans $\mathbb{S}$, TRSM aims to select a set of tiles with specific qualities and finds the lowest MCS among users for multicast such that *QoE, registration rotation, registration overlap, rate requirement*, and *quality of tile selection* constraints are satisfied. The objective is to minimize the total bandwidth cost $\sum_{z\in\mathbb{Z}}\sum_{e\in\mathbb{E}}\alpha_z\times y_{z,e}\times b(e)$.

Theorem 1. *TRSM is NP-hard.*

Proof. Due to the space constraint, the detailed proof of NP-hardness is provided in [17].

III. ALGORITHM

An intuitive approach for TRSM is to iteratively choose the tile with the minimum bandwidth cost and multicast it to a group of users requesting it. In each iteration, if the tile can be registered by the chosen tiles, the approach serves the tile by registration. Otherwise, the approach multicasts the tile to a group of users who request it, using the lowest MCS among

the users. However, choosing the tile with the least bandwidth cost may not effectively minimize the total bandwidth cost when fewer users request adjacent tiles. Moreover, the user with the lowest MCS may severely degrade the transmission rate, and some users' rate requirements may not be met.

To address these issues, we propose MMTR with three phases: 1) socially-aware tile selection (STS), 2) registration replacement (RR), and 3) tile selection adjustment (TSA). STS first builds a tile registration graph (TRG) by constructing each quality of each tile as a node to represent the registration relation between each pair of scans in the nodes. Then, we iteratively select the node with the maximum OCI for multicasting and register other nodes by the selected multicasting nodes.[7] OCI measures the contribution of each node for registration. Unlike previous research [15], [16] that multicast each tile individually without considering point cloud registration, RR examines RAI to replace multicasting nodes with registered nodes that can be utilized to register more scans. RAI is the ratio of the number of scans that can be registered by a registered node and the multicasting nodes to the bandwidth cost of the registered node. To minimize the bandwidth cost, TSA iteratively chooses the tile with the largest preference weight for removal, as it serves the fewest socially related users. Then, TSA removes the node (corresponding to the chosen tile) with the maximum QRI, which estimates the number of users it can satisfy. The time complexity of MMTR is $O(|\mathbb{Z}|^2|\mathbb{E}|^2|\mathbb{P}|^2+|\mathbb{Z}||\mathbb{E}||\mathbb{U}|)$. Due to the space constraint, the complexity analysis and pseudocode are presented in [17]. We also discuss how MMTR supports dynamic user requests and layered bitstream point cloud videos in [17].

1) Socially-aware Tile Selection (STS): To deal with **C1**, STS builds an undirected TRG by constructing each quality of each tile as a node to represent whether the pairs of scans between the nodes can meet the *registration overlap constraint*. Specifically, STS constructs the TRG $G^{\text{TRG}}=(V^{\text{TRG}},E^{\text{TRG}})$ with a set of nodes $V^{\text{TRG}}=\{n_{z,e}|z\in\mathbb{Z},e\in\mathbb{E}\}$ for each tile with each quality, where each node with a weight $\text{NW}_{z,e}$ indicating the number of users requesting $n_{z,e}$. A weighted edge $(n_{z,e},n_{z',e'})\in E^{\text{TRG}}$ exists if the scans in tiles z and z' share a common object and have an overlap ratio above $\tau_{\text{threshold}}$, where edge weight $\text{EW}_{z,e,z',e'}$ indicates the number of scan pairs that meet the *registration overlap constraint*. Then, STS iteratively selects the node with the maximum OCI in TRG to be multicasted. Specifically, let $N(n_{z,e})$ be the set of adjacent nodes of $n_{z,e}$ in G^{TRG}. We define OCI of node $n_{z,e}$ as $\text{OCI}_{z,e}=\frac{\sum_{n_{z',e'}\in N(n_{z,e})}\text{EW}_{z,e,z',e'}\times\text{NW}_{z,e}}{\alpha_z\times b(e)}$, and a larger $\text{OCI}_{z,e}$ means that tile z with quality e can be used to register more adjacent tiles and be multicasted to more users with a smaller bandwidth cost. Also, the tile with a smaller preference weight α_z is more popular among socially related users such that the tile can be multicasted to serve many

---

[6] Since multicast serves a group of users simultaneously, we count the bandwidth cost only once for each transmission of a tile with a quality.

[7] For ease of presentation, we define the "multicasting node" to represent the node on the TRG selected for multicasting, and the "registered node" is the node that can be registered (i.e., the *registration rotation* and *registration overlap* constraints are satisfied) by the multicasting nodes.
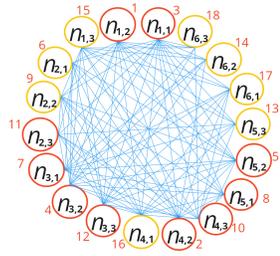
(a) The result of STS

User	u_1	u_2	u_3	u_4	u_5	u_6	u_7
$n_{z,e}$	$n_{1,2}, n_{2,1},$ $n_{4,2}$	$n_{1,1}, n_{4,3},$ $n_{5,2}$	$n_{4,2}, n_{5,3}$	$n_{3,1}, n_{5,1}$	$n_{3,2}, n_{5,2}$	$n_{2,2}, n_{3,3},$ $n_{6,2}$	$n_{1,2}, n_{2,3},$ $n_{3,2}$
MCS	2	2	3	2	2	2	3

(b) Configuration of users

|	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$	$n_{3,1}$	$n_{3,2}$	$n_{3,3}$	$n_{4,1}$	$n_{4,2}$	$n_{4,3}$	$n_{5,1}$	$n_{5,2}$	$n_{5,3}$	$n_{6,1}$	$n_{6,2}$	$n_{6,3}$
$n_{1,1}$	2	2	2	2	2	2	3	3	3	2	2	2	1	1	1
$n_{1,3}$	2	2	2	2	2	2	3	3	3	2	2	2	1	1	1

(c) Pairs of scans that can meet the registration overlap constraint.

Fig. 2: An illustrative example of STS.

users and be leveraged to register many adjacent tiles in those users' FoV. Let $\mathbb{T}$ and $\mathbb{C}$ be the sets of multicasting nodes and registered nodes, respectively. STS iteratively selects node $n_{z,e}$ in G^{TRG} with the largest $\text{OCI}_{z,e}$ for multicasting and adds it to $\mathbb{T}$. Then, for each node $n_{z,e} \notin \mathbb{T} \cup \mathbb{C}$, we iteratively add it to $\mathbb{C}$ if the node can be registered by the nodes in $\mathbb{T}$ and will not be selected to $\mathbb{T}$ until the end of this phase. STS stops when every node on G^{TRG} has been added to the sets $\mathbb{T}$ or $\mathbb{C}$, i.e., $\mathbb{T} \cup \mathbb{C} = V^{\text{TRG}}$.

Example 1. Due to the space constraint, the complete information of node weights, overlap relation between scans, and tile registration relationship are provided in [17]. We assume that $|\mathbb{U}| = 7, |\mathbb{Z}| = 6, |\mathbb{P}| = 4$, and $|\mathbb{S}| = 4$. For each $P \in \mathbb{P}, |S_P|$ is 3, 4, 4, and 5, respectively. $\tau_{\text{threshold}} = 30\%, \xi_{\text{threshold}} = 90°$, and $|\mathbb{E}| = 3$ with the corresponding bandwidth costs of 10, 20, and 30. $\alpha_z = 1, \forall z \in \mathbb{Z}$. G^{TRG} is constructed as shown in Fig. 2(a), where $n_{z,e}$ is the node that represents the tile z with quality e, and the edge is shown as a blue line. Figs. 2(b) shows the nodes requested by the users and the MCS which is mapped from the user's channel state. The number of pairs of scans that satisfy the *registration overlap constraint* between the nodes (i.e., $\text{EW}_{z,e,z',e'}$) is shown in Fig. 2(c). We select $n_{1,2}$ with the maximum $\text{OCI}_{1,2} = \frac{(2+2+2+2+2+2+3+3+3+2+2+2+1+1+1) \times 2}{1 \times 20} = 3$ for multicasting. STS iteratively examines the node in descending order of OCI, as indicated by the red numbers in Fig. 2(a). Because none of the nodes can be registered by the existing multicasting nodes, STS adds $n_{1,2}$ and $n_{4,2}$ to $\mathbb{T}$. STS adds $n_{6,1}$ to $\mathbb{C}$, as it can be registered by the nodes in $\mathbb{T}$. The result of STS for G^{TRG} is shown in Fig. 2(a), where the red and orange circles on G^{TRG} represent the multicasting node and the registered node, respectively. The total bandwidth cost is $10 + 20 + 30 + 10 + 20 + 30 + 20 + 30 + 10 + 20 = 200$ with the multicasting nodes $\{n_{1,1}, n_{1,2}, n_{2,3}, n_{3,1}, n_{3,2}, n_{3,3}, n_{4,2}, n_{4,3}, n_{5,1}, n_{5,2}\}$.

2) Registration Replacement (RR): In addition to the overlap
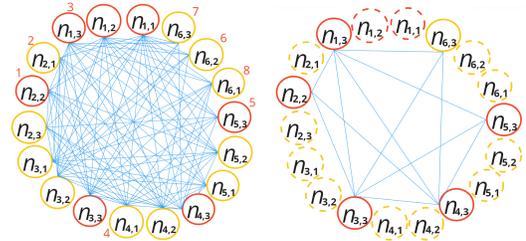
(a) The result of RR   (b) The result of TSA

Fig. 3: An illustrative example of RR and TSA.

ratio, the point cloud registration is also affected by the quality of scans and the rotation angle of scans between the nodes. To address **C2** and minimize bandwidth costs, RR iteratively replaces selected multicasting nodes with registered nodes by examining RAI. Specifically, for each registered node $n_{z,e} \in \mathbb{C}$, RR calculates the ratio of the number of scans that can be registered by $n_{z,e}$ and the nodes in $\mathbb{T}$ to the induced bandwidth cost according to $n_{z,e}$'s quality. Let $I_{z',e'}$ represent the number of scans in $n_{z',e'} \in \mathbb{T}$ that can be registered by using $n_{z,e}$ and the other nodes in $\mathbb{T}$. The ratio is denoted by $\text{RAI}_{z,e} = \frac{\sum_{n_{z',e'} \in \mathbb{T}} I_{z',e'}}{b(e)}, \forall n_{z,e} \in \mathbb{C}$. RR iteratively chooses the node $n_{z,e} \in \mathbb{C}$ with the largest $\text{RAI}_{z,e}$ to be multicasted. Once $n_{z,e}$ is selected, RR defines π as the set of multicasting nodes that can be registered by $n_{z,e}$ and the other multicasting nodes in $\mathbb{T} \setminus \pi$. If the bandwidth cost of $n_{z,e}$ is smaller than the bandwidth cost of the set π, RR replaces all the nodes in π with $n_{z,e}$ and swaps the sets in which they are (from $\mathbb{T}$ to $\mathbb{C}$, and vice versa). Otherwise, RR iteratively examines the two-hop node $n_{z',e'}$ of $n_{z,e}$ with the largest $\text{RAI}_{z',e'}$ to register multicasting nodes and update π, because the node that is more than two hops away from $n_{z,e}$ cannot be used for registration with $n_{z,e}$. If the bandwidth cost of $n_{z,e}$ and its two-hop node $n_{z',e'}$ is higher than the bandwidth cost of π, $n_{z',e'}$ will not be selected to be a multicasting node. RR then examines the next two-hop node of $n_{z,e}$. This process stops until all the two-hop nodes of $n_{z,e}$ have been examined. Note that in the above steps, if the bandwidth cost can be decreased, RR replaces the nodes in π with $n_{z,e}$ and $n_{z',e'}$ and swaps the sets in which they are (from $\mathbb{T}$ to $\mathbb{C}$, and vice versa). If the bandwidth cost still cannot be decreased after examining all the two-hop nodes of $n_{z,e}$, $n_{z,e}$ will not be chosen as a multicasting node. RR then examines the next node in $\mathbb{C}$. This process stops until all the nodes in $\mathbb{C}$ have been examined. Note that we only swap the multicasting nodes in π when the node in $\mathbb{C}$ can still be registered by the nodes in $\mathbb{T}$.

Example 2. We first choose $n_{2,2}$ with the largest $\text{RAI}_{2,2} = \frac{2+2+0+2+2+0+2+0+3+3}{20} = 0.8$ to be multicasted, where the numerator represents the sum of the number of scans in each multicasting node that can be registered by $n_{2,2}$ and the other multicasting nodes. Then, we replace $n_{5,1}$ and $n_{5,2}$ in π with $n_{2,2}$ and swap the sets in which they are, as $n_{2,2}$ has lower bandwidth cost than $n_{5,1}$ and $n_{5,2}$, i.e., $20 < 10 + 20$. RR chooses the node in descending order of RAI, as indicated by the red numbers in Fig. 3(a). Fig. 3(a) shows the result of RR.

530

The total bandwidth cost decreases from 200 to $10 + 20 + 30 + 20 + 30 + 30 + 30 = 170$ with the multicasting nodes $\{n_{1,1}, n_{1,2}, n_{1,3}, n_{2,2}, n_{3,3}, n_{4,3}, n_{5,3}\}$.

3) Tile Selection Adjustment (TSA): To deal with **C3**, TSA iteratively removes the selected multicasting nodes by examining QRI, which evaluates the two factors for a tile of each quality: 1) the number of users who request this quality and can also be satisfied by other qualities and 2) the number of users who cannot be served by this quality. Specifically, let U_z be the set of users who request tile z and $R_{z,e} \subseteq U_z$ be the set of users who request tile z with quality e. We define QRI as $\mathrm{QRI}_{z,e} = \frac{(\sum_{e' \in \mathbb{E} \setminus e} |R_{z,e} \cap R_{z,e'}|) \times |U_z \setminus R_{z,e}|}{b(e)}, \forall z \in \mathbb{Z}$. A higher QRI indicates that a node can be removed without violating the users' requirements, as they can be satisfied by other nodes. TSA iteratively chooses the tile z with the largest α_z for removal, as it is less likely to be requested by socially related users and cannot effectively minimize the bandwidth cost. Then, for each quality of z, we iteratively remove the node with the highest QRI and update QRI for the remaining nodes. Note that we remove the node only when all users' quality requirements can still be satisfied.

Each tile of a specific quality is multicasted for a multicast group, and a user can be in multiple multicast groups to receive tiles. To address **C3**, for each selected node $n_{z,e} \in \mathbb{T}$ in descending order of the quality, TSA first groups the users whose rate requirements are satisfied and can be served by $n_{z,e}$ into a multicast group. TSA will not group the user into the multicast group that receives the same tile with different qualities to ensure that the *quality of tile selection constraint* is satisfied. Then, because a user can belong to multiple groups to receive multiple tiles for registration, we select the nodes with lower bandwidth cost among the multiple pairs of tiles that can be used to register the same requested tile to minimize bandwidth cost.

Example 3. We first examine tile z_1 with the largest preference weight and $U_1 = \{u_1, u_2, u_7\}$. Then, TSA examines QRI for each quality of z_1, where $\mathrm{QRI}_{1,1} = \frac{2 \times 2}{10} = 0.4$, $\mathrm{QRI}_{1,2} = \frac{4 \times 0}{20} = 0$, and $\mathrm{QRI}_{1,3} = \frac{4 \times 0}{30} = 0$. TSA first removes $n_{1,1}$ while ensuring that the quality requirements of all users in U_1 can still be satisfied. Then, TSA removes $n_{1,2}$ with the maximum $\mathrm{QRI}_{1,2} = \frac{4 \times 0}{20} = 0$. Afterward, TSA cannot remove $n_{1,3}$ because we cannot ensure that the quality requirement of all users in U_1 can still be satisfied since $n_{1,3}$ is used to register some other nodes. Fig. 3(b) shows the final result of TSA, where the red solid circles and orange dotted circles represent the nodes being multicasted and removed, respectively. We group u_1, u_2, and u_7 into a multicast group for multicasting $n_{1,3}$ with the lowest $\mathrm{MCS} = 2$ among them. The total bandwidth cost is reduced from 170 to $30 + 20 + 30 + 30 + 30 = 140$ with the selected nodes $\{n_{1,3}, n_{2,2}, n_{3,3}, n_{4,3}, n_{5,3}\}$.

IV. EXPERIMENT

Due to the space constraint, more experimental results are presented in [17].

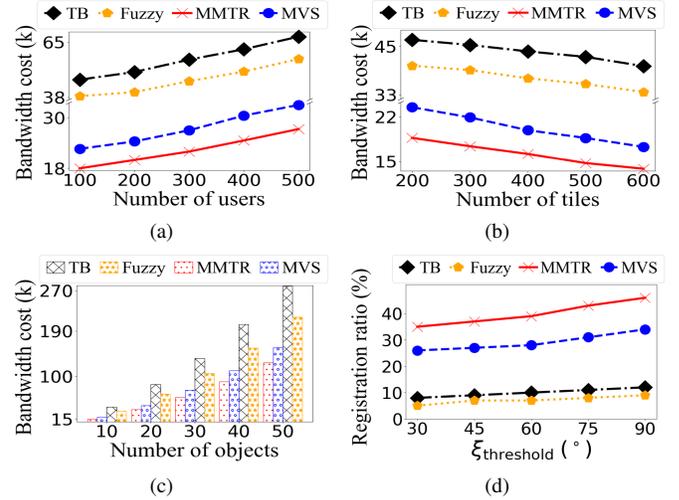
Fig. 4: Experimental results.

A. Experimental Setup

We evaluate MMTR in a single base station (BS) network with a video server. The maximum BS transmit power is 46 dBm [26]. Following 3GPP specifications [20], [25], the path loss is the macro propagation for outdoor urban areas [25]. The shadowing model is log-normal fading with 8 dB standard deviation [25]. The MCS ranges from QPSK to 256QAM [20], and the additive white Gaussian noise power spectral density is -174 dBm/Hz [26]. The users are randomly distributed over the BS's coverage area, and the default number of users is 100. The default number of objects is 10 [18], with each object having up to 24,000 points [10]. A point cloud video with a fixed size is divided into $25 \times 4 \times 2 = 200$ tiles by default [4]. Following [5], [6], we consider five qualities with $100\%, 80\%, 60\%, 40\%$, and 20% of the original number of points in a tile, which require 300Mbps, 240Mbps, 180Mbps, 120Mbps, and 60Mbps data rates, respectively. We set rotation angle threshold $\xi_{\mathrm{threshold}}$ and overlap ratio threshold $\tau_{\mathrm{threshold}}$ to $30°$ and 30% [11], respectively. The preference weight of each tile is set according to [19]. We compare MMTR with 1) multi-view synthesis (MVS) algorithm [8], fuzzy logic-based tile selection (Fuzzy) [4], and tile-based (TB) transmission [5]. We change the parameters: 1) number of users, 2) number of tiles, 3) number of objects, and 4) $\xi_{\mathrm{threshold}}$ to evaluate 1) bandwidth cost, 2) video quality, and 3) registration ratio, which is the number of tiles for point cloud registration over the total number of tiles.[8] To evaluate video quality, we follow [10] to utilize the OMNet registration model and the Stanford 3D Scan dataset to measure the metrics WS-PSNR and WS-SSIM [27]. Each result is averaged over 1000 times.

B. Experimental results

As shown in Fig. 4(a), MMTR achieves significant performance improvement over the baselines as the number of users increases, because it groups users based on the TRG

---

[8]Although Fuzzy and TB did not consider point cloud registration, we measure the number of tiles that can be registered by their selected tiles to evaluate the registration ratio without loss of generality.
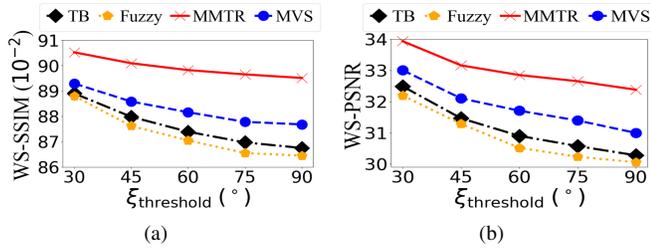
Fig. 5: Registered video quality.

and selects multicasting and registered tiles. TB and Fuzzy do not examine the chance of registration to decrease the number of transmitted tiles and induce a higher bandwidth cost. Despite the increase in the number of nodes that need to be chosen as the number of tiles increases, MMTR generates the smallest bandwidth cost in Fig. 4(b), because it examines the QRI to reduce the number of transmitted tiles to minimize the bandwidth cost. As the number of objects increases in Fig. 4(c), MMTR has more chance to register tiles by examining RAI, which evaluates the node and its two-hop neighbors to find a better set of multicasting nodes for point cloud registration with a smaller cost. In Fig. 4(d), when the registration rotation constraint becomes loose, MMTR registers more tiles than the baselines because we examine OCI to select tiles to allow more point cloud registration. However, MVS registers fewer tiles since it selects multicasting and registered tiles without examining the tiles within the rotation angle constraint to explore the possibility of registration. In Fig. 5, we evaluate the video quality with the real point cloud dataset Standard 3D Scan. WS-SSIM and WS-PSNR decrease as $\xi_{\text{threshold}}$ grows because more scans within larger rotation angles (i.e., the scans are less similar) can be registered. However, the baselines drop rapidly because they will select the scan (tile) with poor quality for registration when its cost is lower for serving an individual user. MMTR provides better video quality since it tends to select higher-quality tiles for registration, which enables registering more tiles within a smaller rotation angle. Compared Figs. 4 with 5, with a relaxed registration rotation constraint, MMTR can reduce the bandwidth cost by more than 20% with only a slight video quality loss.

V. CONCLUSION

For tile-based point cloud video transmission, this paper exploits point cloud registration for multicast to increase the flexibility of tile selection. To address the challenges of minimizing the bandwidth cost for point cloud video transmission, we formulate TRSM and prove the NP-hardness. Then, we propose MMTR, a novel algorithm with the ideas of STS, RR, and TSA designed for tile selection with point cloud registration and multicast. Experimental results manifest that MMTR can achieve better registration quality and reduce more than 20% of bandwidth cost compared to the baselines.

REFERENCES

[1] H. Yu, T. Taleb, K. Samdanis, and J. Song, "Towards supporting holographic services over deterministic 6g integrated terrestrial & non-terrestrial networks," *IEEE Network*, pp. 1–10, 2023.

[2] T. Fujihashi, T. Koike-Akino, and T. Watanabe, "Soft 2d-to-3d delivery using deep graph neural networks for holographic-type communication," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2023.

[3] M. HoloLens, "Hololens 2 x healthcare," 2023. [Online]. Available: https://reurl.cc/NyLL4q

[4] Z. Liu *et al.*, "Fuzzy logic-based adaptive point cloud video streaming," *IEEE Open J. Comput. Soc.*, vol. 1, pp. 121–130, 2020.

[5] L. Wang, C. Li, W. Dai, J. Zou, and H. Xiong, "Qoe-driven and tile-based adaptive streaming for point clouds," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2021.

[6] X. Yu *et al.*, "Pointr: Diverse point cloud completion with geometry-aware transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 12 478–12 487.

[7] M. Yeh, C.-H. Wang, D.-N. Yang, J.-T. Lee, and W. Liao, "Mobile proxy caching for multi-view 3d videos with adaptive view selection," *IEEE Trans. Mob. Comput.*, vol. 21, no. 8, pp. 2909–2921, 2022.

[8] W. Xu, Y. Cui, and Z. Liu, "Optimal multi-view video transmission in multiuser wireless networks by exploiting natural and view synthesis-enabled multicast opportunities," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1494–1507, 2020.

[9] N. Jain, S. Kumar, and L. Van Gool, "Enhanced stable view synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2023, pp. 13 208–13 217.

[10] H. Xu *et al.*, "Omnet: Learning overlapping mask for partial-to-partial point cloud registration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, October 2021, pp. 3132–3141.

[11] S. Huang *et al.*, "Predator: Registration of 3d point clouds with low overlap," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 4265–4274.

[12] X. Bai *et al.*, "D3feat: Joint learning of dense detection and description of 3d local features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 6358–6366.

[13] H. Wang *et al.*, "A survey on the metaverse: The state-of-the-art, technologies, applications, and challenges," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14 671–14 688, 2023.

[14] A. Nartker, "A first look at project starline's new, simpler prototype," May 2023. [Online]. Available: https://reurl.cc/Y0Vnjx

[15] Z. Zhang *et al.*, "Joint user grouping, version selection, and bandwidth allocation for live video multicasting," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 350–365, 2022.

[16] T. Okamoto *et al.*, "Edge-assisted multi-user 360-degree video delivery," in *Proc. IEEE 20th Consum. Commun. Netw. Conf. (CCNC)*, 2023.

[17] H.-R. Lai *et al.*, "Socially-aware tile-based point cloud multicast with registration (full version)," Nov 2023. [Online]. Available: http://mnet.cs.nthu.edu.tw/NTHU-ICC2024FullVersion.pdf

[18] S. Fan *et al.*, "Scf-net: Learning spatial contextual features for large-scale point cloud segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2021, pp. 14 504–14 513.

[19] Y. Dai and S. Han, "Sight guidance enhanced vr video transmission," in *IEEE Int. Conf. Signal Process. (ICSP)*, vol. 1, 2022, pp. 305–309.

[20] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP, Technical report (TR) 38.901, Jun 2018.

[21] D. Lee *et al.*, "Deeppro: Deep partial point cloud registration of objects," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021.

[22] H. Cai *et al.*, "Enable deep learning on mobile devices: Methods, systems, and applications," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 27, no. 3, mar 2022.

[23] X. Huang, G. Mei, and J. Zhang, "Cross-source point cloud registration: Challenges, progress and prospects," *Neurocomputing*, vol. 548, p. 126383, 2023.

[24] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Mahalanobis based point to distribution metric for point cloud geometry quality evaluation," *IEEE Signal Processing Letters*, vol. 27, pp. 1350–1354, 2020.

[25] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); radio frequency (RF) requirements for LTE pico node B," 3GPP, Technical Report (TR) 36.931, Jan 2016.

[26] M. A. Abu-Rgheff, *5G Physical Layer Technologies*. Wiley, 2019.

[27] X. Chai *et al.*, "Super-resolution reconstruction for stereoscopic om-nidirectional display systems via dynamic convolutions and cross-view transformer," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.