CEA Comité Español de Automática

XLV JORNADAS AUTOMÁTICA MÁLAGA 2024

UNIVERSIDADE DA CORUÑA

Jornadas de Automática

GeneraRX: framework de generación automática de modelos 3D para RV

Mora, P.[a,*], Ivorra, E.[a], Alcañiz, M.[a]

[a] *Instituto Universitario de Investigación de Tecnologías Centradas en el Humano, Ciudad Politécnica de la Innovación - Cubo Azul - Edif. 8B - Acceso N - Camino de Vera, s/n 46022 Valencia, España.*

To cite this article: Mora, P., Ivorra, E., Alcañiz, M. 2024. GeneraRX: framework for automatic generation of 3D models for VR. Jornadas de Automática, 45. https://doi.org/10.17979/ja-cea.2024.45.10826

Resumen

Recientemente, la realidad extendida (RX) ha tomado importancia en diferentes áreas como la educación, la salud, y la industria, aumentando la necesidad de la generación de contenido 3D de calidad personalizado. Sin embargo, esta tecnología presenta un alto nivel de complejidad técnica y grandes costes asociados. Para abordar estos problemas, presentamos GeneraRX, un framework de generación automática de modelos 3D, que busca democratizar esta tecnología implementando Inteligencia Artificial Generativa (IAG) y Modelos de Lenguaje a Gran Escala (LLM). Esta herramienta incluye todos los pasos necesarios para la generación de un objeto 3D y está completamente automatizada. Además, integra métodos del estado del arte como Zero123++ y InstantMesh, así como el novedoso Llama 3. Finalmente, GeneraRX se ha validado mediante un experimento que evalúa la usabilidad a través de un test SUS, demostrando que se ha conseguido simplificar la creación de contenido 3D, permitiendo una interacción más natural para todo tipo de usuarios y eliminando las barreras técnicas.

Palabras clave: Técnicas de inteligencia artificial, Aprendizaje automático, Interfaces hombre-máquina, Trabajo en entornos reales y virtuales, Internet de las cosas.

GeneraRX: framework for automatic generation of 3D models for VR

Abstract

Recently, extended reality (XR) has gained importance in various fields such as education, healthcare, and industry, increasing the need for high-quality, customized 3D content generation. However, this technology presents a high level of technical complexity and significant associated costs. To address this issues, we present GeneraRX, an automatic 3D model generation framework aimed at democratizing this technology by implementing Generative Artificial Intelligence (GAI) and Large Language Models (LLM). This tool implements all the necessary steps for generating a 3D object and is fully automated. Additionally, it integrates state-of-the-art methods such as Zero123++ and InstantMesh, as well as the novel Llama 3. Finally, GeneraRX has been validated through an experiment that evaluates usability via a SUS test, demonstrating that it has successfully simplified the creation of 3D content, enabling more natural interaction for all types of users and eliminating technical barriers.

Keywords: Artificial intelligence techniques, Machine Learning, Man-machine interfaces, Work in real and virtual environments, Internet of things.

1. Introducción

En los últimos años, la RX, que incluye realidad virtual (RV) y realidad aumentada (RA), ha ganado relevancia en múltiples áreas como educación, entretenimiento, industria y salud, aumentando la demanda de modelos 3D personalizados de alta calidad. Estos contenidos se generan actualmente de forma manual por un equipo técnico con conocimientos de modelado, lo que hace de este proceso un gran gasto económico y temporal, obstaculizando la expansión de las tecnologías

*Autor para correspondencia: pamogar@upv.edu.es

de RX. Recientemente, gracias a la IAG, han surgido métodos capaces de generar modelos 3D a partir de descripciones de texto o imágenes de referencia. Aunque esto reduce la necesidad de habilidades técnicas especializadas, su implementación suele ser complicada para usuarios no expertos, y requiere aplicaciones separadas para generar y posicionar los modelos.

Para abordar estas limitaciones, se presenta GeneraRX, un framework que utiliza LLM y reconocimiento de voz como interfaz persona-computador para la generación de modelos 3D. Además, está completamente automatizado e incluye todos los métodos necesarios para que el usuario pueda generar un objeto 3D personalizado a partir de su voz. Esto facilita el acceso a esta tecnología a un público más amplio. Además, GeneraRX está integrado en las gafas de RV Meta Quest 3, lo que permite a los usuarios describir y editar modelos directamente en la escena de RV, eliminando la necesidad de aplicaciones adicionales. Las principales aportaciones de nuestro proyecto son las siguientes:

- Modelos de alta calidad: GeneraRX implementa los últimos métodos del estado del arte, que han demostrado resultados de alta calidad.
- Proceso completamente automático: El flujo de trabajo está completamente automatizado una vez se proporciona la descripción por voz.
- Bajo coste computacional: El modelo se genera y carga en la escena en aproximadamente 1 minuto.
- Integración en RV: GeneraRX está integrado en las gafas Meta Quest 3. Se pueden generar y editar la posición y orientación de los modelos en la escena de RV.
- Accesibilidad mejorada: Nuestro proyecto consigue democratizar la tecnología de generación de contenido 3D personalizado y de alta calidad.

En resumen, GeneraRX proporciona una herramienta para generar modelos 3D personalizados y de calidad a usuarios de todos los niveles, interaccionando únicamente con las gafas de RV y de forma completamente automatizada. Estas aportaciones permiten a GeneraRX superar las limitaciones de los métodos tradicionales de generación de modelos 3D, promoviendo una mayor adopción de tecnologías de realidad extendida en diversas áreas como la educación, el entretenimiento, la industria y la salud.

2. Estado del arte

El área de generación automática de modelos 3D tiene gran importancia a día de hoy, y las técnicas y métodos aplicados siguen avanzando continuamente, los primeros intentos de conversión de imágenes a 3D se centraron principalmente en reconstrucciones desde una sola vista, como el proyecto Pixel2Mesh (Wang et al., 2018). Sin embargo, al disponer de únicamente una vista, este tipo de métodos presentan problemas para reconstruir el objeto completamente. Con la aparición de nuevos modelos de difusión como Stable Diffusion (Rombach et al., 2022), algunos trabajos han investigado la generación de modelos 3D condicionados por imágenes, como el trabajo de Zhou et al. (Zhou et al., 2021). Sin embargo, las múltiples vistas generadas no son consistentes. En esta línea de investigación aparece Zero123 (Liu et al., 2023), que ajusta el modelo de Stable Diffusion (Rombach et al., 2022) para generar nuevas vistas condicionadas por las posiciones relativas de las cámaras, manteniendo la consistencia y coherencia entre vistas. Tomando como base este trabajo, algunos métodos consiguen mejorar la consistencia tridimensional de los objetos, como Zero123++ (Shi et al., 2023), que es el que decidimos implementar en GeneraRX.

Una vez solucionado el problema de generar múltiples imágenes, el siguiente desafío es la reconstrucción 3D del modelo. Para este paso existen múltiples técnicas, por ejemplo, FlexiDreamer (Zhao et al., 2024) implementa FlexiCubes, que presentan más flexibilidad, pero pueden tener menos precisión. Otro enfoque es mediante el uso de optimización por Neural Radiance Fields (NeRF) (Chen et al., 2023), que proporcionan más detalle al precio de un coste computacional mayor. Por otra parte, existen las técnicas de Gaussian Splatting, esta técnica destaca por su bajo coste computacional, donde destaca el trabajo de SplatterImage (Szymanowicz et al., 2024). Finalmente, gracias a la gran disponibilidad de datos 3D a gran escala, como el dataset de Objaverse (Deitke et al., 2023) nacen los modelos de reconstrucción altamente generalizables (LRM), como por ejemplo InstantMesh (Xu et al., 2024). Este tipo de modelos han demostrado funcionar generalmente bien, y en concreto este trabajo demuestra resultados de gran calidad con un coste computacional bajo, por lo que se decide implementar en GeneraRX.

Respecto a los LLM, son un tema de investigación que también tiene gran importancia y un avance muy rápido. Los primeros modelos como GPT-1 (Radford et al., 2018) comenzaron a demostrar la habilidad de generar texto coherente y de compresión lectora, y las siguientes versiones de GPT mejoraron estas habilidades. El modelo BERT (Kenton and Toutanova, 2019) introdujo la idea de un preentrenamiento bidireccional y Text-to-Text Transfer Transformer (T5). Más recientemente, ha aparecido Llama 3 (Meta Platforms, 2024a), con su modelo de 70 mil millones de parámetros, que ha demostrado una capacidad superior en generación y comprensión de texto, así como en tareas complejas de lenguaje, es por estos resultados que se implementa en GeneraRX.

3. Hardware y software

En esta sección se va a exponer el material utilizado para el desarrollo y la ejecución de GeneraRX. El primer elemento es una workstation con un procesador Intel Xeon W-3235, con una GPU NVIDIA TITAN RTX y 128 GB de memoria RAM DDR4. Este ordenador se utiliza para ejecutar el script principal que transforma la descripción proporcionada por el usuario a un modelo 3D de calidad, adicionalmente funciona como servidor de una RestAPI con la que se gestiona la comunicación entre dicha workstation y las gafas de RV.

Respecto a las gafas de RV, se utilizan las Meta Quest 3 de 512 GB. Estas gafas presentan una excelente relación calidad-precio, siendo mucho más asequibles que otros dispositivos de RV como las Vision Pro de Apple, manteniendo una alta calidad. Adicionalmente, las Meta Quest 3 son completamente

autónomas y disponen de 2 a 3 horas de batería, tiempo suficiente para generar un entorno de RV personalizado.

Por otra parte, se utiliza Python 3.10 para ejecutar tanto el script principal que genera el modelo 3D como para abrir el servidor de RestAPI. Y para el desarrollo y ejecución de la aplicación de RV se utiliza Unity3D 2022.3.24f1s.

4. Método

En este apartado se va a explicar el funcionamiento de GeneraRX. En concreto, se van a explicar los métodos implementados, la implementación de estos en un mismo flujo, la comunicación entre el hardware y la aplicación de Unity3D. El flujo de trabajo de GeneraRX está dividido entre dos entornos de ejecución, un script de Python y una aplicación de Unity3D. Dicho flujograma se puede ver en la Figura 1.

Como se puede ver en la imagen, la aplicación comienza cuando el usuario activa la captura de voz y da una descripción del objeto a generar. Debido a medidas de privacidad de las Meta Quest 3, este audio se procesa mediante Wit.ai (Meta Platforms, 2024b) y se proporciona la transcripción de texto. Una vez se obtiene el texto, se ha de enviar a la aplicación de Python, esta comunicación se realiza mediante una RestAPI, este texto se utiliza como entrada a la aplicación de Python que empieza su ejecución. El primer paso es generar una imagen de referencia para la reconstrucción 3D, para esto se utiliza StableDiffusion XL (SDXL) (Podell et al., 2023), un modelo de IAG de acceso abierto, seguidamente se le retira el fondo para aislar completamente el objeto a generar, utilizando esta imagen se generan 6 diferentes puntos de vista (POV) mediante Zero123++ (Shi et al., 2023). De forma paralela, utilizando Llama 3 y el texto transcrito, se obtiene una escala aproximada del objeto que se quiere generar. Mediante las 6 imágenes obtenidas se genera el modelo 3D con InstantMesh (Xu et al., 2024), y se utiliza la escala para modificar el resultado obtenido. Finalmente, este modelo 3D se envía a la aplicación de las gafas a través de WiFi 5G, otorgando mayor libertad de movimiento al usuario. A continuación, se carga el modelo en el entorno, donde el usuario es libre de modificar la posición y orientación de los objetos (pose), para generar el entorno 3D deseado.

4.1. StableDiffusion XL

El primer paso del script de Python utiliza SDXL (Podell et al., 2023), una versión avanzada del modelo de inteligencia artificial de difusión de imágenes StableDiffusion (Rombach et al., 2022). SDXL emplea técnicas de difusión para generar imágenes de alta calidad a partir de descripciones textuales. Destaca por su capacidad para producir imágenes de mayor resolución y detalle, manteniendo coherencia semántica con las descripciones proporcionadas. Esto lo hace ideal para aplicaciones que requieren imágenes visualmente impactantes y precisas, como el diseño gráfico, la publicidad y la creación de contenido multimedia.

Para nuestra implementación, utilizamos el prompt de texto que da el usuario como base para generar la imagen, y añadimos prompts como “realista”, “fondo blanco” y “un solo objeto”, para generar una imagen de alta calidad donde únicamente se pueda ver el objeto a reconstruir y que presente un fondo que sea fácil de retirar automáticamente. A partir de la imagen generada se obtendrán los 6 POV para realizar la reconstrucción, por lo que este paso es de vital importancia para garantizar la calidad del modelo a generar.

4.2. Zero123++

Una vez obtenida la imagen, se requiere obtener múltiples puntos de vista del objeto. Para ello, se emplea Zero123++ (Shi et al., 2023), un modelo de difusión condicional que mejora las inconsistencias geométricas al generar vistas múltiples de un objeto, basado en StableDiffusion. Zero123++ utiliza técnicas avanzadas de condicionamiento global y local para mejorar la calidad de las imágenes generadas. En lugar de generar cada vista de manera independiente, modela la distribución conjunta de las vistas, logrando una representación más coherente del objeto en 3D. Además, ajusta finamente los pesos y parámetros del modelo original, integrando mecanismos de atención y escalado de referencia para mantener la coherencia y calidad de las imágenes. Este enfoque supera las limitaciones de los modelos anteriores, proporcionando imágenes 3D multi-vista de alta calidad y consistencia.

Para su implementación en el trabajo de GeneraRX, se utiliza una versión reentrenada de Zero123++ que genera seis vistas con un fondo blanco, proporcionada como parte de InstantMesh (Xu et al., 2024). Para reentrenar esta versión utilizan los modelos 3D de Objaverse (Deitke et al., 2023), generando imágenes sintéticas del objeto desde las seis diferentes vistas y con un fondo blanco, consiguiendo que las imágenes generadas por Zero123++ utilicen estos mismos puntos de vista y generen un fondo blanco que facilita aislar el objeto. Utilizando este método, se generan las seis imágenes del objeto 3D generado en el paso anterior, que se utilizan para generar el modelo 3D.

4.3. Llama 3

De forma paralela a la obtención de las imágenes, se procesa la entrada de texto para obtener la escala aproximada del objeto, debido a que la variedad de objetos que se pueden pedir es muy amplia, se utiliza Llama 3 (Meta Platforms, 2024a) para obtener las medidas. Llama 3 es un LLM desarrollado por Meta, diseñado para mejorar las capacidades de comprensión textual, que se entrena con un enorme conjunto de datos que abarca más de 15 billones de entradas.

En concreto, se utiliza el modelo Llama 3 de 70 mil millones de parámetros para obtener las dimensiones aproximadas de varios objetos. Se introduce un prompt específico para que devuelva una respuesta consistente en el formato ”Dimensiones=[x, y, z]”, donde x es el ancho, y es el largo y z es el alto en centímetros del objeto descrito. Al sustituir la descripción proporcionada por el usuario, se obtiene una respuesta coherente que describe las dimensiones del objeto deseado, lista para procesar y escalar el modelo 3D.

4.4. InstantMesh

Para transformar las imágenes generadas en un modelo 3D, se emplea InstantMesh (Xu et al., 2024), un método que ha demostrado una alta calidad con un coste computacional bajo. El proceso comienza alineando y fusionando las vistas para crear una representación coherente del objeto. Luego, se extraen características relevantes de cada vista mediante una

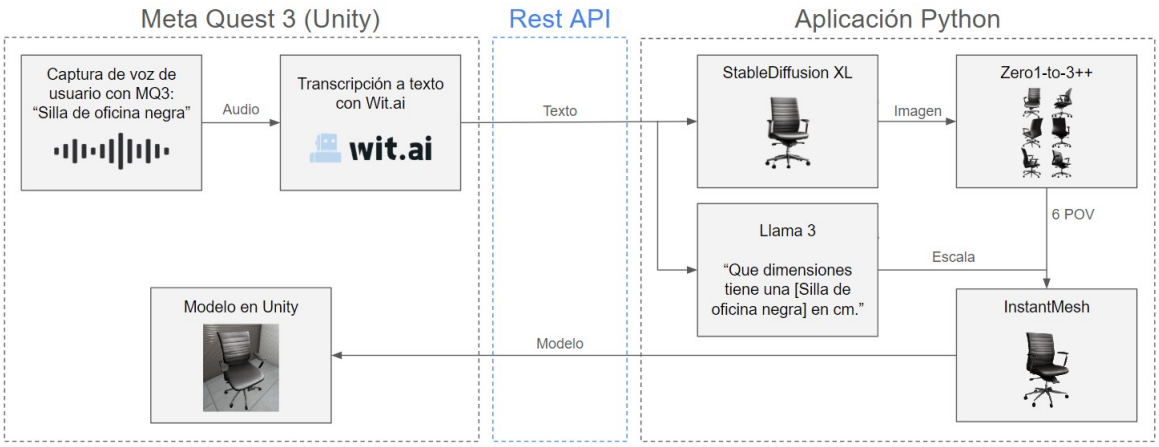

Figura 1: Flujograma del funcionamiento de GeneraRX.

arquitectura de transformador. InstantMesh predice una malla 3D inicial y la refina para suavizar superficies, corregir artefactos y ajustar detalles finos, utilizando datos geométricos como mapas de profundidad y normales. Además, integra un módulo de extracción de FlexiCubes, para mejorar la eficiencia y calidad de la reconstrucción. El resultado es una malla 3D final de alta calidad y coherencia, adecuada para aplicaciones de realidad virtual, entre otras. Las métricas completas de PSNR y SSIM del método se pueden consultar en su artículo, situándolo entre los mejores del estado del arte.

Utilizando este método, se consigue una reconstrucción 3D de alta calidad y realista, que se asemeja a la descripción dada por el usuario. Una vez se obtiene el objeto, se guarda en una carpeta, desde donde la aplicación en Unity3D podrá acceder y cargarlo en el entorno virtual. La información de la ruta de la malla, junto con la escala extraída de Llama3, se comunica a la aplicación de Unity3D mediante RestAPI.

4.5. *Comunicación mediante RestAPI*

La aplicación consta de dos entornos distintos: un código en Python que genera los modelos 3D y una aplicación en Unity3D que interactúa con el usuario y carga los modelos generados. Para conectar estos dos entornos, se utiliza una RestAPI, que es una interfaz de programación que emplea el protocolo HTTP para la comunicación entre aplicaciones. En nuestro caso, esta comunicación se compone de un servidor HTTP, alojado en el ordenador descrito en la sección 3 y dos clientes, siendo estos los entornos de Python y Unity3D.

Las RestAPI implementada facilita la transferencia de dos paquetes de información. Primero, la aplicación de Unity3D envía la descripción de texto al servidor, donde se guarda en la base de datos. Mientras tanto, la aplicación de Python lanza peticiones de la descripción de texto al servidor cada 5 segundos, una vez hay una disponible, la obtiene y comienza a generar el objeto. El otro paquete de información es la información del modelo, que contiene la ruta y la escala del modelo generado. Cuando la aplicación de Python termina de generar el modelo, este la envía al servidor, que almacena la información. Simultáneamente, la aplicación de Unity3D consulta cada 5 segundos al servidor para obtener la información del modelo, una vez hay un paquete de información disponible, la obtiene y carga y escala el modelo en la escena de RV.

4.6. *Aplicación de Unity3D*

Para finalizar este apartado se va a explicar el funcionamiento de la aplicación de Unity3D que se ejecuta en la Meta Quest 3. La escena inicialmente presenta una habitación simple y genérica que se encuentra completamente vacía, únicamente con paredes, techo y suelo. Una vez el usuario carga en el entorno VR, puede pulsar el botón principal de los controladores para que el programa comience a escuchar el audio, en este momento el usuario deberá proporcionar una breve descripción de audio del objeto que quiere generar, que se transcribe automáticamente a texto y se envía a la aplicación de Python mediante RestAPI.

Una vez enviada la descripción de texto, la aplicación consulta cada 5 segundos si se han añadido nuevos modelos a la base de datos, en el momento en el que se añade un nuevo modelo a la lista, este se carga automáticamente accediendo mediante WiFi 5G a la ruta del modelo y utilizando un plugin de Unity3D llamado TriLib 2 (Reis, 2021). Una vez se carga el modelo se le asocian unos scripts prefabricados para hacer que el objeto se pueda agarrar y mover, adicionalmente se le asigna una malla y un cuerpo rígido para colisionar con el resto de objetos de la escena y simular la gravedad.

Repitiendo este proceso se obtiene una escena poblada de objetos personalizados que tiene físicas y colisiones realistas y se puede modificar la distribución libremente.

4.7. *Test de usabilidad*

Con la finalidad de validar el funcionamiento y obtener resultados sobre la usabilidad del framework de GeneraRX, se ha diseñado un experimento que se ha evaluado mediante un test SUS (Brooke et al., 1996). Este test nos proporciona información valiosa sobre la complejidad de la herramienta, y nos ayuda a identificar puntos débiles de nuestro proyecto.

El test se ha realizado a 10 personas de perfiles técnicos variados, entre los que se encuentran desarrolladores de RV, psicólogos y programadores. El escenario de la prueba consiste en una habitación poblada con tres modelos previamente generados con GeneraRX y posicionados. El experimento comienza con una explicación detallada de los controles de aplicación, con los que podrán generar los modelos e interaccionar con ellos. Una vez en la escena de RV, se le pide a los usuarios que generen al menos 3 modelos y que interaccionen

(a) Imagen del objeto (b) Múltiples vistas del objeto (c) Modelo 3D cargado en RV

Figura 2: Resultados paso a paso de la generación de [Silla de oficina negra].

Figura 3: Resultados de múltiples reconstrucciones 3D.

libremente con ellos y con los objetos previamente cargados, se recomienda que prueben las interacciones de agarre y la colisión entre objetos. La prueba finaliza cuando los usuarios han generado al menos los 3 objetos y han interactuado suficientemente con la escena, aproximadamente en unos 10 minutos. Para finalizar, los usuarios rellenan un test SUS, cuyos resultados se procesan y se muestra en la sección de resultados.

5. Resultados

En esta sección se van a mostrar resultados obtenidos durante la validación de la aplicación de GeneraRX, específicamente, se van a mostrar los resultados paso a paso de la generación de un modelo, ilustrado en la Figura 2, también se van a mostrar varios modelos generados junto con la descripción de texto dada por el usuario, y finalmente se van a mostrar los resultados del test SUS realizado para analizar la usabilidad de la aplicación y detectar posibles puntos de mejora.

La generación del modelo comienza con el usuario describiendo el objeto deseado, en este caso, "Silla de oficina negra", la transcripción del audio y su envío tardan aproximadamente 2 segundos. A continuación, se genera una imagen inicial utilizando SDXL. Posteriormente, se elimina el fondo de la imagen para aislar completamente el objeto, como se muestra en la Figura 2(a), este paso tarda alrededor 2 segundos. Como se puede ver en la imagen, se consigue generar una imagen realista que se corresponde con la descripción dada con el usuario, además, se consigue aislar retirando el fondo para poder realizar correctamente la reconstrucción.

El siguiente paso es generar diferentes puntos de vista a partir de la imagen, esto se realiza mediante una versión adaptada de Zero123++ que genera seis imágenes desde puntos de vista fijados y con fondos blancos en aproximadamente 15 segundos. El resultado de este paso se puede ver en la Figura 2(b). Analizando la imagen, se puede observar que se consiguen generar diferentes puntos de vista realistas y que mantienen la estructura del objeto de forma consistente. En paralelo a los dos anteriores pasos, se utiliza Llama3 para obtener una escala aproximada del objeto, en concreto, se pregunta a Llama3 por las dimensiones que podría tener un objeto de ese tipo y devuelve, en unos 5 segundos, la siguiente respuesta: "Dimensiones=[50, 60, 100]", donde el primer valor es el ancho, el segundo la profundidad y el tercero la altura, con lo que nos proporciona una escala realista.

Para finalizar el procesado, se utiliza InstantMesh para generar el objeto 3D, que se obtiene en aproximadamente 30 segundos. Una vez se genera el modelo, se envía su información a la aplicación de RV, donde se carga y se escala en cerca de 4 segundos, haciendo un total de 58 segundos. Una vez en la escena, el usuario puede modificar su pose para conseguir la distribución deseada. En la Figura 2(c) se puede observar el modelo cargado en la escena virtual. Adicionalmente, la Figura 3 muestra más modelos generados con GeneraRX junto con su descripción de texto, aquí se puede observar que estos modelos también tienen un buen nivel de detalle y se asemejan a la descripción. Para finalizar el apartado de resultados, la Figura 4 muestra los resultados del test SUS.

Se puede observar que los resultados del test SUS son satisfactorios, ya que siempre superan la puntuación promedia de 68 (Sauro and Lewis, 2016). Respecto al valor medio, se obtiene una puntuación de 85,75, un resultado que indica una muy buena usabilidad y bajo nivel de complejidad de la herra-

mienta GeneraRX. Adicionalmente, se preguntó a los usuarios si preferían que los modelos se pareciera más a la descripción dada, o que fueran de mejor calidad, y 9 de los 10 participantes respondieron que preferirían que los modelos se asemejaran más a la descripción proporcionada, lo que se alinea con la finalidad de GeneraRX de crear modelos personalizados, en vez de utilizar modelos genéricos de gran calidad.

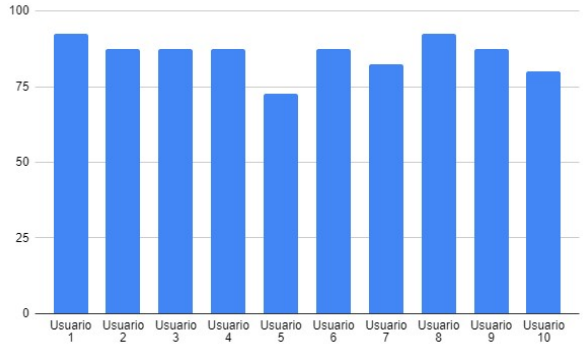

Figura 4: Resultados del test SUS.

Finalmente, gracias a comentarios de los participantes, se identificaron algunos puntos débiles de la aplicación. En concreto, al usar un agarre con los controladores y a poca distancia, hace que mover objetos grandes como sofás resulte complicado, ya que tapan parcialmente la vista de los usuarios. Otro comentario fue sobre la escala de los objetos, ya que algunas veces aparecían más pequeños de lo que es usuario hubiera esperado, por ejemplo, al generar "Un caballo", se escaló a tamaño de un juguete, en vez de al tamaño de un caballo real. Además, se identificó que durante la generación de la imagen de referencia, la IAG genera mejores resultados si el prompt se proporciona en inglés en vez de en español.

6. Conclusiones

En conclusión, este trabajo presenta una nueva aplicación que es capaz de generar modelos 3D personalizados de calidad durante la ejecución de una escena de realidad virtual, que habilita a usuarios sin conocimientos técnicos de modelado a generar escenas personalizadas para su uso en aplicaciones de realidad virtual. Adicionalmente, el proceso está completamente automatizado una vez el usuario proporciona la descripción del objeto, y se genera y carga en la escena en aproximadamente un minuto.

Para finalizar, como trabajos futuros se pretende trabajar sobre los comentarios de los usuarios, añadiendo la posibilidad de realizar un agarre a distancia, procesando la descripción de texto para intentar mejorar la predicción de la escala y proporcionando una herramienta a los usuarios para modificar el tamaño de los objetos. Además, se pretende procesar la descripción del objeto para eliminar muletillas y otros elementos de la frase que puedan empeorar la generación del modelo. Finalmente, se podría contextualizar Llama 3 e integrarlo como un asistente para el usuario.

Agradecimientos

Este trabajo ha sido subvencionado por el proyecto PAID-06-23 "GeneraRX: Democratización de la Creación de Contenidos 3D para Realidad Extendida mediante Inteligencia Artificial" concedido por la Universitat Politècnica de Valencia. El autor Mora. P. es beneficiario de una ayuda de Formación de Profesorado Universitario concedida por el Ministerio de Universidades.

Referencias

Brooke, J., et al., 1996. Sus-a quick and dirty usability scale. Usability evaluation in industry 189 (194), 4–7.

Chen, H., Gu, J., Chen, A., Tian, W., Tu, Z., Liu, L., Su, H., 2023. Single-stage diffusion nerf: a unified approach to 3d generation and reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2416–2425.
DOI: 10.48550/arXiv.2304.06714

Deitke, M., Schwenk, D., Salvador, J., Weihs, L., Michel, O., VanderBilt, E., Schmidt, L., Ehsani, K., Kembhavi, A., Farhadi, A., 2023. Objaverse: a universe of annotated 3d objects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13142–13153.
DOI: 10.48550/arXiv.2212.08051

Kenton, J. D. M.-W. C., Toutanova, L. K., 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of naacL-HLT. Vol. 1. p. 2.
DOI: 10.48550/arXiv.1810.04805

Liu, R., Wu, R., Van Hoorick, B., Tokmakov, P., Zakharov, S., Vondrick, C., 2023. Zero-1-to-3: zero-shot one image to 3d object. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9298–9309.
DOI: 10.48550/arXiv.2303.11328

Meta Platforms, I., 2024a. Introducing meta llama 3: the most capable openly available llm to date.
URL: https://ai.meta.com/blog/meta-llama-3/

Meta Platforms, I., 2024b. Wit.ai.
URL: https://wit.ai/

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., Rombach, R., 2023. Sdxl: improving latent diffusion models for high-resolution image synthesis. arXiv preprint arXiv:2307.01952.
DOI: 10.48550/arXiv.2307.01952

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al., 2018. Improving language understanding by generative pre-training. Preprint.

Reis, R., 2021. Trilib 2.
URL: https://ricardoreis.net/trilib-2/

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695.
DOI: 10.48550/arXiv.2112.10752

Sauro, J., Lewis, J. R., 2016. Quantifying the user experience: practical statistics for user research. Morgan Kaufmann.

Shi, R., Chen, H., Zhang, Z., Liu, M., Xu, C., Wei, X., Chen, L., Zeng, C., Su, H., 2023. Zero123++: a single image to consistent multi-view diffusion base model. arXiv preprint arXiv:2310.15110.
DOI: 10.48550/arXiv.2310.15110

Szymanowicz, S., Rupprecht, C., Vedaldi, A., 2024. Splatter image: Ultra-fast single-view 3d reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10208–10217.
DOI: 10.48550/arXiv.2312.13150

Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., Jiang, Y.-G., 2018. Pixel2mesh: generating 3d mesh models from single rgb images. In: Proceedings of the European conference on computer vision (ECCV). pp. 52–67.
DOI: 10.48550/arXiv.1804.01654

Xu, J., Cheng, W., Gao, Y., Wang, X., Gao, S., Shan, Y., 2024. Instantmesh: efficient 3d mesh generation from a single image with sparse-view large reconstruction models. arXiv preprint arXiv:2404.07191.
DOI: 10.48550/arXiv.2404.07191

Zhao, R., Wang, Z., Wang, Y., Zhou, Z., Zhu, J., 2024. Flexidreamer: single image-to-3d generation with flexicubes. arXiv preprint arXiv:2404.00987.
DOI: 10.48550/arXiv.2404.00987

Zhou, L., Du, Y., Wu, J., 2021. 3d shape generation and completion through point-voxel diffusion. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5826–5835.
DOI: 10.48550/arXiv.2104.03670