Interactive Perception & Graphics for A Universally Accessible Metaverse



Ruofei Du

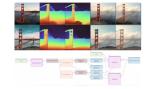
Senior Research Scientist / Manager Google AR www.ruofeidu.com Twitter: DuRuofei@ me@duruofei.com

Self Intro



🛗 🏛 in 🛩 f 💿 🗘 🔯

Featured Publications



Rapsai: Accelerating Machine Learning Prototyping of Multimedia Applications Through Visual Programming

Ruofei Du, Na Li, Jing Jin, Michelle Carney, Scott Miles, Maria Kleiner, Xiuxiu Yuan, Yinda Zhang, Anuva Kulkarni, Xingvu "Bruce" Liu, Ahmed Sabie, Sergio Escolano, Abhishek Kar, Ping Yu, Ram Iyengar, Adarsh Kowdle, and Alex Olwal

pdf, lowres, doi | project, video | cited by, cite



DepthLab: Real-Time 3D Interaction With Depth Maps for Mobile Augmented Reality 14K Installs

Ruofei Du, Eric Turner, Maksym Dzitsiuk, Luca Prasso, Ivo Duarte, Jason Dourgarian, Joao Afonso, Jose Pascoal, Josh Gladstone, Nuno Cruces, Shahram Izadi, Adarsh Kowdle, Konstantine Tsotsos, and David Kim Software and Technology (UIST), 2020.

pdf, doi | website, project, video, slides, code, demo, supp | cited by, cite



Stylization of Multiview Video Textures Microsoft TechFest 2018

Ruofei Du, Ming Chuang, Wayne Chang, Hugues Hoppe, and Amitabh Varshney

pdf, lowres, doi | website, project, video, slides | cited by, cite



Visual Captions: Augmenting Verbal Communication With On-the-Fly Visuals

Xingyu "Bruce" Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Peggy Chi, Xiang "Anthony" Chen, Alex Olwal, and Ruofei Du Proceedings of the 2023 CHI Conference on Human Factors in

Computing Systems (CHI), 2023.

pdf, lowres, doi | project, video | cited by, cite



Geollery: A Mixed Reality Social Media Platform Online Demo of a Metaverse of Mirrored World

Ruofei Du, David Li, and Amitabh Varshney Proceedings of the 2019 CHI Conference on Human Factors in

pdf, doi | website, project, video, slides, demo | cited by, cite



Montage4D: Real-Time Seamless Fusion and





Social Street View: Blending Immersive Street Views With Geo-Tagged Social Media Best Paper Award

Ruofei Du and Amitabh Varshney Technology (Web3D), 2016.

pdf, lowres, doi | website, project, video, slides | cited by, cite

Self Intro Ruofei Du (杜若飞)

Ruofei Du is a Senior Research Scientist at Google and works on creating novel interactive technologies for virtual and augmented realityresearch covers a wide range of topics in VR and AR, including AR interaction (**DepthLab**, **Ad** hoc UI), augmented communication (**CollaboVR**), mixed-reality social platforms (**Geollery**), video-based rendering (**Montage4D**), gaze-based interaction (**GazeChat**, **Kemel Foveated Rendering**), and deep learning in graphics (**3D Representation**, **HumanGPS**, **Sketch Colorization**). His research has been featured by Engadget, The Verge, PC Magazine, VOA News, cnBeta, etc. Du serves as an **Associate Editor** for IEEE Transactions on Circuits and Systems for Video Technology and Frontiers in Virtual Reality. He also served as a committee member in CHI 2021-2023, UIST 2022, and SIGGRAPH Asia 2020 XR. He holds 3 US patents and has published over 30 peer-reviewed publications in top venues of HCI, Computer Graphics, and Computer Vision, including CHI, SIGGRAPH Asia, UIST, TVCG, CVPR, ICCV, ECCV, ISMAR, VR, and I3D. Du holds a Ph.D. and an M.S. in Computer Science from University of Maryland, College Park, and a B.S. from ACM Honored Class, Shanghai Jiao Tong University. Website: https://duruofei.com

Google publications

24 publications

(i)

Personal Website Google Scholar

Ruofei Du

About

 Research
 Image: Human-Computer Interaction and Visualization
 Image: Human-Computer Interaction and Visualizat

Authored publications coog Fiters Sort by: Year ~ Research areas * Research areas * Year * ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Triggs in Video Meetings Ericht Lus and Unit Research areas * Year * ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Trings in Video Meetings Ericht Hu, Jens Emil Grantax, Wen Ying, Budel Bu, Seopkock Heo - Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI), ACM (to appear) Visual Capitions: Augmenting Verbal Copies of Physical Objects for Sharing Things in Video Meetings Ericht Hu, Jens Emil Grantax, Wen Ying, Budel Bu, Seopkock Heo - Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI), ACM (to appear)

		Autanate Kai, Filig Tu, Kain iyengai, Autanat Kowute, Alex owai • Frodeedings of the 2023 of the otherence on Human Factors in comparing systems (only, Kow (to appear)				
	+	ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Things in Video Meetings Erzhen Hu, Jens Emil Grønbæk, Wen Ying, <u>Rudfel Du</u> , Seongkook Heo 🔹 Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI), ACM (to appear)	i			
		Visual Captions: Augmenting Verbal Communication with On-the-fly Visuals Xngyu Tance Lu, Viadimir Kirlyik, Xuxiu Yuan, Eeggy.Chi, Xiang Xnthony' Chen, <u>Alex Clivial, Ruofel Du</u> · Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI, ACM (to appear)	i			
		"Slurp" Revisited: Using 'system re-presencing' to look back on, encounter, and design with the history of spatial interactivity and locative media Shengzhi Wu, Darsgh Byrne, <u>Buofel Du</u> , Molly Steenson · ACM Conference on Designing Interactive Systems, ACM (2022)	i			
		OmniSyn: Synthesizing 360 Videos with Wide-baseline Panoramas David Li, Yinda Zhang, Christian Haene, Danhang "Danny" Tang, Amitabh Varshney, <u>Ruofei Du</u> 🔹 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE	i			
		Opportunistic Interfaces for Augmented Reality: Transforming Everyday Objects into Tangible 6DoF Interfaces Using Ad hoc UI Budel Lu, Mathieu Le Go., <u>Blac Ukual</u> , Shenghi Wu, Danhang Danhary Tang, Yinda Zhang, Jun Zhang, David Joseph New Tan, <u>Eddarico Tombar</u> , David Kim · Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, ACM	i			
		PRIF: Primary Ray-based Implicit Function	i			

Self Intro _{Ruofei Du} (杜若飞)

for	Interactive Perception & Graphics Lead / Manager, <u>Google</u> Verified email at google.com - <u>Homepage</u>					
J.J.	VR / AR Interactive Perception Interactive Graphics Human Computer In	teraction Meta	verse	Cited by		VIEW
					All	Since 2
TITLE		CITED BY	YEAR	Citations h-index	870 16	
R Du, EL Turner, M I	time 3D Interaction with Depth Maps for Mobile Augmented Reality Dztisluk, L Prasso, I Duarte, J Dourgarian, J Atonso, edings of the 33rd Annual ACM Symposium on User Interface	104	2020	i10-index	23	
Kernel Foveated Rendering X Meng, R Du, M Zwicker, A Varshney Proceedings of the ACM on Computer Graphics and Interactive Techniques			2018			
Language-based Colorization of Scene Sketches C Zou, H Mo, C Gao, R Du, H Fu # ACM Transactions on Graphics (SIGGRAPH Asia 2019) 38 (6), 1-16			2019		111	
Montage4D: Rea R Du, M Chuang, W	al-time Seamless Fusion and Stylization of Multiview Video Textures Chang, H Hoppe, A Varshney Graphics Techniques (ACM I3D 2018) 8 (1), 1-34	57 *	2019	2016 2017 2018 201	9 2020 2021 202	2 2023
C Zou, Q Yu, R Du,	Ri <mark>chly-Annotated Scene Sketches</mark> H Mo, YZ Song, T Xiang, C Gao, B Chen, H Zhang e on Computer Vision (ECCV), 2018	55	2018	Public access		VIEW
R Du, D Li, A Varshn	d Reality Social Media Platform ey fings of the 2019 CHI Conference on Human Factors in	47	2019	not available Based on funding m	andatas	availa
Z He, R Du, K Perlin	configurable Framework for Creative Collaboration in Virtual Reality	44	2020	Dased on funding in	anuates	
X Meng, R Du, A Va	guided Foveated Rendering rshney n Visualization and Computer Graphics (TVCG) 26 (5), 1972	38	2020	Co-authors	rshney	VIEW
Video Fields: Fu R Du, S Bista, A Var	sing Multiple Surveillance Videos into a Dynamic Virtual Environment	28	2016	Alex Olwal	ge of Computer, Research Scien	
Multiresolution D Z Chen, Y Zhang, K	leep Implicit Functions for 3D Shape Representation Genova, S Fanello, S Bouaziz, C Haene, R Du, mational Conference on Computer Vision (ICCV)	27	2021		f Maryland, Colle	ege
A Log-Rectilinea	r Transformation for Foveated 360-degree Video Streaming CD Brumar, A Varshney s on Visualization and Computer Graphics (TVCG Honorable	26	2021	Danhang Ta Research S Que yinda Zhan Google Res	icientist, Google g	
Evaluating Hapti Text Using Finge	c and Auditory Directional Guidance to Assist Blind People in Reading Printed r-Mounted Cameras	d 26	2016	Adarsh Kov		eer a
ACM Transactions o	Jh, C Jou, L Findlater, DA Ross, JE Freehlich n Accessible Computing (TACCESS) 9 (1), 1-38			David Kim Staff Softwa	are Engineer at G	Google
R Du, A Varshney	w: Blending Immersive Street Views with Geo-tagged Social Media he 21st International Conference on Web3D Technology (Best	26	2016		Allen School of C	omp
Y Jiang, R Du, C Lut	aptive GUI Layout with OR-Constraints teroth, W Stuerzlinger	25	2019	Sean Ryan Research S	Fanello icientist and Man	ager
CHI '19: Proceeding	s of the 2019 CHI Conference on Human Factors in			🕋 Haoran Mo	(莫浩然)	

Follow

GET MY OWN PROFILE

VIEW ALL

Since 2018

821

15

0

VIEW ALL 19 articles available

VIEW ALL

>

>

>

>

>

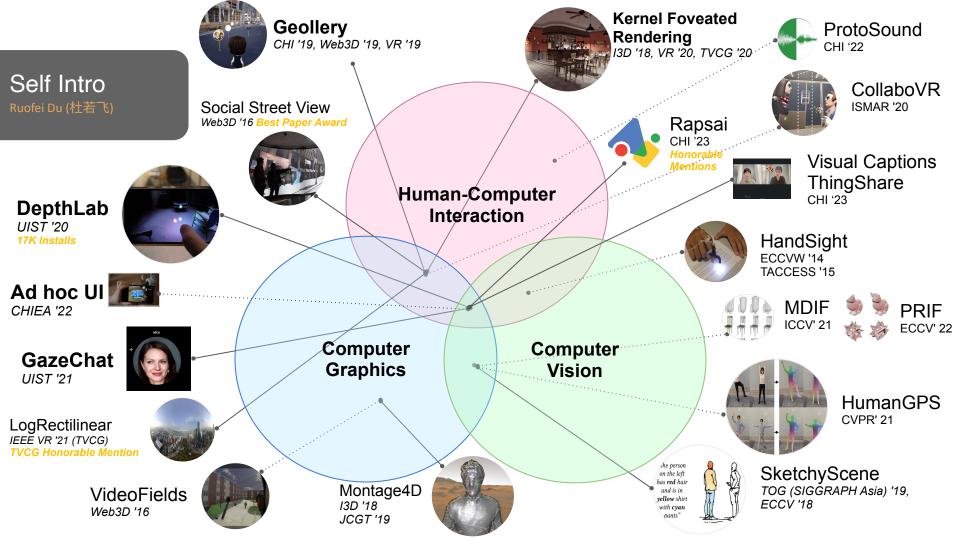
>

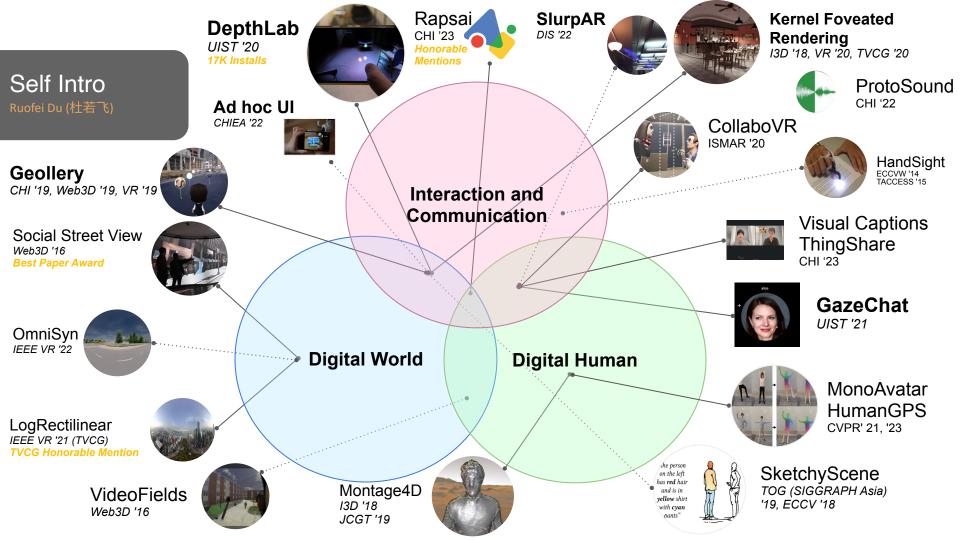
>

>

~

Ruofei Du





Interactive Perception & Graphics for A Universally Accessible Metaverse



Ruofei Du

Senior Research Scientist / Manager Google AR www.ruofeidu.com Twitter: DuRuofei@ me@duruofei.com

Metaverse

How Metaverse is defined by academia and industry?

Snow Cras

Neal Stephenson, 1992.

A STEVEN SPIELBERG FILM

Origin Ready Player One







Co-presence Gaming Accessibility Privacy

Security

<u>Blockchain</u> **Extended Reality (XR)** Things **Economics Digital Twin** NFT **Augmented Reality** of Metaverse nternet **The Future of Internet**

Virtual Reality

VR OS Mirrored World

Avatars Wearable AI Vision

Decentralization

Web 3.0

Neural

What about a non-technical perspective?

Metaverse envisioned a *persistent* digital world,

where people are fully connected as virtual representations.

More importantly, what research directions shall we devote to Metaverse?



metaverse → a medium to *make information more useful and* accessible and help people to live a better physical life Interactive Perception & Graphics for a Universally Accessible Metaverse

Chapter One · Mirrored World & Real-time Rendering

Chapter Two · Computational Interaction: Algorithm & Systems

Chapter Three · Digital Human & Augmented Communication

Chapter One · Mirrored World & Real-time Rendering



Social Street View Web3D '16 Best Paper Award Geollery CHI '19, Web3D '19, VRW '19 Kernel Foveated Rendering I3D '18, VR '20, TVCG '20 LogRectilinear, OmniSyn IEEE VR '21 (TVCG), VRW '22 TVCG Honorable Mention

Geollery.com & Social Street View: Reconstructing a Live Mirrored World With Geotagged Social Media

Hi, friends!

Ruofei Du⁺, David Li⁺, and Amitabh Varshney

{ruofei, dli7319, varshney}@umiacs.umd.edu | www.Geollery.com | ACM CHI 2019 & Web3D 2016 Best Paper Award & 2019



Greetings!

UMIACS

THE AUGMENTARIUM VIRTUAL AND AUGMENTED REALITY LAB AT THE UNIVERSITY OF MARYLAND



Hello!









Motivation





1

10

20

m)

100

199)

1

Liuti Chan ful Irre Like

Uke - Comment - Share

June Che The

81

Leyla Norooz





image courtesy: instagram.com, facebook.com. twitter.com

Motivation 2D layout

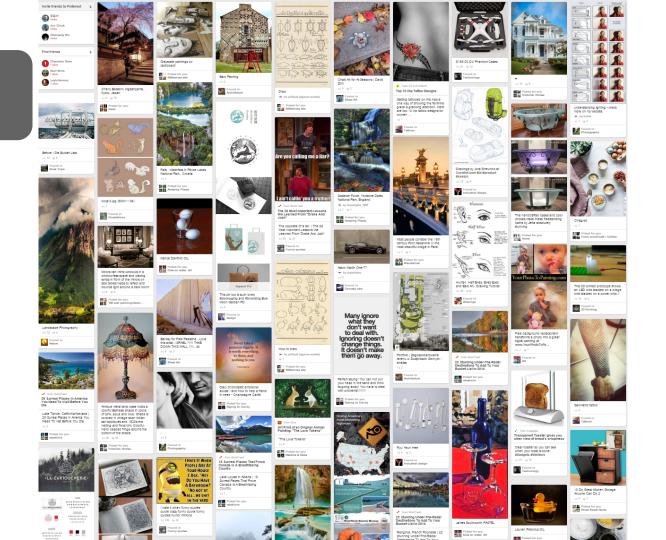


image courtesy: pinterest.com Motivation



image courtesy: viralized.com



Motivation Pros and cons of the classic



City Selation





Related Work Social Street View, *Du* and *Varshney* Web3D 2016 Best Paper Award

Related Work

-

REAL BROWN

36

3D Visual Popularity Bulbul and Dahyot, 2017



Related Work Immersive Trip Reports Brejcha et al. UIST 2018





What's Next? Research Question 1/3

What may a social media platform look like in mixed reality?

11111111111111111111

What's Next? Research Question 2/3

What if we could allow social media sharing in a live mirrored world?



What use cases can we benefit from social media platform in XR?





System Overview Geollery Workflow



2D polygons and metadata from OpenStreetMap



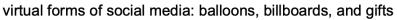
internal and external geotagged social media



shaded 3D buildings with 2D ground tiles









added avatars, clouds, trees, and day/night effects

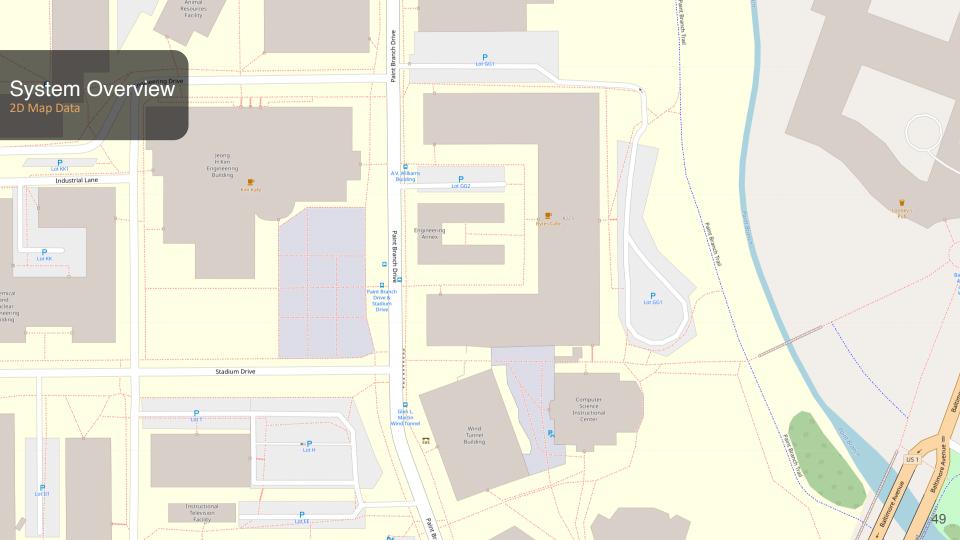


Geollery fuses the mirrored world with geotagged data, street view 360° images, and virtual avatars.













System Overview Street View Panoramas

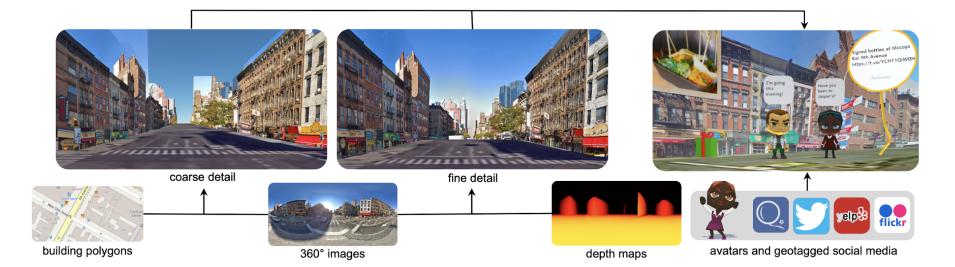






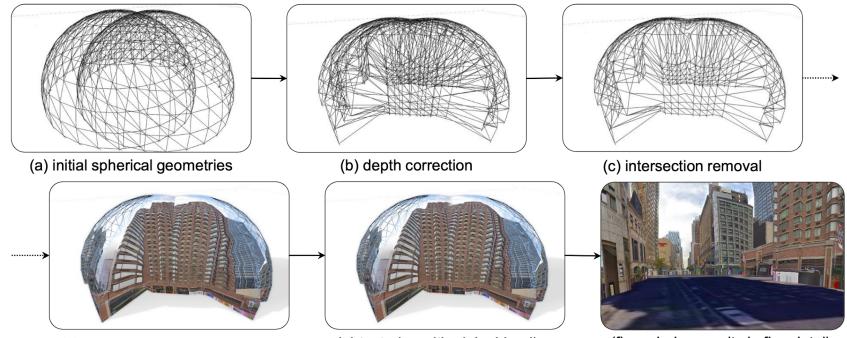


System Overview Geollery Workflow



All data we used is publicly and widely available on the Internet.



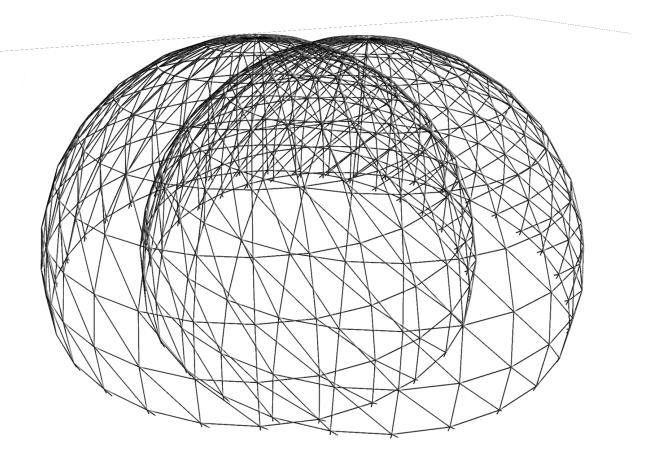


(d) texturing individual geometry

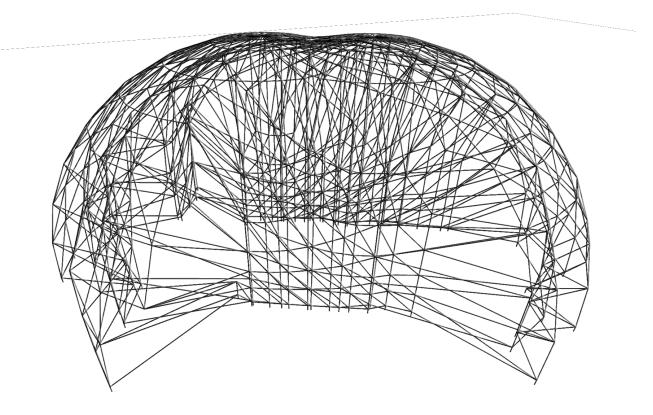
(e) texturing with alpha blending

(f) rendering results in fine detail

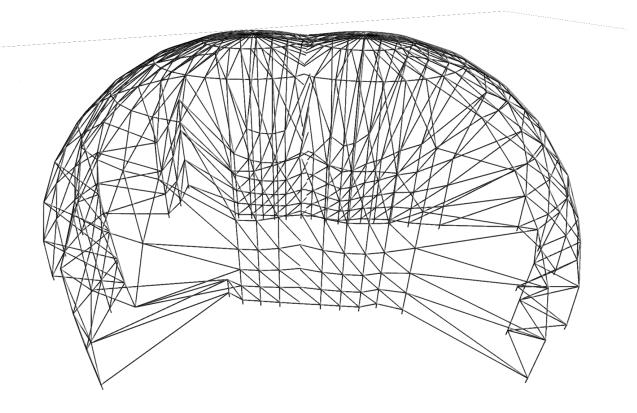
Rendering Pipeline Initial spherical geometries



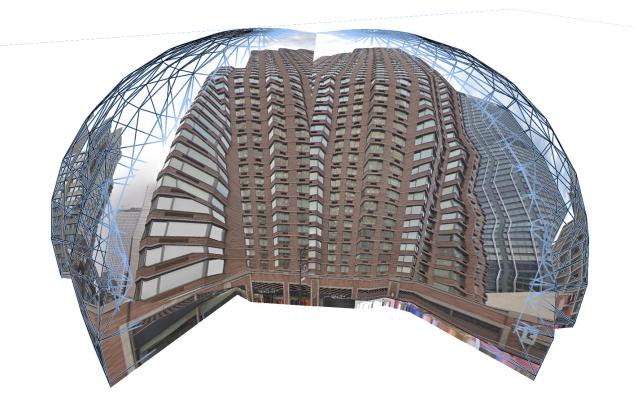
Rendering Pipeline



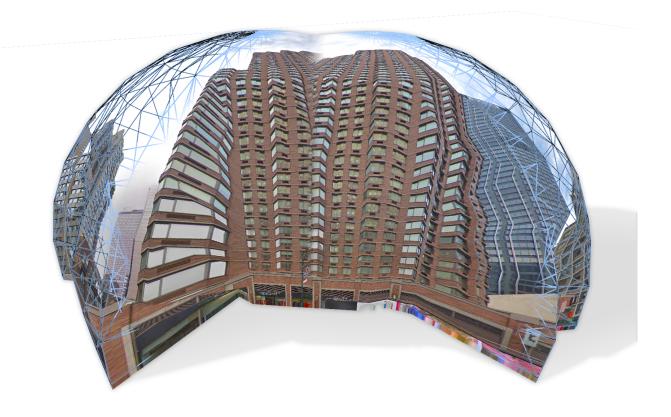
Rendering Pipeline Intersection removal



Rendering Pipeline Texturing individual geometry



Rendering Pipeline Texturing with alpha blending



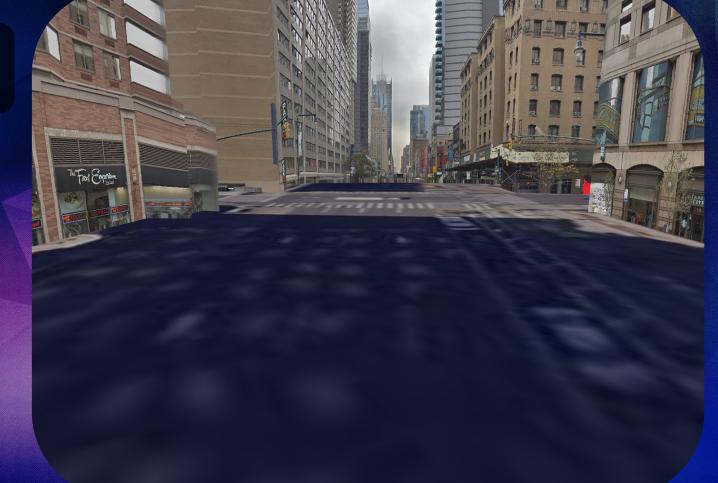
Rendering Pipeline Rendering result in the fine

detail



Rendering Pipeline Rendering result in the fine

detail



Rendering Pipeline Rendering result in the fine detail







I would like to use it for the food in different restaurants. I am always hesitating of different restaurants. It will be very easy to *see all restaurants with street views*. In Yelp, I can only see one restaurant at a time.



"

[I will use it for] exploring *new places*. If I am going on vacation somewhere, I could *immerse myself* into the location. If there are avatars around that area, I could *ask questions*.



I think it (Geollery) will be useful for families. I just taught my grandpa how to use Facetime last week and it would great if I could teleport to their house and meet with them, then we could chat and share photos with our avatars.



if there is a way to unify the interaction between them, there will be more realistic buildings [and] you could have more roof structures. Terrains will be interesting to add on.

P18/M

Rendering Pipeline Experimental Features

A V Williams Building

55

17

What wonderful five years in Maryland!

88 84

and the second

Grant









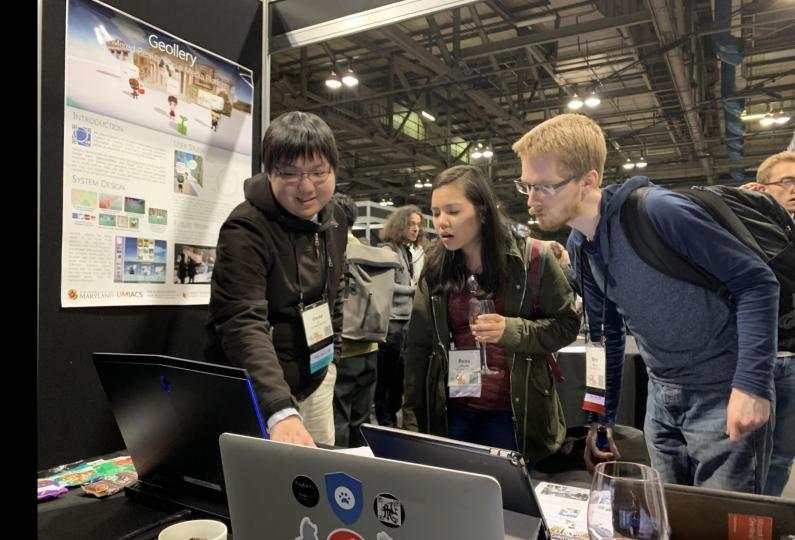
User Study

The same is sense in sends and sense is differences in the sense is th

THE WORK AND A STATE OF A STATE O



Landing Impact Demos at ACM CHI 2019



Instant Panoramic Texture Mapping with Semantic Object Matching for Large-Scale Urban Scene Reproduction

TVCG 2021, Jinwoo Park, Ik-beom Jeon, Student Members, Sung-eui Yoon, and Woontack Woo

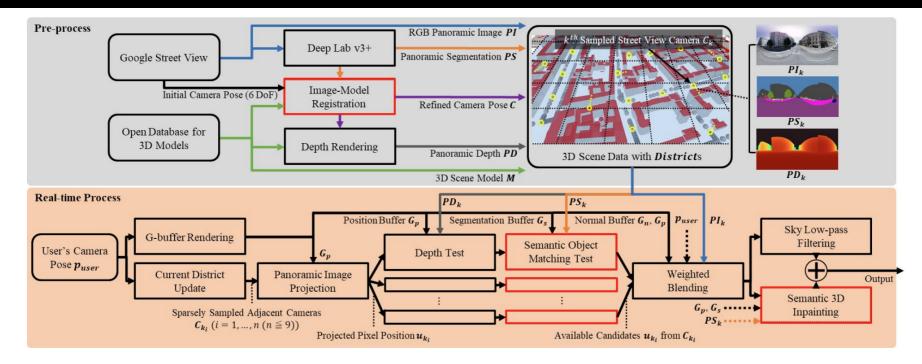


Fig. 2: Overview. In a pre-process, our system constructs *3D Scene Data*, which contains five different input resources: street-view images, 3D models, estimated panoramic-segmentation images, synthetic panoramic-depth images, and refined extrinsic camera parameters. For sparse

Instant Panoramic Texture Mapping with Semantic Object Matching for Large-Scale Urban Scene Reproduction

TVCG 2021, Jinwoo Park, Ik-beom Jeon, Student Members, Sung-eui Yoon, and Woontack Woo

A more applicable method for constructing walk-through experiences in urban streets was employed by Geollery [16], which adopted an efficient transformation of a dense spherical mesh to construct a local proxy geometry based on the depth maps from Google Street View



Park et al. Virtual Reality 2022



Yeom et al. IEEE VR 2021

Freeman *et al.* ACM PHCI 2022



He et al. ISMAR 2020

What's Next?

Video Fields: Fusing Multiple Surveillance Videos into a Dynamic Virtual Environment

Ruofei Du, Sujal Bista, Amitabh Varshney

The Augmentarium | UMIACS | University of Maryland, College Park {ruofei, varshney} @ cs.umd.edu www.Video-Fields.com



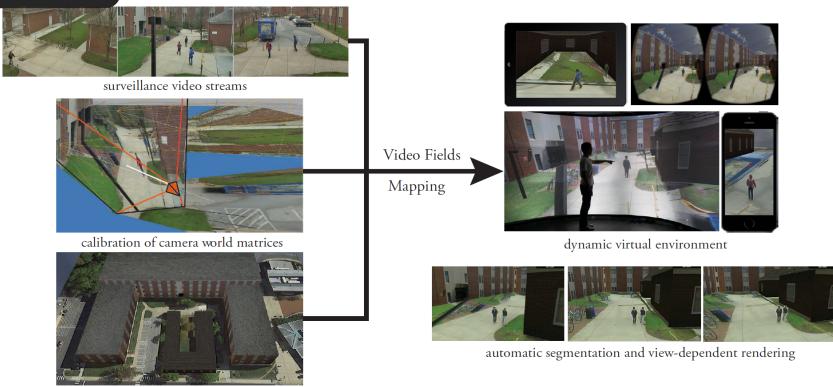
Introduction Surveillance Videos

UNIVERSITY OF MARYLAND • DEPARTMENT OF PUBLIC SAFETY



image courtesy: University of Maryland, College Park





static 3D models and satellite image



In this paper we introduce, Video Fields, a novel web-based interactive system to create, calibrate, and render ...

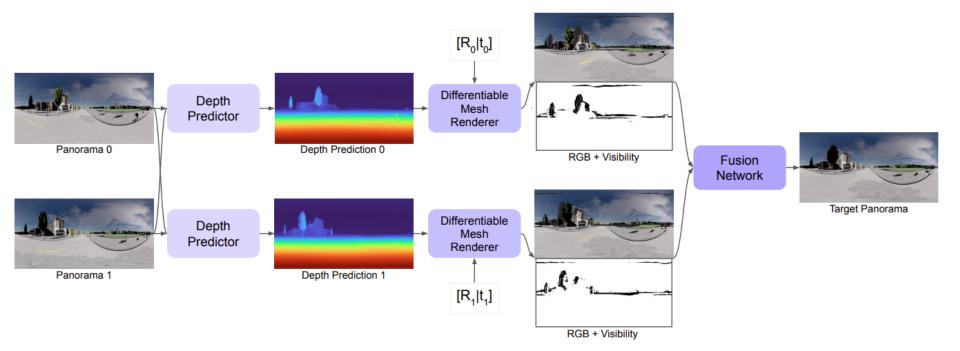
OmniSyn: Intermediate View Synthesis Between Wide-Baseline Panoramas

David Li, Yinda Zhang, Christian Häne, Danhang Tang, Amitabh Varshney, and Ruofei Du, VR 2022



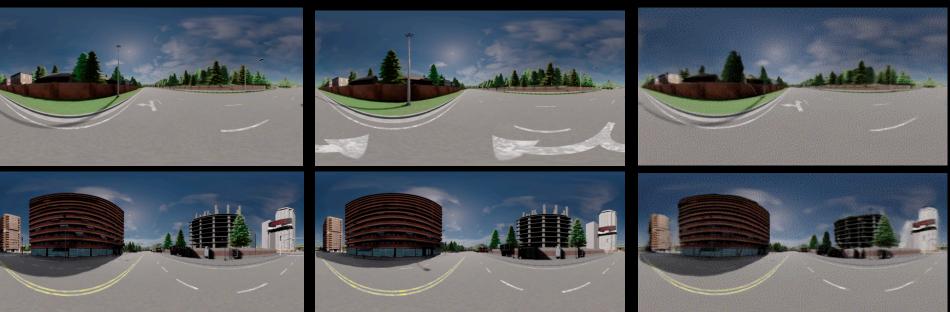
OmniSyn: Intermediate View Synthesis Between Wide-Baseline Panoramas

David Li, Yinda Zhang, Christian Häne, Danhang Tang, Amitabh Varshney, and Ruofei Du, VR 2022



input 1

input 2









Block-NeRF

Scalable Large Scene Neural View Synthesis

Matthew Tancik^{*} UC Berkeley

Ben Mildenhall Google Research Vincent Casser Waymo

Pratul Srinivasan Google Research Xinchen Yan Waymo

Jonathan T. Barron Google Research Sabeek Pradhan Waymo

Henrik Kretzschmar Waymo



Chapter Two · Computational Interaction: Algorithm & Systems



DepthLab UIST '20 17K Installs & deployed in Tiktok, Snap, Teamviewer etc.



Ad hoc UI CHI EA '22



SlurpAR DIS '22

Rapsai CHI '23 Honorable Mentions

DepthLab: Real-time 3D Interaction with Depth Maps for Mobile Augmented Reality

Ruofei Du, Eric Turner, Maksym Dzitsiuk, Luca Prasso, Ivo Duarte, Jason Dourgarian, Joao Afonso, Jose Pascoal, Josh Gladstone, Nuno Cruces, Shahram Izadi, Adarsh Kowdle, Konstantine Tsotsos, David Kim

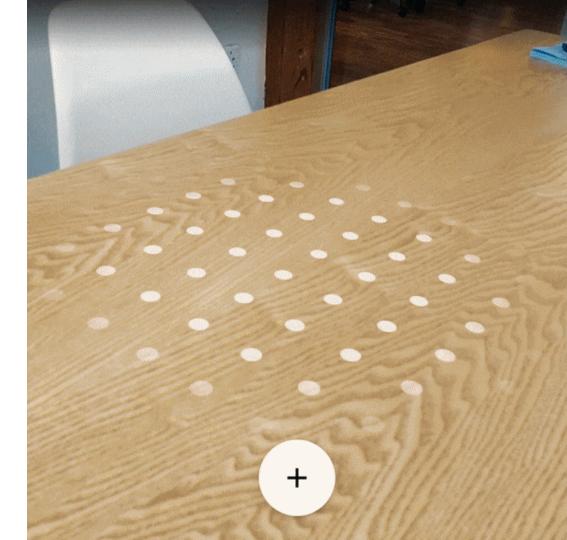
Google | ACM UIST 2020





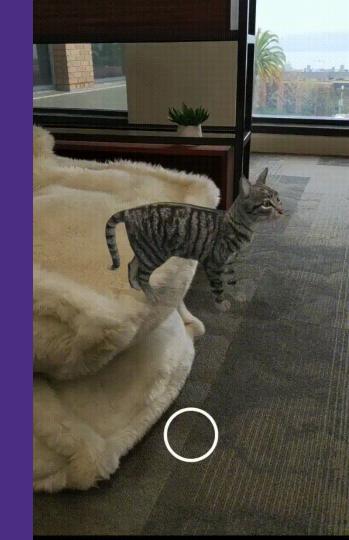




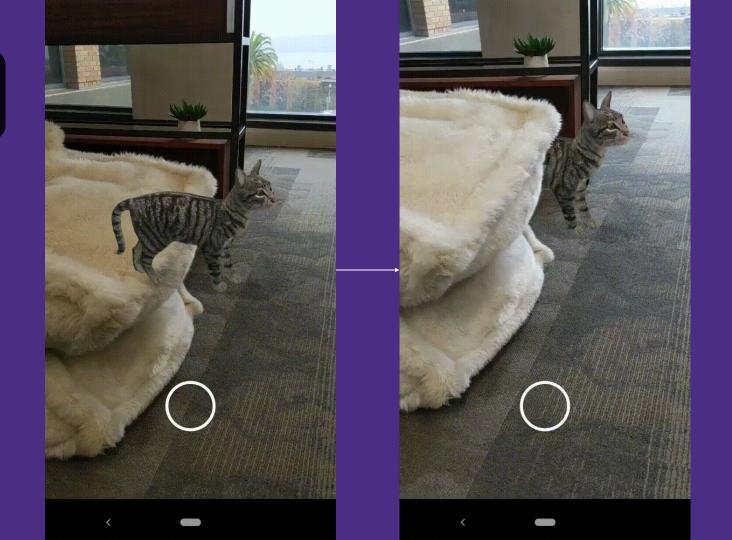


Introduction Mobile Augmented Reality Introduction Depth Lab

> Virtual content looks like it's *"pasted on the screen"* rather than *"in the world"*!



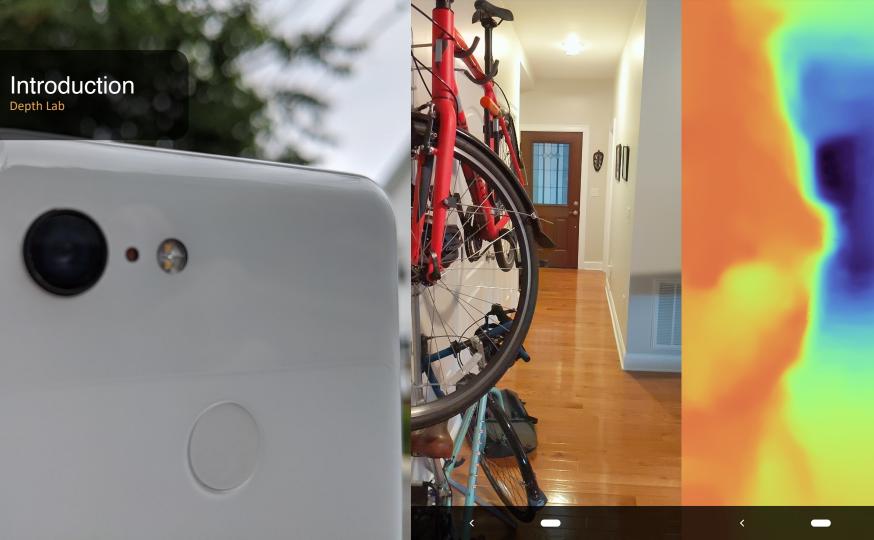
Introduction



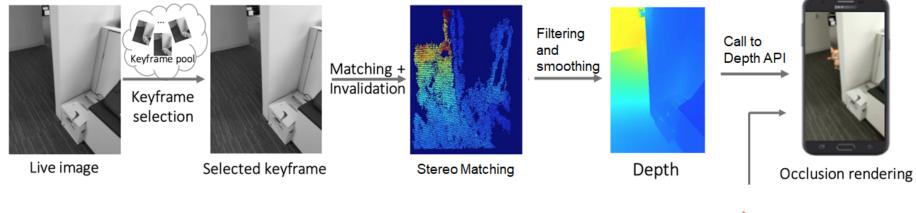
SPATIAL MAPPING

2.10

How can we bring these advanced features to mobile AR experiences WITHOUT relying on dedicated sensors or the need for computationally expensive surface reconstruction?

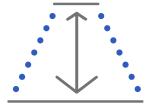








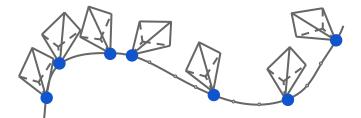
Depth Maps

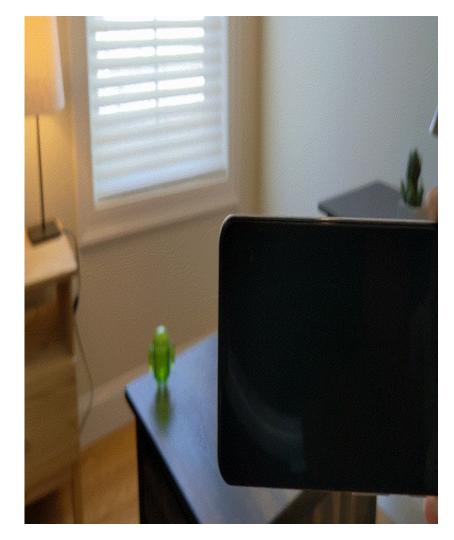




Depth From a Single Camera

Depth from Motion

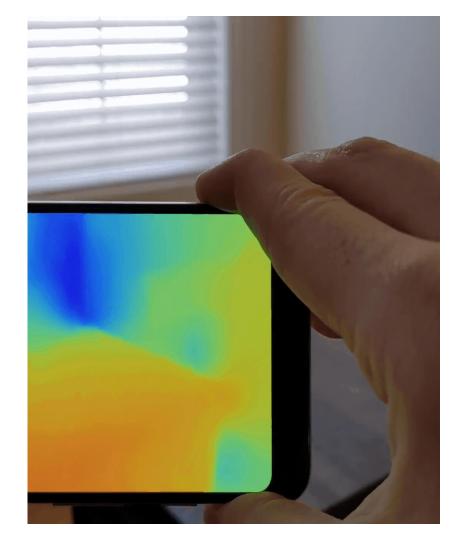




Optimized to give you the best depth

Depth from Motion is fused with stateof-the-art Machine Learning

Depth leverages specialized hardware like a Time-of-Flight sensor when available





Introduction Depth Lab



Mobile AR

developers



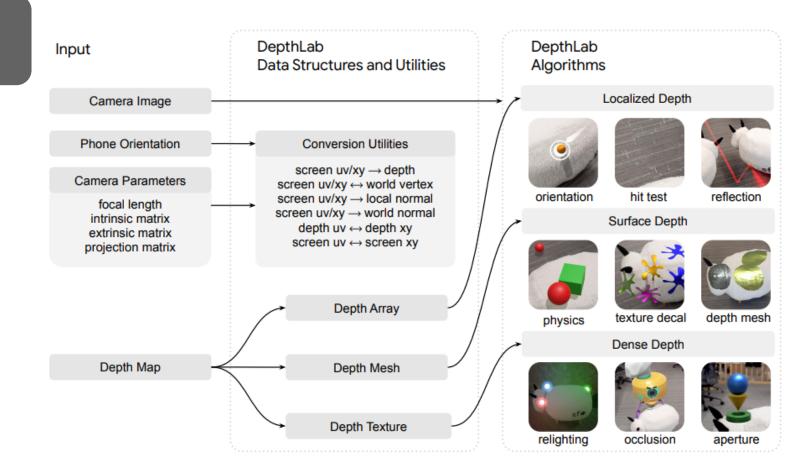
Design Process 3 Brainstorming Sessions

0

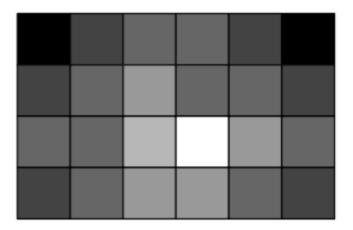
3 brainstorming sessions18 participants39 aggregated ideas





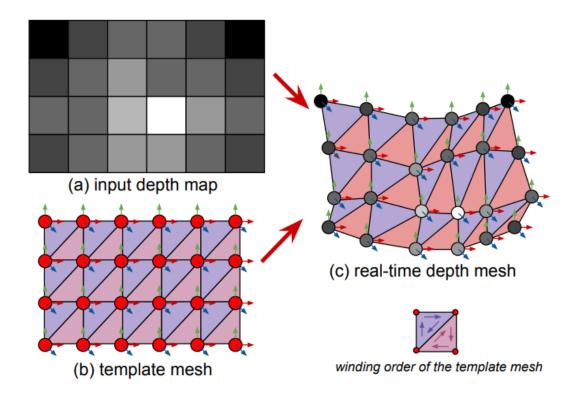


Data Structure Depth Array



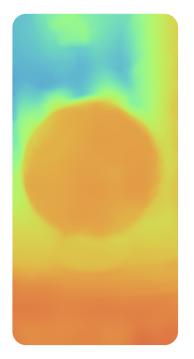
2D array (160x120 and above) of 16-bit integers

Data Structure Depth Mesh



Data Structure





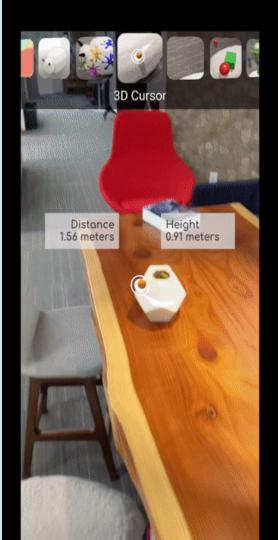


	Localized Depth	Surface Depth	Dense Depth
CPU	1	1	X (non-real-time)
GPU	N/A	✓ (compute shader)	✓ (fragment shader)
Prerequisite	point projection normal estimation	depth mesh triplanar mapping	anti-aliasing multi-pass rendering
Data Structure	depth array	depth mesh	depth texture
Example Use Cases	physical measure oriented 3D cursor path planning	collision & physics virtual shadows texture decals	scene relighting aperture effects occluded objects



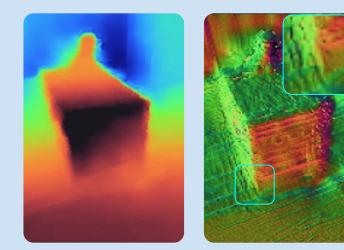
Conversion Utilities

screen uv/xy → depth screen uv/xy ↔ world vertex screen uv/xy → local normal screen uv/xy → world normal depth uv ↔ depth xy screen uv ↔ screen xy



Localized Depth Normal Estimation

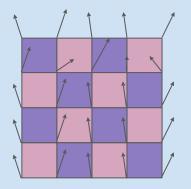
$$\mathbf{n_p} = \left(\mathbf{v_p} - \mathbf{v_{p+(1,0)}}\right) \times \left(\mathbf{v_p} - \mathbf{v_{p+(0,1)}}\right)$$



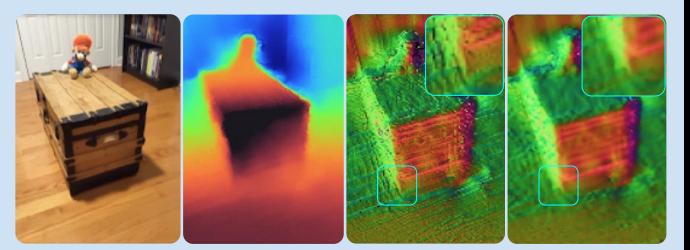
Localized Depth

Point in DepthLab. **Input** : A screen point $\mathbf{p} \leftarrow (x, y)$ and focal length f. Output : The estimated normal vector **n**. 1 Set the sample radius: $r \leftarrow 2$ pixels. 2 Initialize the counts along two axes: $c_X \leftarrow 0, c_Y \leftarrow 0$. 3 Initialize the correlation along two axes: $\rho_X \leftarrow 0, \rho_Y \leftarrow 0$. 4 for $\Delta x \in [-r,r]$ do for $\Delta y \in [-r, r]$ do 5 **Continue if** $\Delta x = 0$ and $\Delta y = 0$. 6 Set neighbor's coordinates: $\mathbf{q} \leftarrow [x + \Delta x, y + \Delta y]$. 7 Set **q**'s distance in depth: $d_{\mathbf{pq}} \leftarrow \|\mathbf{D}(\mathbf{p}), \mathbf{D}(\mathbf{q})\|$. 8 Continue if $d_{pq} = 0$. 9 if $\Delta x \neq 0$ then 10 $c_X \leftarrow c_X + 1.$ 11 $\rho_X \leftarrow \rho_X + d_{pq} / \Delta x.$ 12 13 end if $\Delta y \neq 0$ then 14 $c_Y \leftarrow c_Y + 1$. 15 $\rho_Y \leftarrow \rho_Y + d_{\mathbf{pq}}/\Delta y.$ 16 end 17 end 18 19 end 20 Set pixel size: $\lambda \leftarrow \frac{\mathbf{D}(\mathbf{p})}{f}$. **21 return** the normal vector **n**: $\left(-\frac{\rho_Y}{\lambda c_Y}, -\frac{\rho_X}{\lambda c_X}, -1\right)$.

Algorithm 1: Estimation of the Normal Vector of a Screen



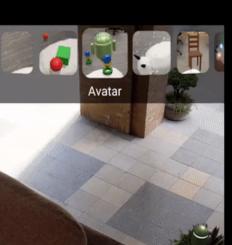
Localized Depth Normal Estimation





Localized Depth Avatar Path Planning

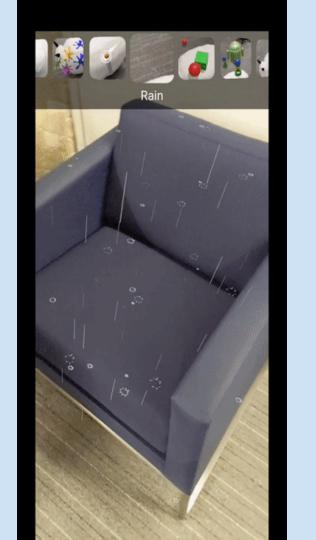






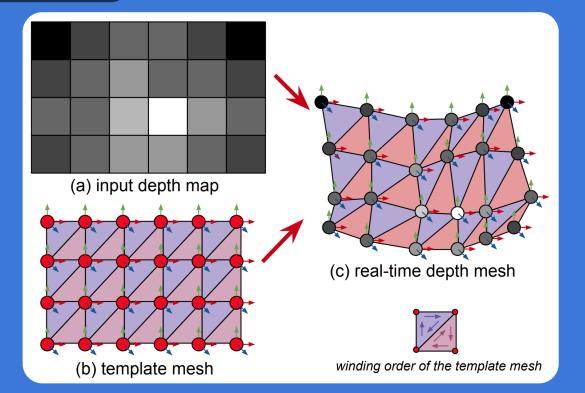
0

Localized Depth Rain and Snow





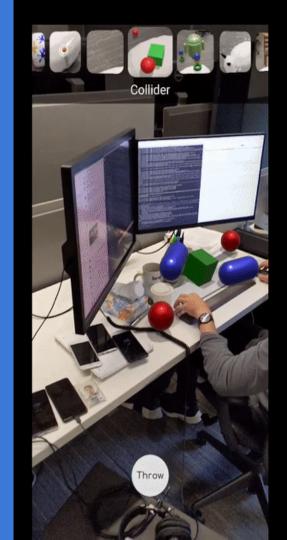








Physics with depth mesh.





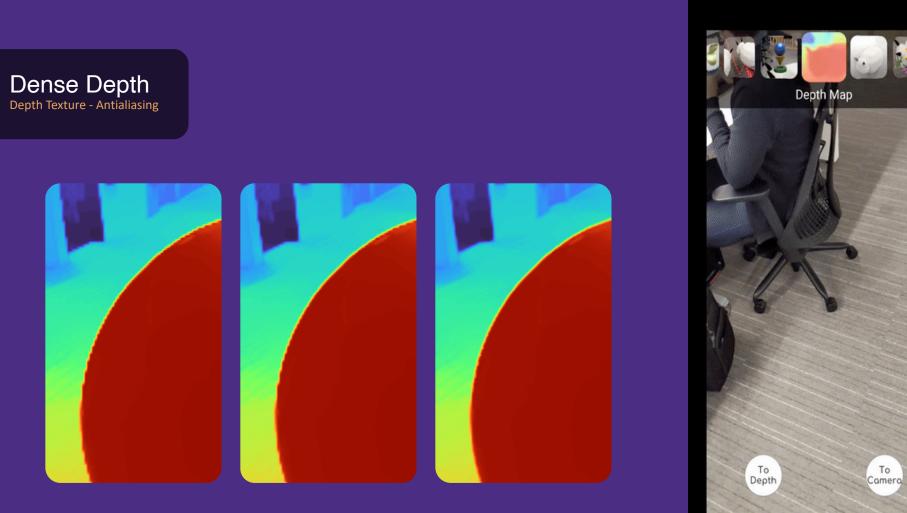
Texture decals with depth mesh.



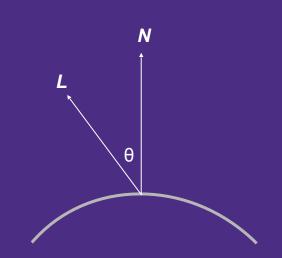


Projection mapping with depth mesh.





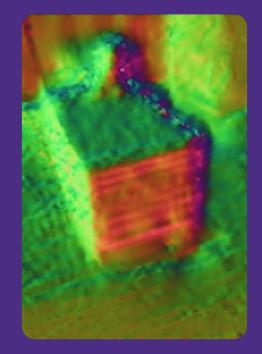






Dense Depth Why normal map does not

work?





Relighting



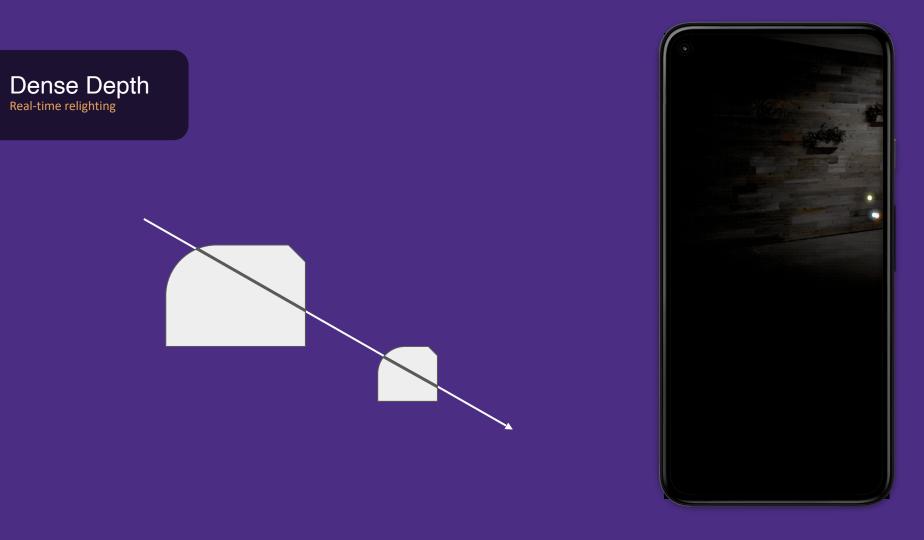


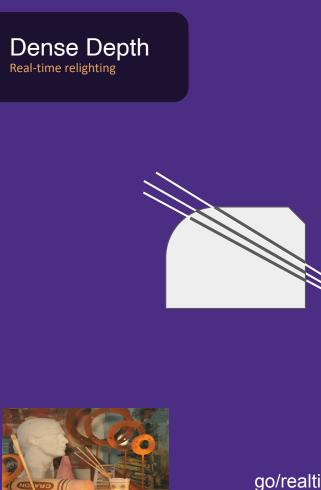
Dense Depth Real-time relighting

Algorithm 3: Ray-marching-based Real-time Relighting.	
Input : Depth map D , the camera image I , camera intrinsic	
matrix K , <i>L</i> light sources $\mathbb{L} = \{\mathscr{L}^i, i \in L\}$ with each	
light's location $\mathbf{v}_{\mathscr{L}}$ and intensity in RGB channels	
$\phi_{\mathscr{L}}.$	
Output : Relighted image O.	
1 for each image pixel $\mathbf{p} \in$ depth map \mathbf{D} in parallel do	
2 Sample p 's depth value $d \leftarrow \mathbf{D}(\mathbf{p})$.	
3 Compute the corresponding 3D vertex $\mathbf{v_p}$ of the screen	
point p using the camera intrinsic matrix $\mathbf{v}_{\mathbf{p}}$ with K :	
$\mathbf{v}_{\mathbf{p}} = \mathbf{D}(\mathbf{p}) \cdot \mathbf{K}^{-1}[\mathbf{p}, 1]$	
4 Initialize relighting coefficients of $\mathbf{v_p}$ in RGB: $\phi_p \leftarrow 0$.	
5 for each light $\mathscr{L} \in light$ sources \mathbb{L} do	
6 Set the current photon coordinates $\mathbf{v}_o \leftarrow \mathbf{v_p}$.	
7 Set the current photon energy $E_o \leftarrow 1$.	
8 while $\mathbf{v}_o \neq \mathbf{v}_{\mathscr{L}}$ do	
9 Compute the weighted distance between the	
photon to the physical environment	
$\Delta d \leftarrow \alpha \mathbf{v}_o^{xy} - \mathbf{v}_{\mathscr{L}}^{xy} + (1 - \alpha) \mathbf{v}_o^z - \mathbf{v}_{\mathscr{L}}^z , \alpha = 0.5.$	
10 Decay the photon energy: $E_o \leftarrow 95\% E_o$	
11 Accumulate the relighting coefficients:	
$\phi_{\mathbf{p}} \leftarrow \phi_{\mathbf{p}} + \Delta dE_o \phi_{\mathscr{L}}.$	
12 March the photon towards the light source:	
$\mathbf{v}_o \leftarrow \mathbf{v}_o + (\mathbf{v}_{\mathscr{L}} - \mathbf{v}_o)/S$, here $S = 10$, depending	
on the mobile computing budget.	
13 end	
14 end	
15 Sample pixel's original color: $\Phi_{\mathbf{p}} \leftarrow \mathbf{I}(\mathbf{p})$.	
16 Apply relighting effect:	
$\mathbf{O}(\mathbf{p}) \leftarrow \gamma \cdot 0.5 - \phi_{\mathbf{p}} \cdot \Phi_{\mathbf{p}}^{1.5 - \phi_{\mathbf{p}}} - \Phi_{\mathbf{p}}, \text{ here } \gamma \leftarrow 3.$	
17 end	









go/realtime-relighting, go/relit

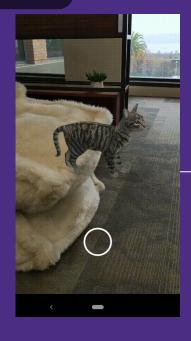
















FOG



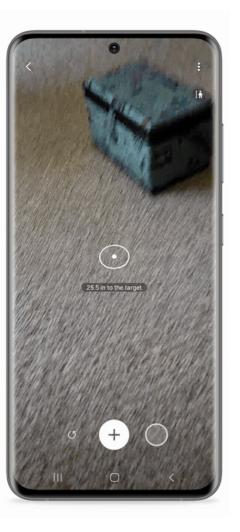
Impact Deployment with partners







Impact Deployment with partners



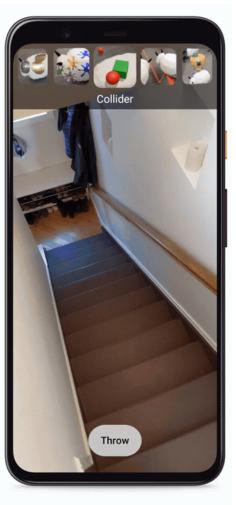




Impact Deployment with partners









AR Realism

In TikTok





AR Realism

Built into Lens Studio for Snapchat Lenses



Snap **Dancing Hotdog**

Kevaid Saving Chelon



Quixotical The Seed: World of Anthrotopia

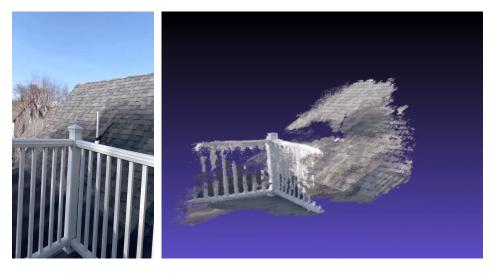
New depth capabilities

Raw Depth API

Provides a more **detailed**

representation of the geometry

of the objects in the scene.



Camera Image

3D Point Cloud



ARCore Depth Lab App



Depth API Codelab



Raw Depth API Codelab

GitHub



Googlesamples / arcore-o	depth-lab		⊙ Unwatch 👻	- 28 🛉 Unstar 333 💱 Fork 66
<> Code (1) Issues 3 (1)	Pull requests 🕟 Actions 🔟 Projects 🖽 Wiki 🕕	Security 🖂 In	isights 🔞 Set	ttings
양 master 👻 양 1 branch 📀 0	O tags Go to file	Add file -	⊻ Code -	About 爺
😨 ruofeidu Updated README.md	with latest UIST 2020 publication.	c111eda on Jul 31	🕄 6 commits	ARCore Depth Lab is a set of Depth API samples that provides assets using depth for advanced geometry-aware
Assets	Added a demo scene of stereo photo mode.		3 months ago	features in AR interaction and
ProjectSettings	Added a demo scene of stereo photo mode.		3 months ago	rendering. (UIST 2020)
CONTRIBUTING.md	Initial commit.		3 months ago	arcore arcore-unity depth mobile
	Initial commit.		3 months ago	ar interaction
README.md	Updated README.md with latest UIST 2020 publication.		2 months ago	🛱 Readme
				বাঁহ্ৰ View license
README.md			R	

README.md

ARCore Depth Lab - Depth API Samples for Unity

Copyright 2020 Google LLC. All rights reserved.

Depth Lab is a set of ARCore Depth API samples that provides assets using depth for advanced geometry-aware features in AR interaction and rendering. Some of these features have been used in this Depth API overview video.

ARCore Depth API is enabled on a subset of ARCore-certified Android devices. iOS devices (iPhone, iPad) are not supported. Find the list of devices with Depth API support (marked with Supports Depth API) here: https://developers.google.com/ar/discover/supported-devices. See the ARCore developer documentation for more information.

Download the pre-built ARCore Depth Lab app on Google Play Store today.



Sample features

The sample scenes demonstrate three different ways to access depth:

1. Localized depth: Sample single depth values at certain texture coordinates (CPU).

Character locomotion on uneven terrain

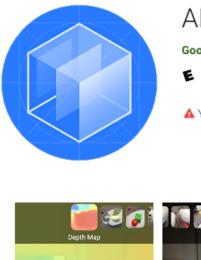
- · Collision checking for AR object placement
- · Laser beam reflections
- Oriented 3D reticles

	eleases published e a new release
Pack	kages
	ackages published sh your first package
	tributors 2 kidavid David Kim ruofeidu Ruofei Du
Lang	guages

Play Store Try it yourself!







ARCore Depth Lab

Google Samples Tools

***** 40 🚊

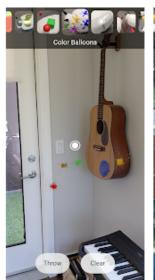
E Everyone

A You don't have any devices.

Installed











KEY QUOTES

"The result is a more believable scene, because the depth detection going on under the hood means your smartphone better understands every object in a scene and how far apart each object is from one another. Google says it's able to do this through optimizing existing software, so you won't need a phone with a specific sensor or type of processor. It's also all happening on the device itself, and not relying on any help from the cloud." - The Verge

"Occlusion is arguably as important to AR as positional tracking is to VR. Without it, the AR view will often "break the illusion" through depth conflicts." - UploadVR

"Alone, that feature (creating a depth map with one camera) would be impressive, but Google's intended use of the API is even better: occlusion, a trick by which digital objects can appear to be overlapped by real-world objects, blending the augmented and real worlds more seamlessly than with mere AR overlays." - VentureBeat

"Along with the Environmental HDR feature that blends natural light into AR scenes, ARCore now rivals ARKit with its own exclusive feature. While ARKit 3 offers People Occlusion and Body Tracking on compatible iPhones, the Depth API gives ARCore apps a level of environmental understanding that ARKit can't touch as of yet." - Next Reality

"More sophisticated implementations make use of multiple cameras...That's what makes this new Depth API almost magical. With just one camera, ARCore is able to create 3D depth maps ... in real-time as you move your phone around." -Slash Gear



COVERAGE LINKS

- A New Wave of AR Realism with the ARCore Depth API. Google Developers. June 25, 2020.
- Google Makes Its AR-Centric Depth API Available to All Developers. Engadget. June 25, 2020.
- AR Realism with the ARCore Depth API (Video). Google Developers. June 25, 2020.
- Introducing the ARCore Depth API for Android and Unity. Google AR & VR. June 25, 2020.
- ARCore's new Depth API is out of beta, bringing the next generation of hide-and-seek to phones. Android Police. June 25, 2020.
- Google is improving its augmented reality tool so virtual cats can hide behind your sofa. ZDNet. December 10, 2019.
- ARCore's Depth API helps create depth maps using a single camera. XDA Developers. December 10, 2019.
- Google's New Phone AR Update Can Hide Virtual Things in the Real World. CNET. December 9, 2019.
- Google Shows off Stunning New AR Features Coming to Web and Mobile Apps Soon. The Verge. December 9, 2019.
- Google's ARCore Depth API Enables AR Depth Maps and Occlusion with One Camera. VentureBeat. December 9, 2019.
- Google's ARCore Is Getting Full Occlusion For More Real AR. UploadVR. December 9, 2019.
- Google ARCore Depth API Now Available, Letting Devs Make AR More Realistic. RoadToVR. December 9, 2019.
- ARCore Depth API Takes Android AR Experiences To A Whole New Level. VRScout. December 9, 2019.
- Google Update Adds Real-World Occlusion to ARCore with Depth API. Next Reality. December 9, 2019.
- ARCore phones can now detect depth with a single camera. 9To5Google. December 9, 2019.
- ARCore Depth API: How it will fundamentally transform your AR experiences. Android Authority. December 9, 2019.
- ARCore Depth API lets you hide cats behind sofas even with one camera. SlashGear. December 9, 2019.
- Google's Latest ARCore API Needs Just One Camera For Depth Detection. HotHardware. December 9, 2019.
- Get Ready for the ARCore Depth API (Video). Google AR & VR. December 9, 2019.
- Blending Realities with the ARCore Depth API. Google Developers. December 9, 2019.

Limitations

Dynamic Depth? HoloDesk, HyperDepth, Digits, Holoportation for mobile AR?





After exploring interaction with the physical world, how shall we interaction in everyday object?

Ad hoc UI: On-the-fly Transformation of Everyday Objects into Tangible 6DOF Interfaces for AR

Ruofei Du, Alex Olwal, Mathieu Le Goc, Shengzhi Wu, Danhang Tang, Yinda Zhang, Jun Zhang, David Joseph Tan, Federico Tombari, David Kim

75°

Mostly cloudy

65°

Google | CHI 2022 Interactivity



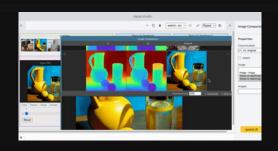


With recent advances of multi-modal machine learning models, how can we accelerate the prototyping efforts?

What if we can build applications as if building Legos?



Accelerating Machine Learning Prototyping of Multimedia Applications through Visual Programming

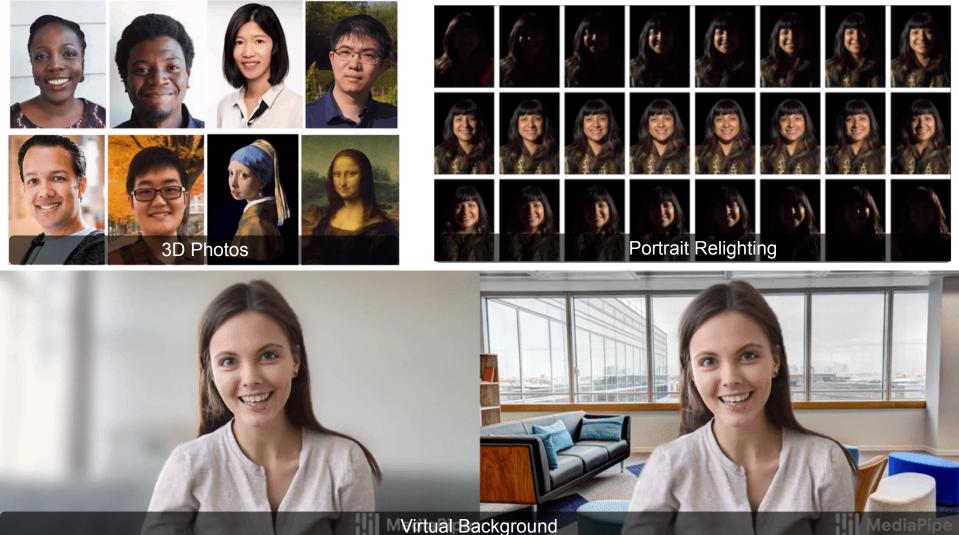


<u>**Ruofei Du**</u>, Na Li, Jing Jin, Michelle Carney, Scott Miles, Maria Kleiner, Xiuxiu Yuan, Yinda Zhang, Anuva Kulkarni, Xingyu "Bruce" Liu, Ahmed Sabie, Sergio Escolano, Abhishek Kar, Ping Yu, Ram Iyengar, Adarsh Kowdle, and Alex Olwal

visualblocksforml.github.io







Virtual Background





Input Foreground



Surface Normals

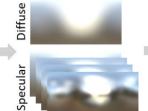
Convolved Light Maps



Albedo



Input HDR Map





Diffuse Light Map



Specular Light Maps



Relit Foreground

Pandey, Rohit, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. "Total relighting: learning to relight portraits for background replacement." ACM Transactions on Graphics (SIGGRAPH 2021) 40, no. 4 (2021): 1-21.

CB-14

101212-2010 H SEAM NULL PLANE

ŧ

N HER CXE NEED

はなな事業に A 16 16 19

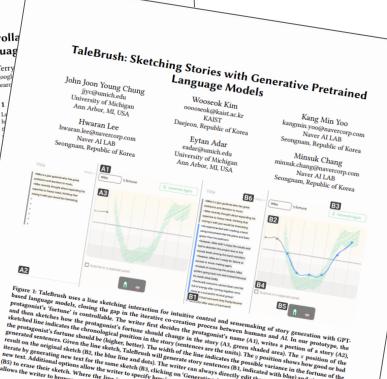
A LANSAGE ž

11

K. S. K **12** 35 15 17

とよま 7





sketched line indicates the chronological position in the story (sentences are the units). The y position shows how good or bad the protagonist's fortune should be (higher, better). The width of the line indicates the possible variance in the fortune of the sensested contences. Given the line sketch. TaleBreach will senseste story sentences (R1 indicated with hige) and visualized to the the protagonist's fortune should be (higher, better). The width of the line indicates the possible variance in the fortune of the generated sentences, (iven the line sketch, TaleBrush will generate story sentences (B1, indicated with blue) and visualize the society of the science of of the sci generated sentences. Given the line sketch, TaleBrush will generate story sentences (B.1, indicated with blue) and visualize the result on the original sketch (B2; the blue line and dots). The writer can always directly edit the generated text. They can also a sentence of the same chetch /B3 elicibies on 'Conversing Again's or revising their sketch and generating and the same chetch /B3 elicibies on 'Conversing Again's or revising their sketch and generating and the same chetch /B3 elicibies on 'Conversing Again's or revising their sketch and generating and the same chetch /B3 elicibies on 'Conversing Again's or revising their sketch and generating again and the same chetch /B3 elicibies on 'Conversing Again's or revising their sketch and generating again and the same chetch /B3 elicibies on 'Conversing Again's or revising their sketch and generating again and the same chetch /B3 elicibies on 'Conversing Again's or revising the same chetch and generating again and the same chetch /B3 elicibies on 'Conversing Again's or revising the same chetch and generating again and the same chetch /B3 elicibies on 'Conversing Again's or revising the same chetch and generating again and the same chetch /B3 elicibies on 'Conversing Again's or revising the same chetch and generating again and the same chetch /B3 elicibies on 'Conversing Again's or revising the same chetch and generating again result on the original sketch (h2, the blue line and dots). The writer can always directly edit the generated text. They can also iterate by generating new text for the same sketch (B3, dicking on 'Generating Again') or revising their sketch and generating new text a distributed ontions allow the writer to snecify how 'surprising' the generation should be (B4) or using the eraset to a single of the generation of the same text of the same text of the same text additional ontions allow the writer to snecify how 'surprising' the generation should be (B4) or using the eraset to a single of the generation of the same text of the same text of the same text additional ontions allow the writer to snecify how 'surprising' the generation should be (B4) or using the eraset to a single of the same text of tex iterate by generating new text for the same sketch (B3, clicking on 'Generating Again') or revising their sketch and generating new text. Additional options allow the writer to specify how 'surprising' the generation should be (B4) or using the eraser tool (R5) to erase their sketch. Where the line is erased. TalePrinch generates unconstrained sentences. A history drondward (6) new text. Additional options allow the writer to specify how 'surprising' the generation should be (B4) or using the eraser tool (B5) to erase their sketch. Where the line is erased, TaleBrush generates unconstrained sentences. A history dropdown (B6) endows the writer to browse previously generated sentences.

Permission to make digital or hard copies of all or part of this work for personal or datarroom use it granted without for provided that copies are not made or distributed for another experimental experimental and any environmental strategies are not stated by a first data and the data strategies are strategies and any environmental strategies are not stated by a first data and the data strategies are any strategies and any environmental strategies are not strategies and any strategies are data strategies and any strategies are any strategies and any strategies are data and any strategies and any strategies are any strategies are data any strategies and any strategies are data any strategies and any strategies are data any strategies and data any strategies are data any strategies and data any strategies are data any strategies and data any strategies any strategies and data any strategies are data any strategies and data any strategies and data any strategies any strategies and data any strategies and data any strategies and data any strategies any strategies and data any strategies and data any strategies any strategies any strategies any strategies and data any strategies any strategies and data any strategies any strategies any strategies any strategies and data any strategies any strategies any strategies any strategies any strategies any strategies and data any strategies any

classroom use is practicel without for provided fract copies are not masse or distributed for prodite commercial advantage and had copies hear this notice and the full classics for prodite commercial advantage and had copies hear this notice and the full classics the state of the state o

Ctrr 22, April 29-May 5, 2022, New Orleans, LA, USA e 2022 (Appropring) the dd by the ownerstandbar(s). Publication rights licensed to ACM. ACM SIRN 97: 4269-455-32200, ...515.00 https://doi.org/10.1165/34911023501619

for predict arcommercial advantages and that copies hear this motion and the full cristian on the far provide the copyrights for composition of this work overally by others than the methods (b) much be copyrights for composition of the premated. To all you obtain that the output of the copyright of the relative term of the premated for the premated and/or to the copyright of the relative term of the premated for the premised of the previous premative term of the premised of the premised of the copyright of the premised of the premised of the premised of the pre-cision of the premised of the premised of the premised of the pre-cision of the premised of the premised of the premised of the pre-cision of the premised of the premised of the pre-cision of the pre-pre-pre-tation of the pre-pre-tation of the pre-tation of the pre-pre-tation of the pre-pre-tation of the pre-tation of th While advanced text generation algorithms (e.g., GPT-3) have enwhile advanced rest generation agostions to the second state of th tive remains a challenge. Existing systems often leverage simple two remains a charge strange systems strenge everyone inter-turn-taking between the writer and the AI in story development. However, writers remain unsupported in intuitively understanding the AT's actions or steering the iterative generation. We introduce TaleBrush, a generative story ideation tool that uses line sketching interactions with a GPT-based language model for control sensemaking of a protagonist's fortune in

AI Chains: Transparent and Controlla by Chaining Large Languag

Michael Terry

Google Researc

USA

1

michaelterry@goog

Tongshuang Wu* wtshuang@cs.washington.edu University of Washington USA

ABSTRACT

Although large language models (LLMs) have demonstrated impressive potential on simple tasks, their breadth of scope, lack of transparency, and insufficient controllability can make them less effective when assisting humans on more complex tasks. In response, we introduce the concept of Chaining LLM steps together, where the output of one step becomes the input for the next, thus aggregating the gains per step. We first define a set of LLM primitive operations useful for Chain construction, then present an interactive system where users can modify these Chains, along with their intermediate results, in a modular way. In a 20-person user study, we found that Chaining not only improved the quality of task outcomes, but also significantly enhanced system transparency, controllability, and sense of collaboration. Additionally, we saw that users developed new ways of interacting with LLMs through Chains: they leveraged sub-tasks to calibrate model expectations, compared and contrasted alternative strategies by observing parallel downstream effects, and debugged unexpected model outputs by "unit-testing sub-components of a Chain. In two case studies, we further explo how LLM Chains may be used in future applications.

CCS CONCEPTS

 Human-centered computing → Empirical studies in Interactive systems and tools; Computing methodologi Machine learning.

KEYWORDS

Human-AI Interaction, Large Language Models, Natural L Processing

ACM Reference Format:

Tongshuang Wu, Michael Terry, and Carrie J. Cai. 2022. AI Ch parent and Controllable Human-AI Interaction by Chaining Lar Model Prompts. In CHI Conference on Human Factors in Comp (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA. ACM, 1 USA, 22 pages. https://doi.org/10.1145/3491102.3517582

"The work was done when the author was an intern at Google In



This work is licensed under a Creative Commons Attribution 4.0 License.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA © 2022 Copyright held by the owner/author(s). ACM ISBN 978-1-4503-9157-3/22/04. https://doi.org/10.1145/3491102.3517582

PromptChainer: Chaining Large Langu through Visual Program Ellen Jiang[†]

ellenj@google.com Google Research Alejandra Molina alemolinata@google.com

USA

Google Research USA

USA

Google Research

USA

Carrie J. Cai

cjcai@google.com

ACM Refere Tongshuang Michael Ter guage Mode Human Fact

that cannot be easily handled via a single run of an LLM. Recent unar cannot oc easily nananess via a single run or an LLM. Revent work has found that chaining multiple LLM runs together (with the work was course that comming manque elsest sums together (with the output of one step being the input to the next) can help users accomplish these more complex tasks, and in a way that is perceived to be

puss mese more compact tasks, and in a way that is perceived to be more transparent and controllable. However, it remains unknown what users need when authoring their own LLM chains - a key step to lowering the barriers for non-Al-experts to prototype Al-infused Large lan to sovering the barriers for non-rol-experts to prototype Arranusco applications. In this work, we explore the LLM chain authoring process. We find from pilot studies that users need support transprocess, we and from past attailer that users new support that forming data between steps of a chain, as well as debugging the chain at multiple granularities. To address these needs, we designed PrompiChainer, an interactive interface for visually programming cromps_namer, an interactive interface tor visually programming chains. Through case studies with four designers and developers, we show that PromptChainer supports building prototypes for a range of applications, and conclude with open questions on scaling tonge on approximation, and conclose with open questions on scaning chains to even more complex tasks, as well as supporting low-fi

chain prototyping. CCS CONCEPTS

chine learning. The work was done when the author was an intern at Google Inc.

[†]Equal contribution.

(c) (i)

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA © 2022 Copyright held by the owner/author(s). ACM ISBN 978-1-4503-9156-6/22/04.

Related Work

Visual programming

TAILOR: Generating

◊Paul G. Allen School of G

{wts

Abstract

Controlled text perturbation is

uating and improving model

However, current techniques

a model for every target pertur

expensive and hard to general

TAILOR, a semantically-control

ation system. TAILOR builds

seq2seq model and produces

conditioned on control codes

mantic representations. We

erations to modify the cont

in turn steer generation toy

tributes. These operations ca

posed into higher-level ones

ible perturbation strategies.

the effectiveness of these pe

tiple applications. First,

automatically create high-q

for four distinct natural la

(NLP) tasks. These contrast

spurious artifacts and are

manually annotated ones

versity. Second, we sho

turbations can improve m

data augmentatio

Alexis Ross*† Tongshuang

in language

Tongshuang Wu*† wtshuang@cs.washington.edu University of Washington

> Jeff Gray jeffgray@google.com Google Research

While LLMs have made it possible to rapidly prototype new ML w nue LLANS nave muse n possible to rapinary prototype new na-functionalities, many real-world applications involve complex tasks

1 INT

Abstracts),

NY, USA,

ML ft

tion a

prom

tation

prod

tom

the

and

appl

for

(wh

au

find

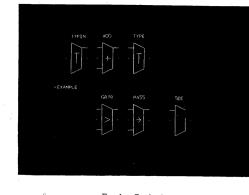
for proto data, mo mation at run t simply For exat LLM to the pro Count the na examp

 Human-centered computing → Empirical studies in HCI; • runnan-centered computing, → Empirical studies in rich; Interactive systems and ools; • Computing methodologies → Ma-

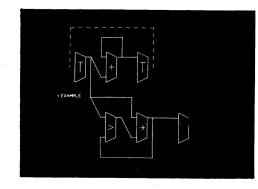
ed under a Creative Commons Attribution International

https://doi.org/10.1145/3491101.3519729

Related Work Visual programming in graphics



Basic Symbols Figure 1.1



Connected Program Figure 1.2



THE ON-LINE GRAPHICAL SPECIFICATION OF COMPUTER FROCEDURES

by

WILLIAM ROBERT SUTHERLAND

B.E.E., Rensselaer Polytechnic Institute

(1957)

M.S., Massachusetts Institute of Technology

(1963)

SUEMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF FHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

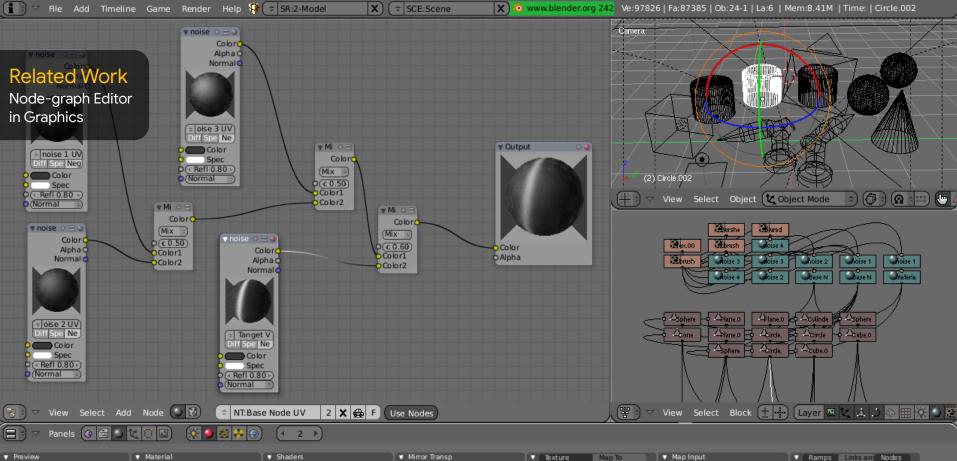
January, 1966

Signature of Author _____ Department of Electrical Engineering, January 10, 1966 Certified by ______ Thesis Supervisor

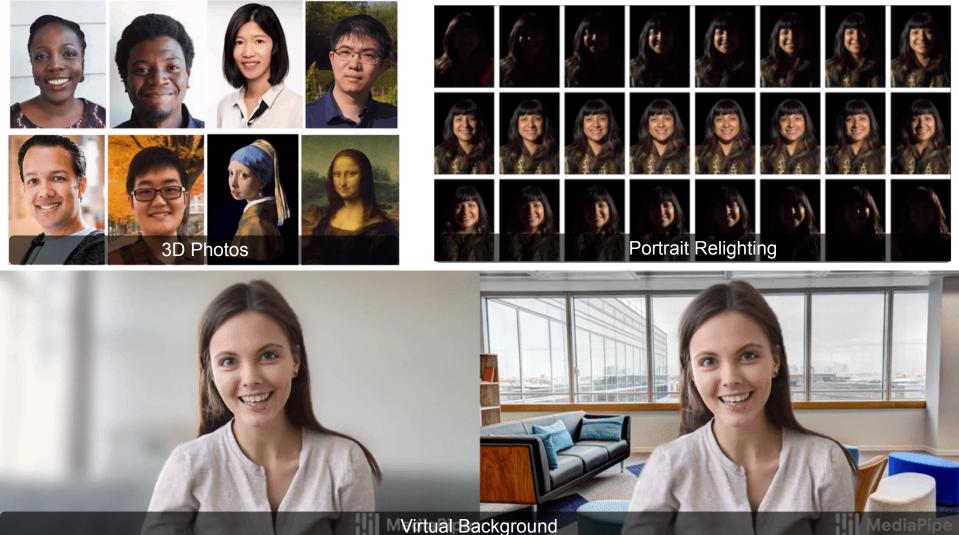
1

WR Sutherland. 1966. On-Line Graphical Specification of Procedures. SJCC, Boston, Mass (1966).

12



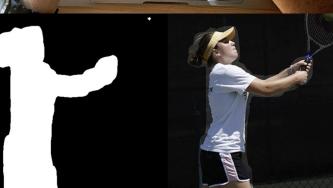
▼ Preview	🔽 🔻 Material	▼ Shaders	▼ Mirror Transp	Texture Map To	▼ Map Input	Ramps Links and Nodes
	•••	Lambe Ref 0.8	(Ray Mirror) (RayMir 0.2 Depth: 4)	Col Nor Csp Cmir Ref Spec Am Hard RayM Alph Emit TransL Disp	Glob Object UV Orco Stick Win Nor Refl	Link to Object
	VCol Lig VCol Pai TexFace Shadele No Mist Env	(Wardls =) (Spec 0.] Shado	Fresnel 2 Fac 1.2	(Sten Ne No RG) (Mix :)	Stress Tangent Flat Cube (ofsX 0.000)	(E:Circle.004 OB ME ← 1 Mat 1 →) Active Material Node
	Col (R 0.140	ms:0.1 OnlySh Bias	IOR 1.00 Depth: 2 Limit 10 Falloff 1	R 0.28	Tube Sphe	(± MA:noise 4 UV Tanget V Render Pipeline
	Spe G 0.140	GR:	Fresnel 0 Fac 1.2	G 0.28 Var 1.00 B 0.28 Disp 0.2	X Y Z X Y Z (sizeX 90.00)	Halo ZTransp offs: 0.000 Full Osa Wire Strands Zinvert
	(BGHSVIDYN A 1 000	Amb 0.5 Emit 0.0	(SpecTra	(War 1 War fac 0	X Y Z 4 size Z 0.30	Radio OnlyCas Traceabl Shadbuf



Virtual Background





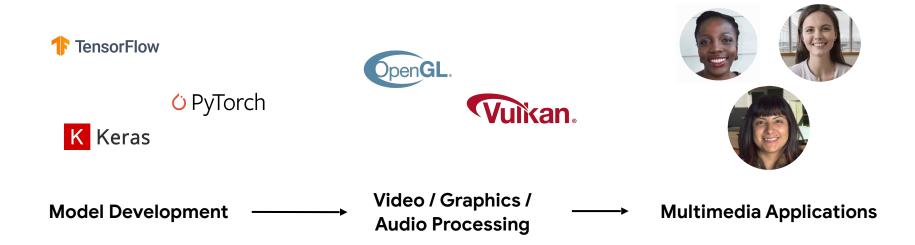




person 0.864

baseball glove 0.769







< Z □ INPUT 4 Image Mockup Camera Raw Sketch MODEL Pre-trained model Custom model EFFECT Effect1 Custom model 1 Effect1 Canvas Effect2 Show in preview param 1 0.8 drop here Image 1 param 2 0.5 OUTPUT 3, 256, 256 input Show in preview output 1024 drop here Canvas Key points 20 offset x Raw Custom model 2 (Keypoint offset y 50 rotation Ø Transformer 0 drop here MISC 2.4 scale overlay on 🔲 Image 1 W Performance 3, 256, 256 Show in preview input output 8 Show in preview

Design goals

Visual programming for rapid prototyping
 Run real-time ML pipelines
 Input in-the-wild
 Interactive data augmentation
 Side-by-side comparison
 Off-the-shelf & customize models

► rapsai → Visual Blocks for ML

A Visual Programming Platform for Rapid and Iterative Development of End-to-end ML-based Applications

N ~ \$	< > 10 ~	rapsai studio	
Q se	earch nodes	< + □ ■ WebML1 v C × Rapsai v S >	
Input	Audio		
Effect	Live Camera	Inputs: urls [ImageURL] image Image Outputs: image Image	
♦ ‡ Model			
⊡ Output			
[] Tensor			
₩ Misc		*	Nothing to inspect
V	ocal: Mich	elle Carney	

speed x6

€

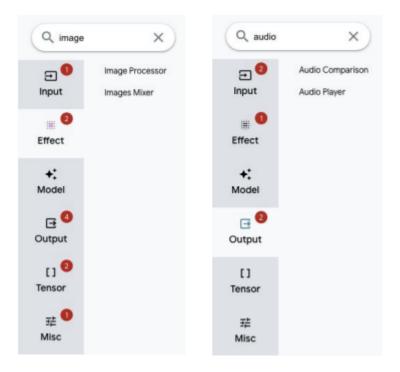
Rapsai System Overview

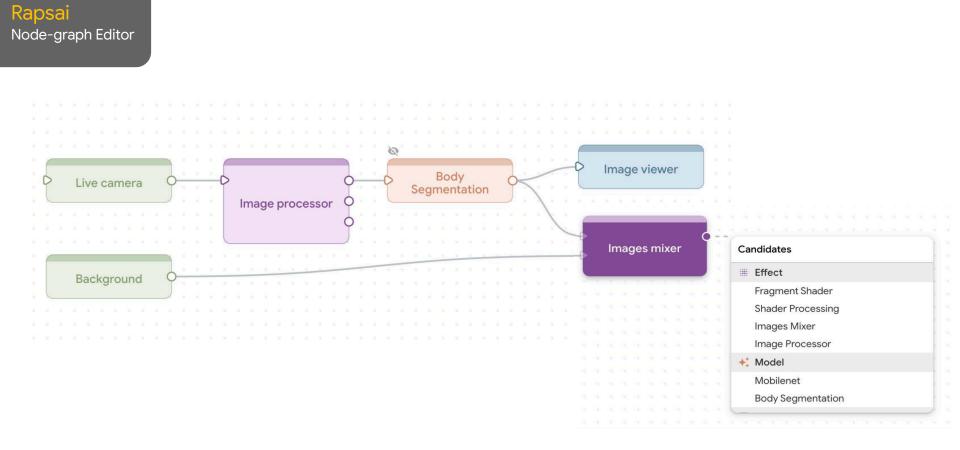


Rapsai Nodes Library

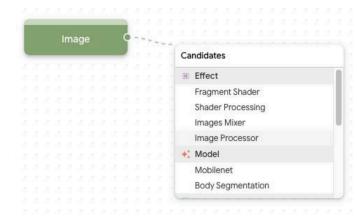
Q Search	h nodes	Q Searc	h nodes	Q Search	n nodes	Q Search	n nodes	Q Search	h nodes	Q Searc	h nodes
∋ Input	Audio Image	Input	Audio Processor Fragment Shader	.∋ Input	Body Segmentation CCA Denoise	于 Input	3D Model Viewer Audio Comparison	.∋ Input	Binary Op Clip By Value	于 Input	Get Image Size Get Size From Rect
iii Effect	Input Stream Live Camera Simple Audio	iii Effect	Image Processor Images Mixer Shader Library	iii Effect	Custom Model Runner Mobilenet	i∰ Effect	Audio Player Bar Viewer Device Image	iiii Effect	Const Tensor Crop And Resize	iiii Effect	Lobby Webpage
✦ ‡ Model	Video	+ ‡ Model	Shader Processing	✦ ‡ Model		+ ‡ Model	Image Comparison	✦ Model	Image To Tensor Postprocess Depth Model	✦ ‡ Model	
⊡ Output		⊡ Output		⊡ Output		⊡ Output	Json Viewer Output Stream	⊡ Output	Preprocess Image Remap Value Range Tensor Picker	⊡ Output	
[] Tensor		[] Tensor		[] Tensor		[] Tensor	Tensor To Depthmap Tensor To Image	[] Tensor	Tensor To ClassifierResults	[] Tensor	
] ∰Misc		辈 Misc		∃≟ Misc] ⊞ Misc	Tensor Viewer	≇ Misc		<u>∓</u> ≟ Misc	

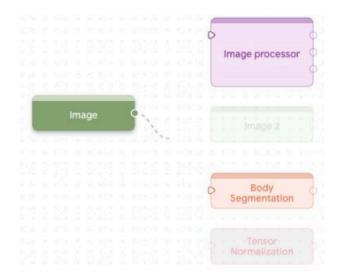
Rapsai Nodes Library





Rapsai Node-graph Editor

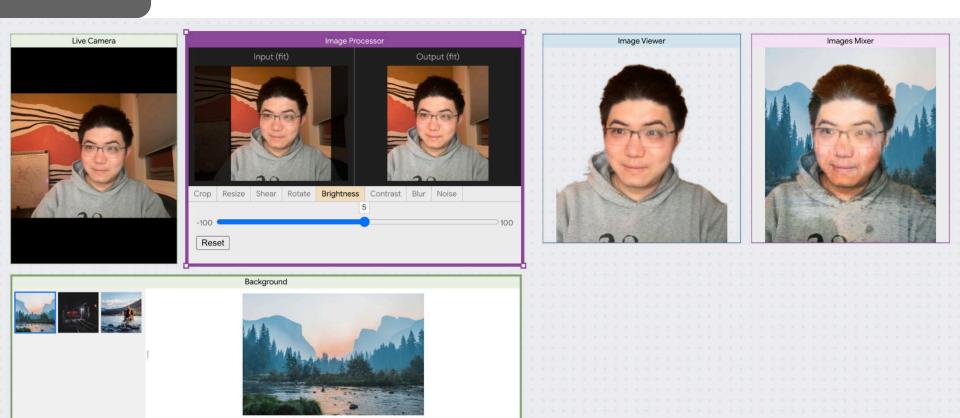




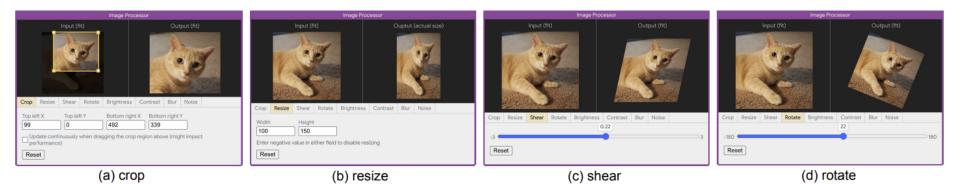
Node Suggestion

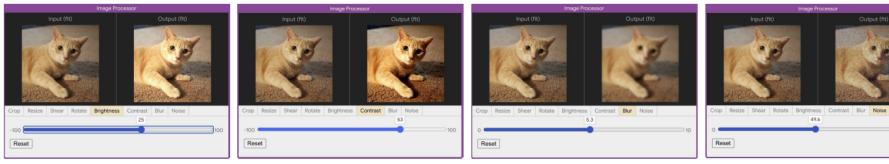
Link Suggestion











(e) brightness

(f) contrast

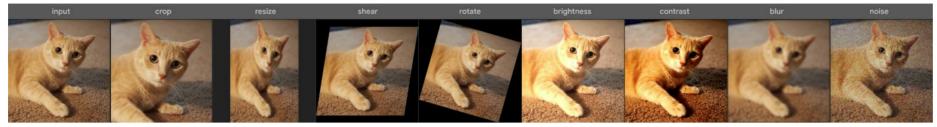
(g) blur

(h) noise

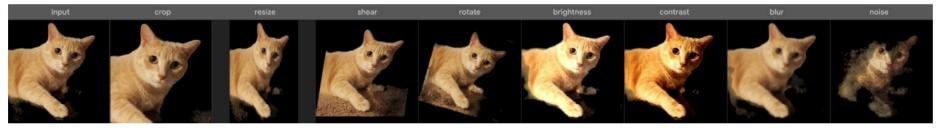
49.6

100

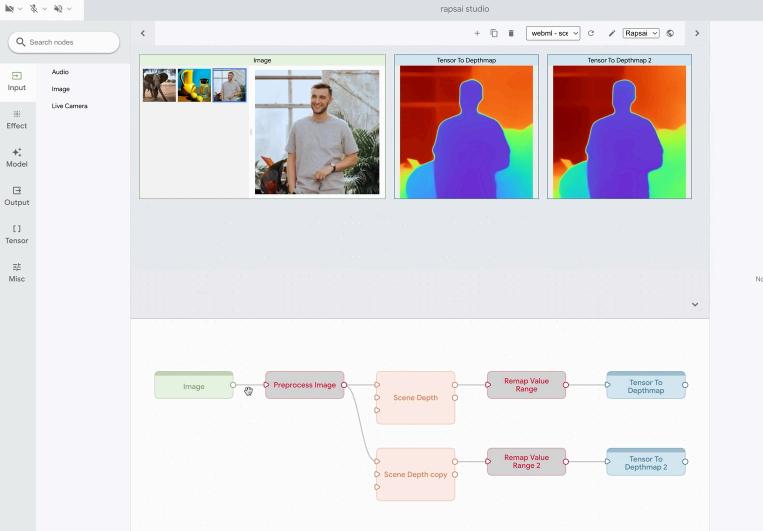




(a) intermediate results of interactive data augmentation applied onto a single input image



(b) side-by-side comparison of segmentation results with different augmentation techniques



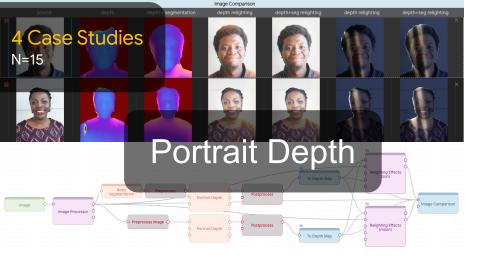
Nothing to inspect

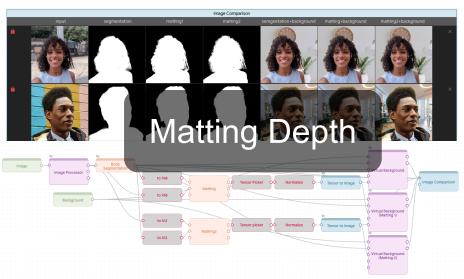
speed x8

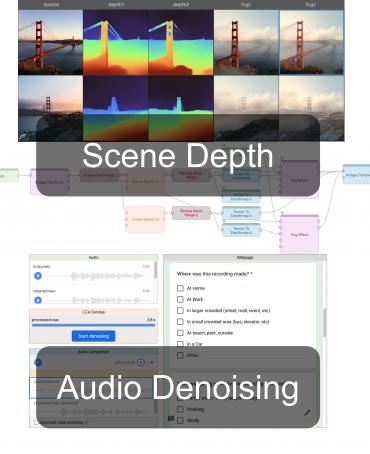
€

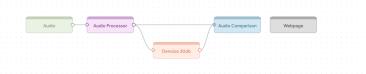




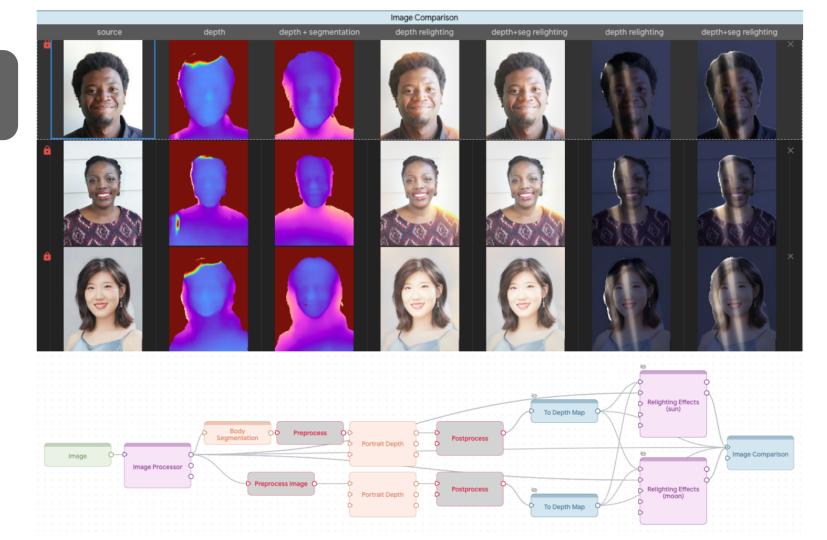




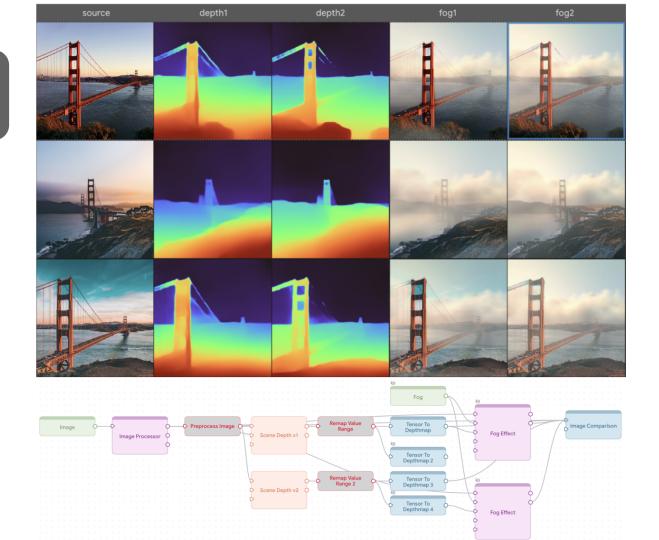




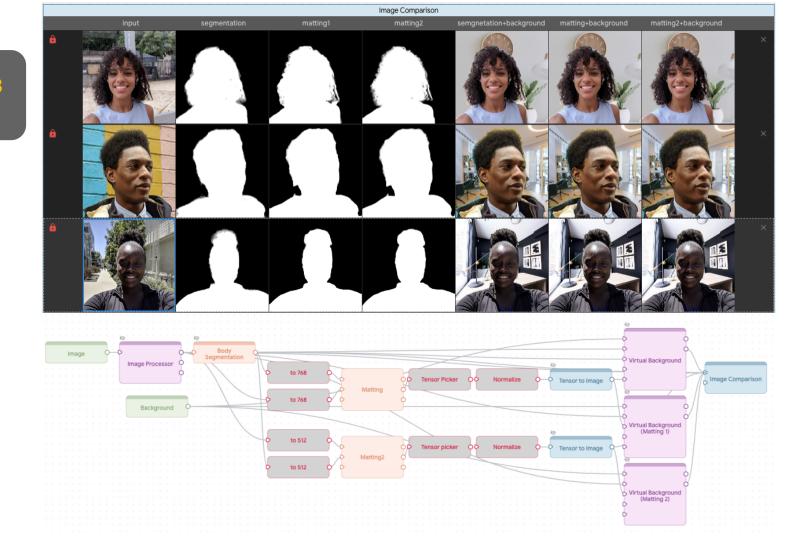








Case Study 3 Alpha Matte



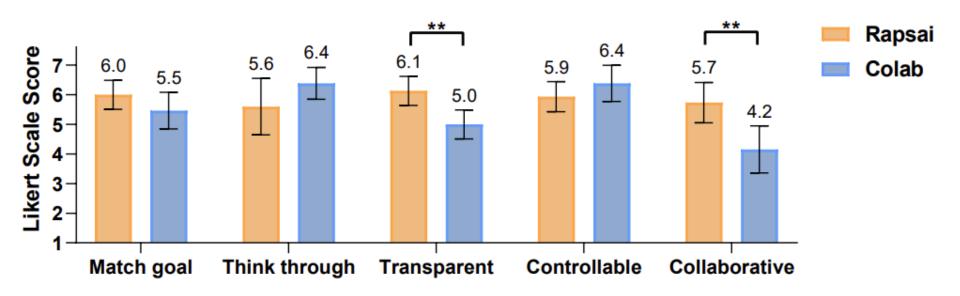
	Audio	Webpage				
	noisy.wav 3.0s	Where was this recording made? *				
Case Study 4 Audio Denoising	cleaned.wav 3.0s					
	CCA Denoise	In larger crowded (street, mall, event, etc)				
	processed.wav 3.8 s	In small crowded area (bus, elevator, etc)				
	Start denoising	At beach, park, outside				
		🗌 In a Car				
	Audio Comparison	Other:				
	Active track •					
	processed.wav	What was going on while this recording was made? *				
	processed.wav (denoised)	Party / event - many people talking				
		Honking				
	Automatic track switching ⑦	P Windy				
	Audio Audio Processor	Audio Comparison Webpage				



Background interview (6.1 ± 0.8 min)
Video tutorial (4 min),
Visual analytics procedure using Rapsai (39.4 ± 4.6 min)
Discussion of Rapsai and perception prototyping in future (10.2 ± 2.0 min)
Post-hoc exit survey to use Rapsai and compare with Colab

Findings 1 Rapsai vs Colab

Less Control but More Transparent and Collaborative



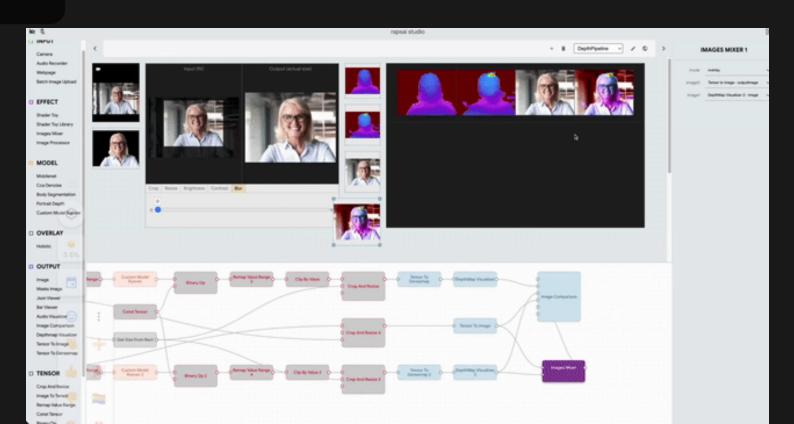
I spent about half a minute to create an image classification pipeline, and I spent 2–3 minutes to build a depth estimation pipeline from scratch, since it took some time to figure out how to preprocess the input and visualize the output... while Colab is more flexible for different tasks, I guess it could range from 1

hour to a day or two.

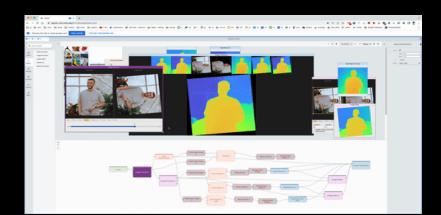
In my case, I started from an existing template but overall it was quite fast, I'd say less than 5 min.

Findings 2

Assist in Identifying Issues with ML Models and Training Sets



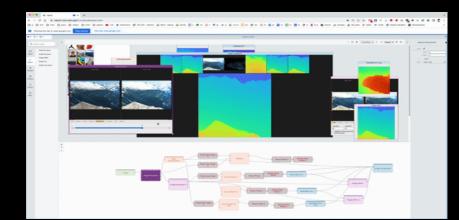
It can help me understand how I should change the model architecture and what training examples to add.



P10

"

I can manipulate the brightness to see when the model fails.



P2

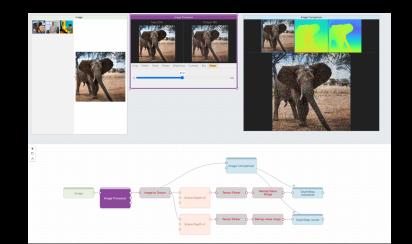
It gives me an intuition about which data augmentation operations that my model is more sensitive, then I can go back to my training pipeline, maybe increase the amount of data augmentation for those specific steps that are making my model more sensitive.





"

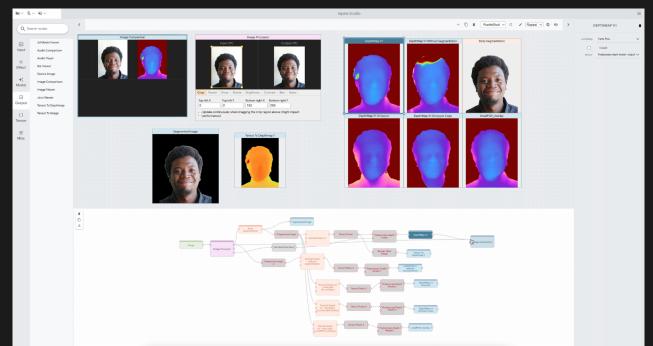
Using a video <as input> helps me get a cross-time feel of how the model performance varies, which is hard to capture with metrics. Comparing various noise parameters in the input to a model is useful to identify augmentation bias.



P8

Findings 3

Rapsai helps Model Selection, Learning From Pipelines, Study deployment





Building a custom webpage as debugging tool [by coding], cost <a junior engineer> over a month to build. This [Rapsai] is easy to distribute and try it immediately. It helps debug the pipeline.

"

It can help me understand how should I change the model architecture and add what training examples.





- 1. Lowers the barriers for ML prototyping
- 2. Empowers users to experiment with no/lowcode environment
- 3. Facilitates collaboration between designers and developers

Google Research

BLOG >

Visual Blocks for ML: Accelerating machine learning prototyping with interactive tools

FRIDAY, APRIL 21, 2023

Posted by Ruofei Du, Interactive Perception & Graphics Lead, Google Augmented Reality, and Na Li, Tech Lead Manager, Google CoreML

Recent deep learning advances have enabled a plethora of high-performance, real-time multimedia applications based on machine learning (ML), such as **human body segmentation** for video and teleconferencing, **depth estimation** for 3D reconstruction, **hand and body tracking** for interaction, and **audio processing** for remote communication.

However, developing and iterating on these ML-based multimedia prototypes can be challenging and costly. It usually involves a cross-functional team of ML practitioners who fine-tune the models, evaluate robustness, characterize strengths and weaknesses, inspect performance in the end-use context, and develop the applications. Moreover, models are frequently updated and require repeated integration efforts before evaluation can occur, which makes the workflow ill-suited to design and experiment.

In "Rapsai: Accelerating Machine Learning Prototyping of Multimedia Applications through Visual Programming",

presented at CHI 2023, we describe a visual programming platform for rapid and iterative development of end-to-end ML-based multimedia applications. Visual Blocks for ML, formerly called Rapsai, provides a no-code graph building experience through its node-graph editor. Users can create and connect different components (nodes) to rapidly build an ML pipeline, and see the results in real-time without writing any code. We demonstrate how this platform enables a better model evaluation experience through interactive characterization and visualization of ML model performance and interactive data augmentation and comparison. Sign up to be notified when Visual Blocks for ML is publicly available.

Visual Blocks for ML

Unleash your creativity

Visual Blocks for ML is an experimental JavaScript framework from Google that helps you add drag-and-drop machine learning blocks to your platform. Only your imagination limits the blocks you give your users. Off the shelf blocks include models, user inputs, processors and visualizations.

Colab experience and open source library will come soon.

Age How to create effects with models and shaders. 🤇 kiving 🗠 🗠 Dave Econt Import					
		V12H00W		Lie comes	Disker proteins to proteins carries to proteins to pro
	Consister Constant (main and/compression) (main and		•		Padar inducid be is backing format. Errore will be displayed in developer console.
	Liste Borney Gest calutaria (C) Hide president (C)	All O brage D brage Model: Foreground image v Hold provine C toos notice D brain code D brain	And a contract of the second sec	Visal dobr V	International Control Con

With Visual Blocks for ML, you can drag and drop to make your ML no coding required

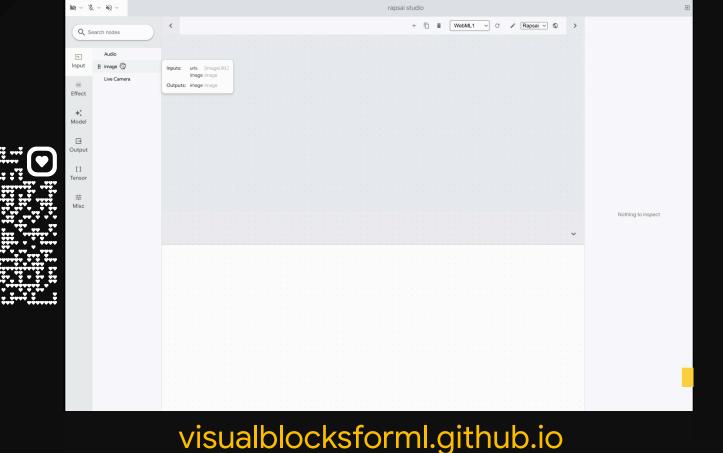
Experiment with Visual Blocks

Click on any of the demos to pull up its graph in the interaction editor below





With the right tools, everyone can unleash your inner creativity.





Ruofei Du, Na Li, Jing Jin, Michelle Carney, Scott Miles, Maria Kleiner, Xiuxiu Yuan, Yinda Zhang, Anuva Kulkarni, Xingyu "Bruce" Liu, Ahmed Sabie, Sergio Escolano, Abhishek Kar, Ping Yu, Ram Iyengar, Adarsh Kowdle, and Alex Olwal



Chapter Three · Digital Human & Augmented Communication



Montage4D I3D '18 JCGT '19 MonoAvatar & HumanGPS CVPR '23, CVPR '21



GazeChat & CollaboVR UIST '21 & ISMAR '20 Visual Captions & ThingShare CHI '23

What is Avatar?



NETROEME TO MAR ALCOMENT, ALLA SA MARTAREN TA ALLAN EMPELIORARI MARTINARI MARTINARI MARTINARI MARTINARI MARTINA Allem al televez an marco astronomente el martina el martinari del servico del servico del servico del servico Allem al televez antinomente astronomente el martinari del servico del servico del servico del servico del servi

THE WAY OF WATER

Сарана сала на собрат са на сала са сала са сала на собрат н Собрат на собр



Avatar is a term used in Hinduism for a material manifestation of a deity: "descent of a deity from a heaven"



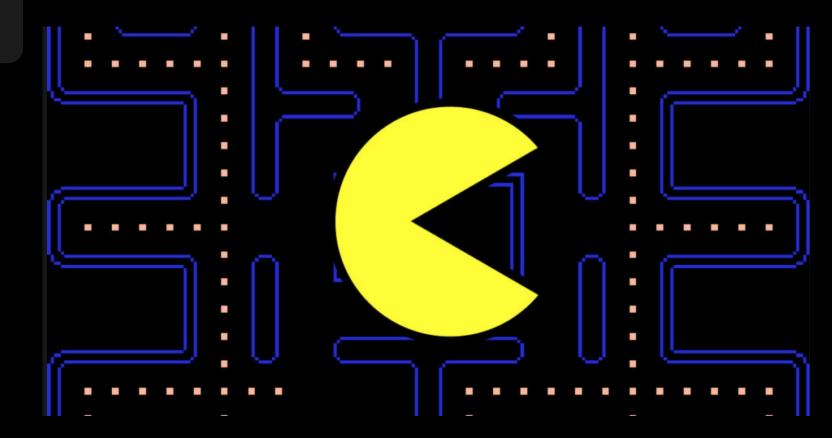
In computing, an avatar is a graphical representation of a user or the user's character or persona.

"



What is the *oldest avatar* in computer history?





Avatar History & Definition



Guo, Kaiwen, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey et al. "The relightables: Volumetric performance capture of humans with realistic relighting." ACM Transactions on Graphics (ToG) 38, no. 6 (2019): 1-19.

Dating back to real-time digital human / avatars...

Shahram Izadi Partner Research Manager

Montage4D: Interactive Seamless Fusion of Multiview Video Textures



Ruofei Du^{†‡}, Ming Chuang^{‡¶}, Wayne Chang[‡], Hugues Hoppe^{‡§}, and Amitabh Varshney[†] [†]Augmentarium | UMIACS | University of Maryland, College Park [‡]Microsoft Research, Redmond [¶]PerceptIn Inc. [§]Google Inc.

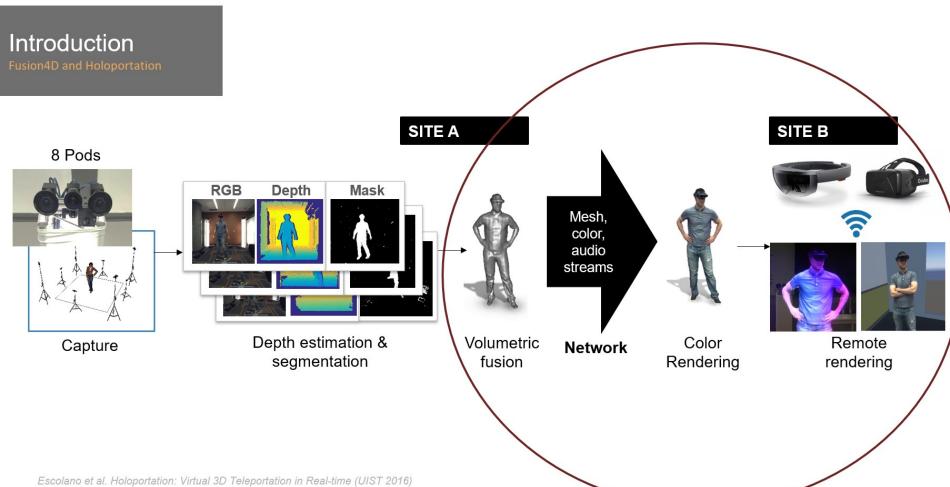


THE AUGMENTARIUM /IRTUAL AND AUGMENTED REALITY LABORA' IT THE UNIVERSITY OF MARYLAND









Fusing multiview video textures onto dynamic task with real-time constraint is **a challenging task**

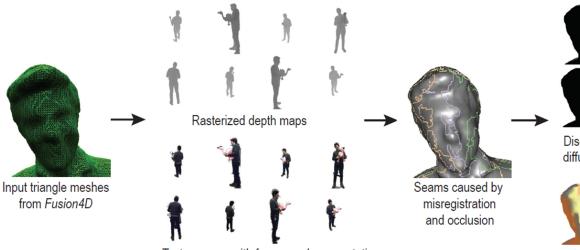


of the users does not believe the 3D reconstructed person looks real









Texture maps with foreground segmentation

Discrete geodesic distance fields to diffuse texture fields from the seams

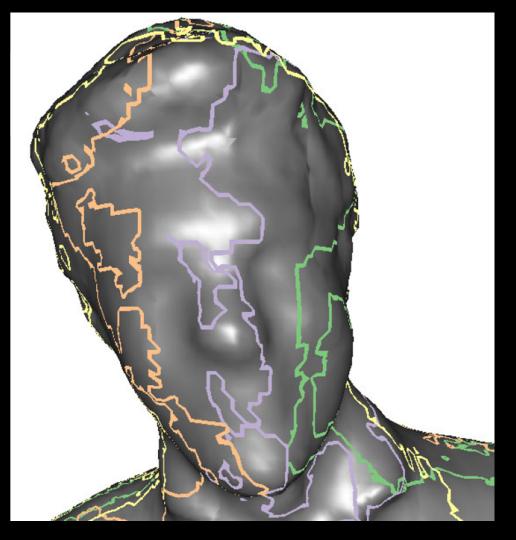


Update temporal texture fields

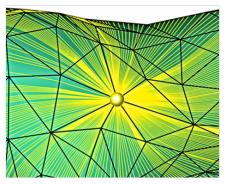


Montage4D Results





Geodesic is the **shortest** route between two points on the surface.



On triangle meshes, this is challenging because of the computation of **tangent directions**. And shortest paths are defined on **edges** instead of the vertices.

Approximate Geodesics

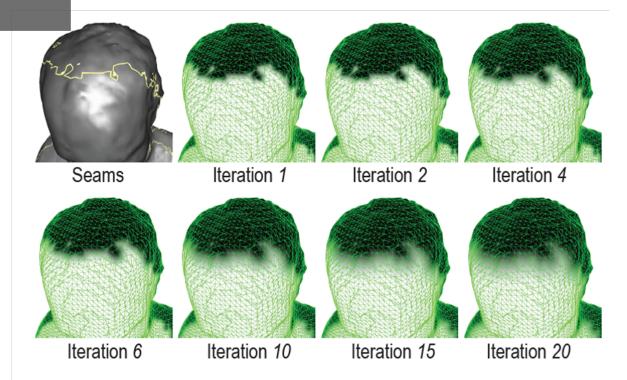
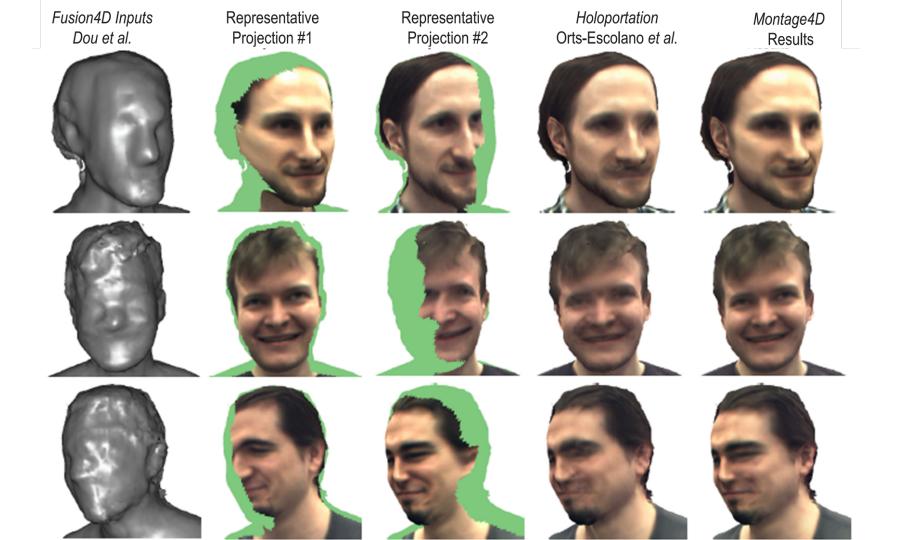
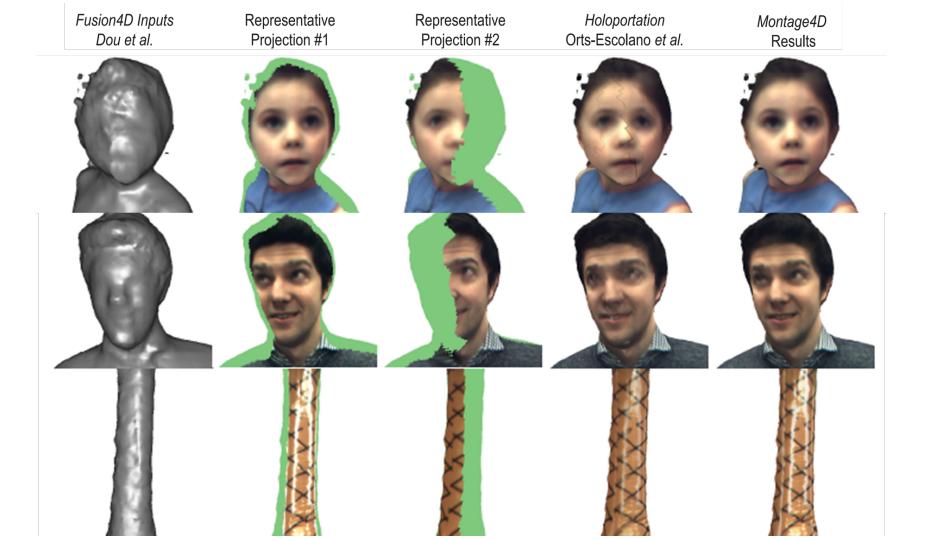


Figure 6: *Examples of the initial seam triangles and the propagation process for updating the geodesics.*



Color Scheme for the Texture Fields





What is the state-of-the-art since then?



Total Relighting: Learning to Relight Portraits for Background Replacement

ROHIT PANDEY^{*}, SERGIO ORTS ESCOLANO^{*}, CHLOE LEGENDRE^{*}, CHRISTIAN HÄNE, SOFIEN BOUAZIZ, CHRISTOPH RHEMANN, PAUL DEBEVEC, and SEAN FANELLO, Google Research

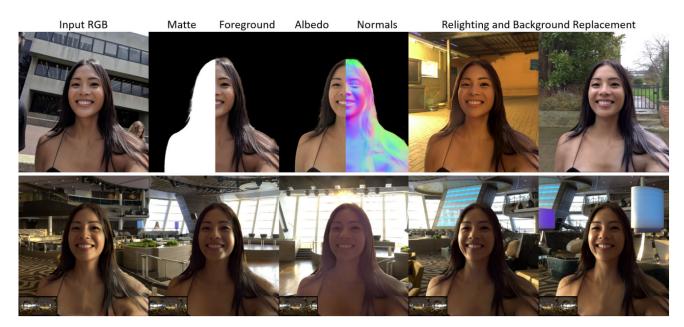
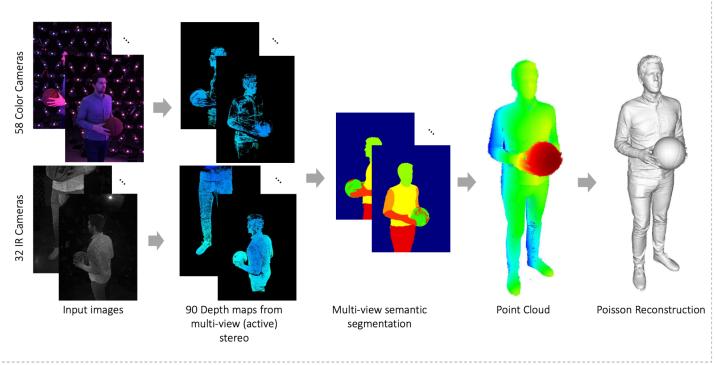


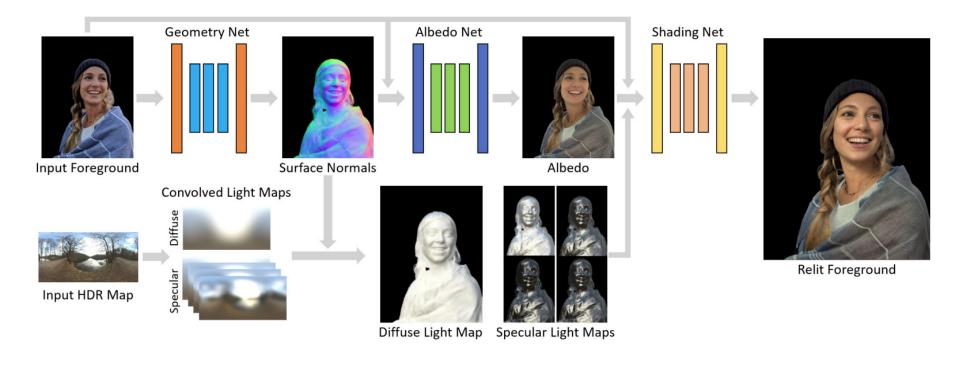
Fig. 1. Given a portrait and an arbitrary high dynamic range lighting environment, our framework uses machine learning to composite the subject into a new scene, while accurately modeling their appearance in the target illumination condition. We estimate a high quality alpha matte, foreground element, albedo map, and surface normals, and we propose a novel, per-pixel lighting representation within a deep learning framework.

Input \rightarrow Mesh

Photorealistic Characters The Relightables



Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, Danhang Tang, Anastasia Tkach, Adarsh Kowdle, Emily Cooper, Mingsong Dou, Sean Fanello, Graham Fyffe, Christoph Rhemann, Jonathan Taylor, Paul Debevec, and Shahram Izadi. 2019. The Relightables: Volumetric Performance Capture of Humans With Realistic Relighting. ACM Transactions on Graphics, pp. . DOI: <u>https://doi.org/10.1145/3355089.3356571</u>



Photorealistic Characters Rocketbox



Mar Gonzalez-Franco, Eyal Ofek, Ye Pan, Angus Antley, Anthony Steed, Bernhard Spanlang, Antonella Maselli, Domna Banakou, Nuria Pelechano, Sergio Orts Escolano, Veronica Orvahlo, Laura Trutoiu, Markus Wojcik, Maria V. Sanchez-Vives, Jeremy Bailenson, Mel Slater, and Jaron Lanier "The Rocketbox library and the utility of freely available rigged avatars." Frontiers in Virtual Reality DOI: 10.3389/frvir.2020.561558





Chen Cao, Tomas Simon, Jin Kyu Kim, Gabe Schwartz, Michael Zollhoefer, Shun-Suke Saito, Stephen Lombardi, Shih-En Wei, Danielle Belko, Shoou-I Yu, Yaser Sheikh, and Jason Saragih. 2022. Authentic Volumetric Avatars From a Phone Scan. ACM Transactions on Graphics, pp. . DOI: <u>https://doi.org/10.1145/3528223.3530143</u>

How can we build dynamic dense correspondence within the same subject and among different subjects?

HumanGPS: Geodesic PreServing Feature for Dense Human Correspondences

CVPR 2021

Feitong Tan^{1,2} Danhang Tang¹ Ruofei Du¹ Deging Sun¹

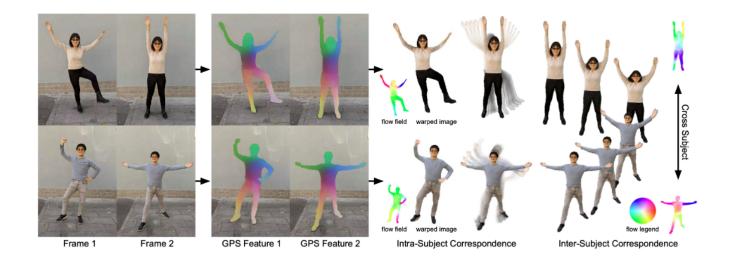
Mingsong Dou¹ Sofien Bouaziz¹

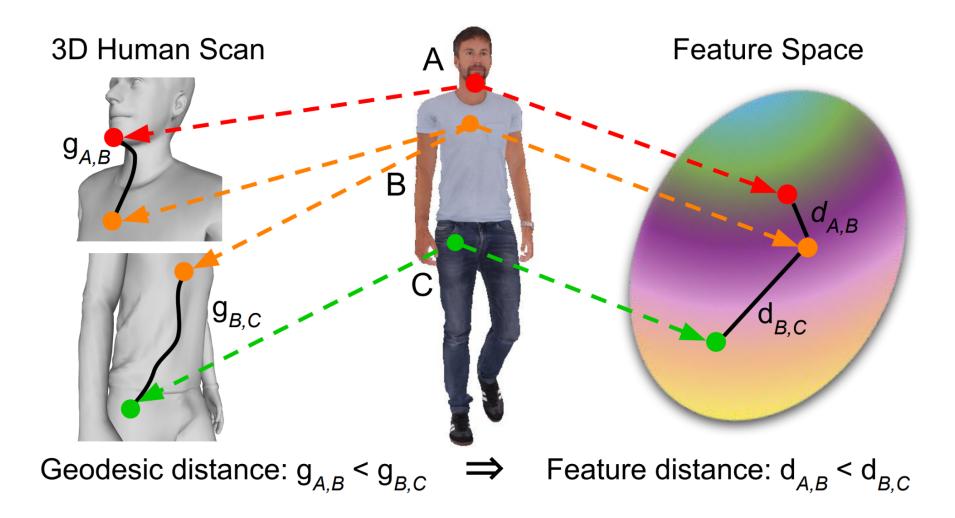
u¹ Kaiwen Guo¹ tiz¹ Sean Fanello¹ Rohit Pandey¹ Cem Keskin¹ Ping Tan² Yinda Zhang¹

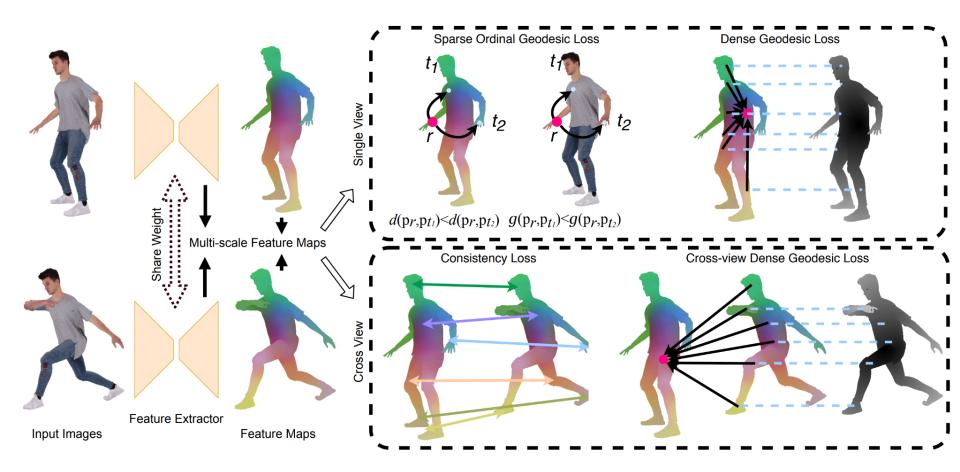
¹ Google

² Simon Fraser University









Live Demo

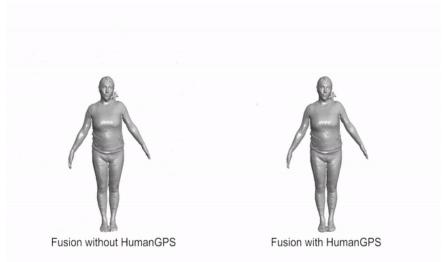
1. Click Choose File to upload a human image and a mask image (recommended w : h = 256 : 384) or use the example images: 1 2 3 2. Click 'Process' button to run the model. You may use Pen, Brush, Eraser, and Clear to doodle on the mask image. Note that the first time of 'Process' takes longer time for initialization. The model may take a few seconds to load in some region.





Choose File No file chosen











Fusion without HumanGPS

Fusion with HumanGPS

How can we leverage real-time Avatars today?

GazeChat

Enhancing Virtual Conferences With Gaze-Aware 3D Photos

Zhenyi He[†], Keru Wang[†], Brandon Yushan Feng[‡], Ruofei Du^{*}, Ken Perlin[†]

- [†] New York University
- [‡] University of Maryland, College Park
- * Google Research





UNIVERSITY OF MARYLAND

















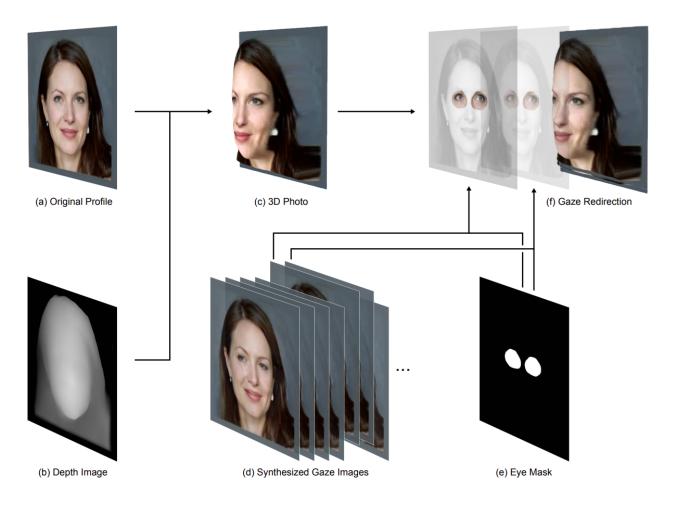
zhenyi



brandon



3D Photo Rendering 3D photos



3D Photo Rendering 3D photos



MonoAvatar: Learning Personalized High Quality Volumetric Head Avatars from Monocular RGB Videos

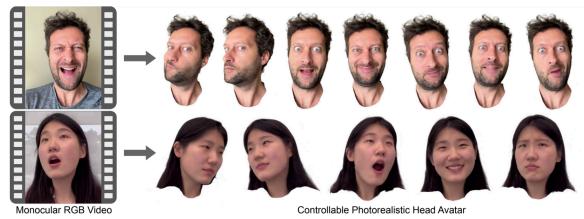
CVPR 2023

Ziqian Bai^{1,2}, Feitong Tan¹, Zeng Huang¹, Kripasindhu Sarkar¹, Danhang Tang¹, Di Qiu¹, Abhimitra Meka¹, Ruofei Du¹, Mingsong Dou¹,

Sergio Orts-Escolano¹, Rohit Pandey¹, Ping Tan², Thabo Beeler¹, Sean Fanello¹, Yinda Zhang¹,

¹ Google, ² Simon Fraser University





MonoAvatar builds a 3D avatar representation of a person using just a single short monocular RGB video (e.g., 1-2 minutes). We leverage a 3DMM to track the user's expressions, and generate a volumetric photorealistic 3D avatar that can be rendered with user-defined expression and viewpoint.



Coming to the new era...



How can we facilitate creative collaboration in XR?

CollaboVR: A Reconfigurable Framework for Creative Collaboration in Virtual Reality







avatars in front of a large virtual interactive board.

How can we further augment communication, in videoconferencing, AR, and XR in future?

Visual Captions Augmenting Verbal Communication with On-the-fly Visuals



Xingyu "Bruce" Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Alex Olwal, Peggy Chi, Xiang "Anthony" Chen, <u>Ruofei Du</u>

github.com/google/archat



Systems to Facilitate Verbal Communication







Visual Augmentations of Spoken Language

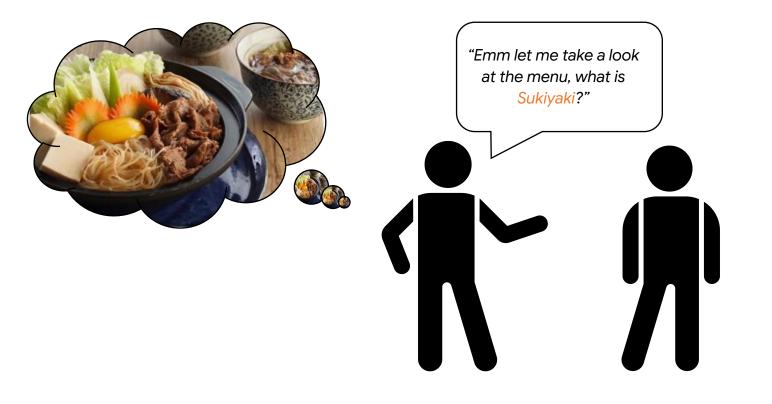


Chinese

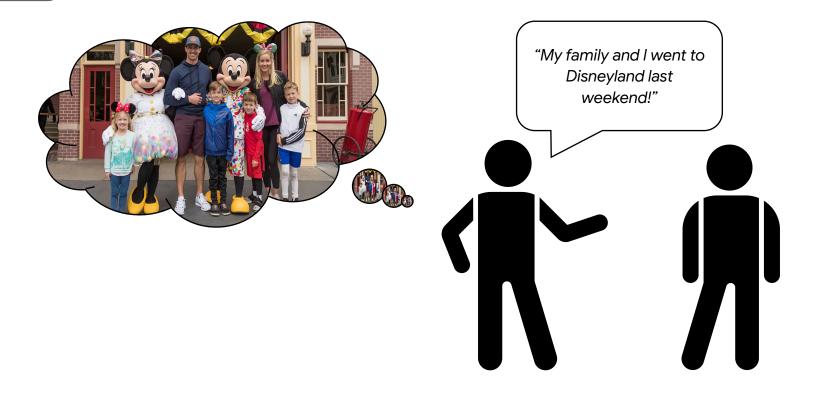
We can communicate with each other, I love you.



Motivatior



Motivation



A. VC1.5K Dataset

"Me and my family went to Disneyland, it was so fun!"
<a photo> of <Disneyland> from <online image search>
<a photo> of <me and my family at Disneyland > from <personal album>
<an emoji> of <happy face> from <emoji search>
"So where do you wnat to visit in LA?"
<a map> of <Los Angeles> from <online image search>

B. Visual Prediction Model

"Tokyo is located in the Kanto region of Japna"

"We spent our weekend in Yosemite"

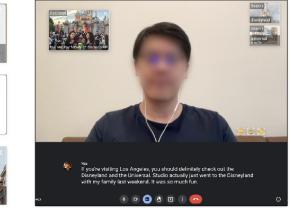
"You know, the triangle building in San Francisco."

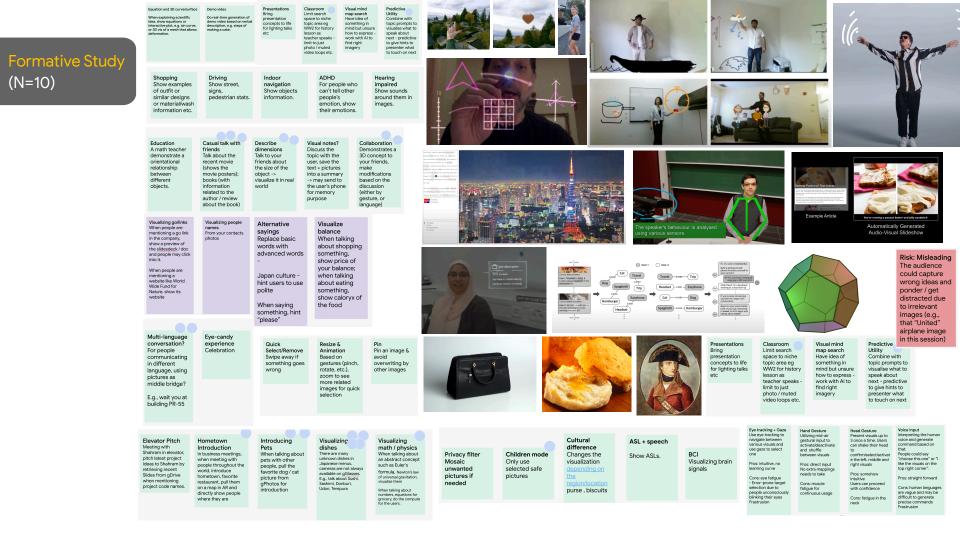




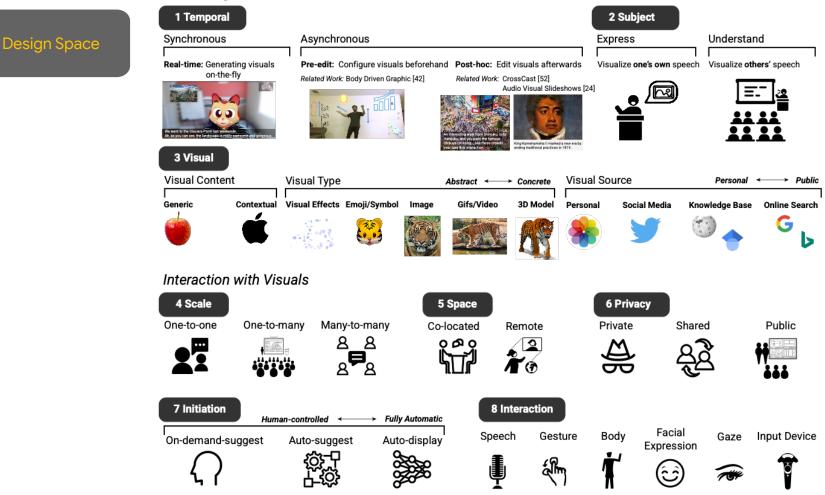


C. Visual Captions Interface





Generating Visuals





1 Temporal

Synchronous

Real-time: Generating visuals on-the-fly



Asynchronous

Pre-edit:Configure visuals beforehandPost-hoc:Edit visuals afterwardsRelated Work:Body Driven Graphic [42]Related Work:CrossCast [52]

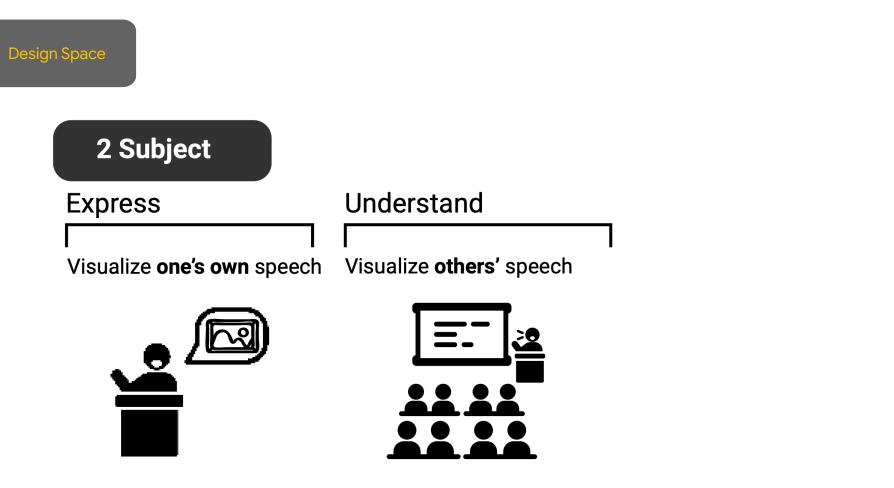






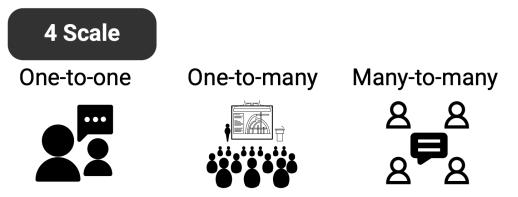
Audio Visual Slideshows [24]

King Kamehameha II marked a new era by ending traditional practices in 1819.



3 Visual Visual Type Visual Content Abstract -→ Concrete Contextual Visual Effects Emoji/Symbol Gifs/Video 3D Model Generic Image **Visual Source** Personal Public -> Personal **Social Media** Knowledge Base **Online Search** Ω 344

Interaction with Visuals



Design Space

5 Space

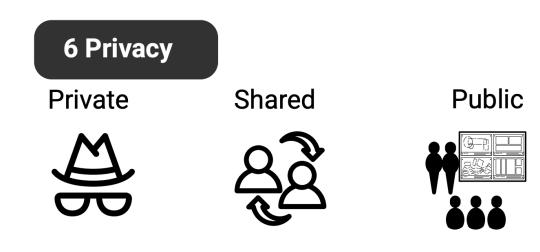
Co-located

Remote

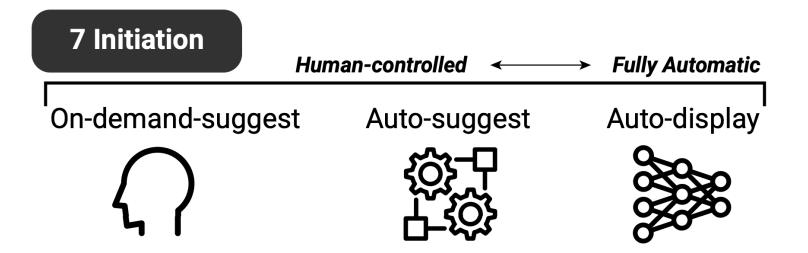




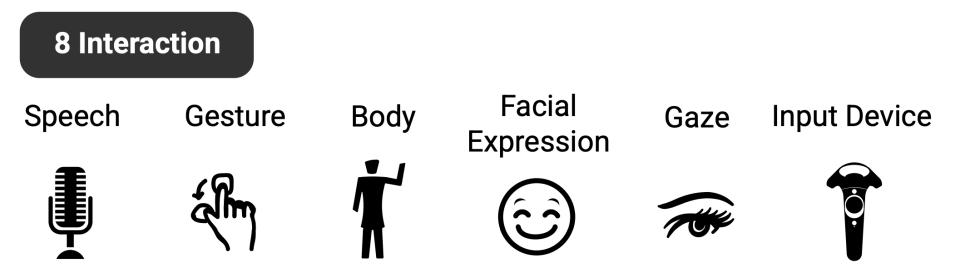
Design Space











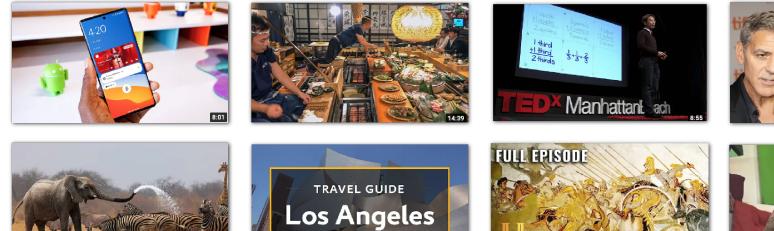


1595 sentence-visual pairs from 42 YouTube videos and the Daily Dialog datasets

246 MTurk workers

Task
Context: Talking about the best electronic products in 2021 Previous:"
Last Sentence: This is the top 10 gadgets that you can actually get your hands on that came out in the last 365 days."
Visuals to supplement the last sentence:
Example: A photo of Disneyland
Format: The visual should be:
(Please select)
Source: The visual should be retrieved from:
(Please select) V
Submit

VC 1.5K



48:05

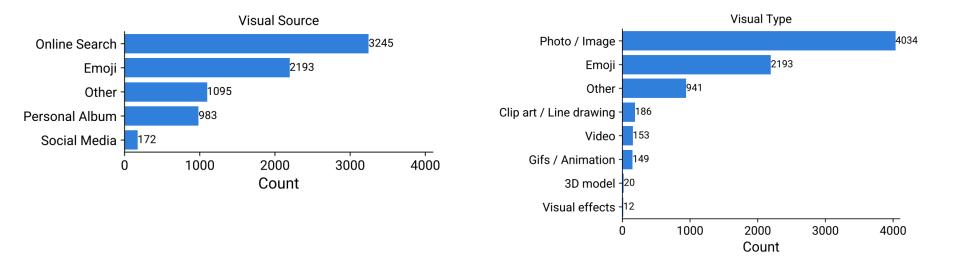
HISTO

8:50



5:48

VC 1.5K

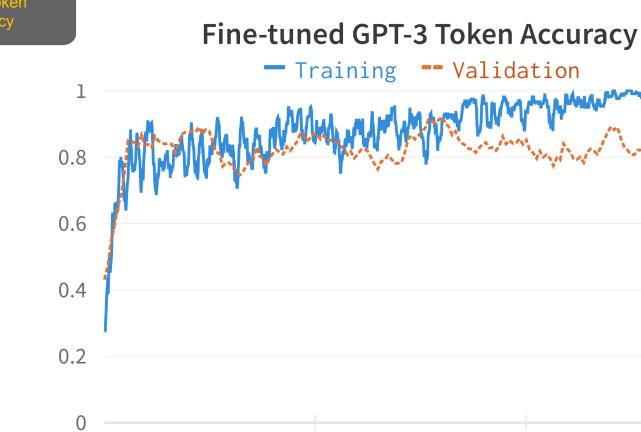


{"prompt": "<Previous Conversation> →",
"completion":
"<Visual Type 1> of <Visual Content 1> from <Visual Source 1>;

<Visual Type 2> of <Visual Content 2> from <Visual Source 2>;

•••

n



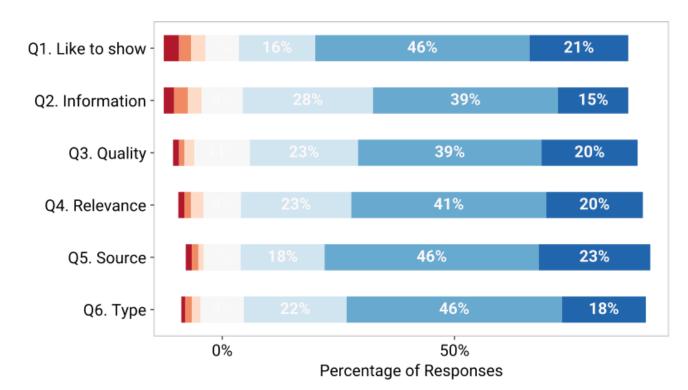
500



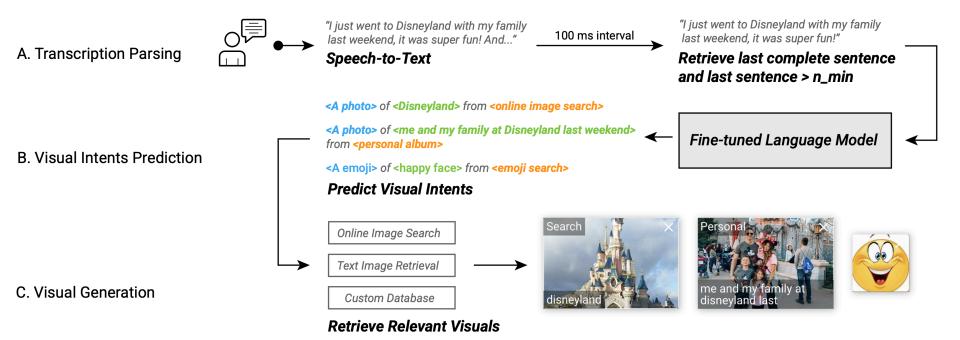
86% Token Accuracy



Crowdsourced Evaluation 846 Tasks









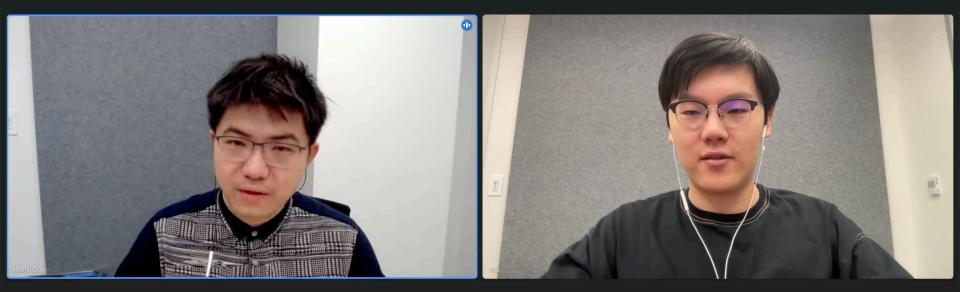
2:03 PM | uys-pdyp-bie



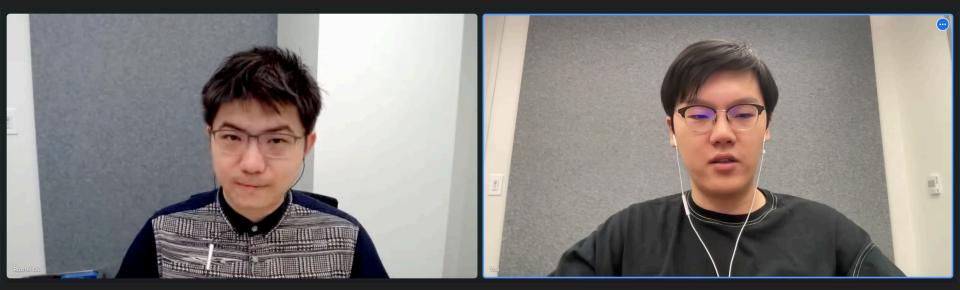
() 2. E & &

...

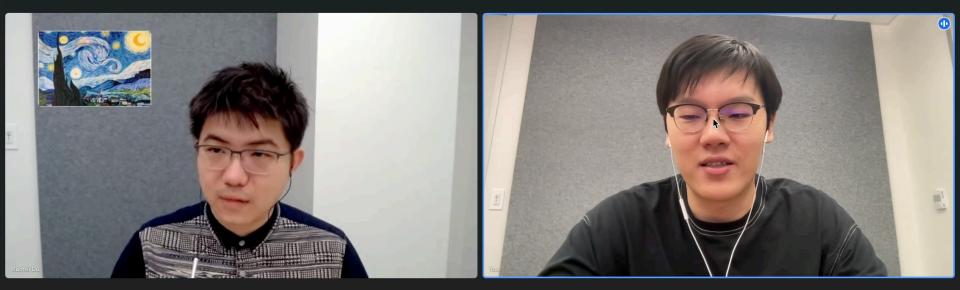










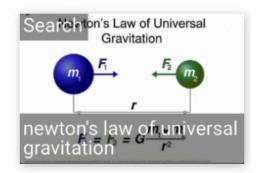




(2)

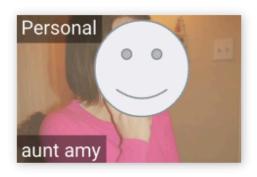
"We will cover the Newton's Law of Universal Gravitation"

(1) → Visual Content: Law of universal gravitation
 Visual Type: Diagram
 Visual Source: Internet Search



"Your aunt Amy will be visiting this Saturday."

Visual Content: Aunt Amy
 Visual Type: Photo
 Visual Source: Personal Album



"Tokyo is in the Kanto region of Japan."

(3) → Visual Content: Tokyo
 Visual Type: Photo
 Visual Source: Internet Search



(4) → Visual Content: Kanto Region of Japan
 Visual Type: Map
 Visual Source: Internet Search



Different Visual Content

"My favorite movie is the Matrix."

(5) → Visual Content: The movie Matrix Visual Type: Poster Visual Source: Internet Search

"In today's lecture, we will learn a mathematical concept, matrix"

(6) → Visual Content: A math matrix
 Visual Type: Diagram
 Visual Source: Internet Search



Searcholumns m rows $a_{1,1}$ $a_{1,2}$ $a_{1,3}$. . . a2,1 a_{2,2} a_{2,3} . . . $a_{3,1}$ a_{3,2} a_{3,3} . . . athematics matrix

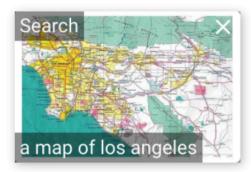
"Welcome to Los Angeles!"

(9) → Visual Content: Los Angeles
 Visual Type: Photo
 Visual Source: Internet Search

Search ×

"So where do you want to visit in LA?"

(10) → Visual Content: Los Angeles Visual Type: Map Visual Source: Internet Search



"Yosemite in the winter is really beautiful."

(7) → Visual Content: Yosemite in Winter
 Visual Type: Photo
 Visual Source: Internet Search



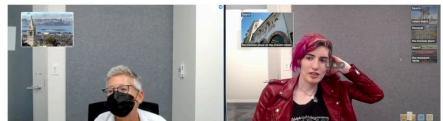
"We spent our weekend in Yosemite."

(8) → Visual Content: Yosemite
 Visual Type: Photo
 Visual Source: Personal Album



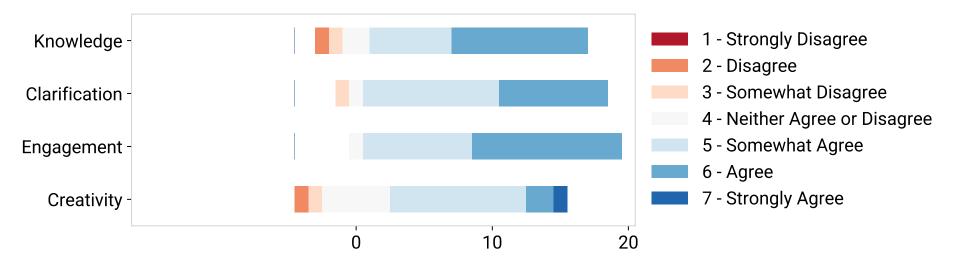
User Studies N=26











"

When I would really want visuals is like people don't know what I was talking about. For example when I just mentioned Santa Monica Pier, it's great that I can easily explain what it is.



"

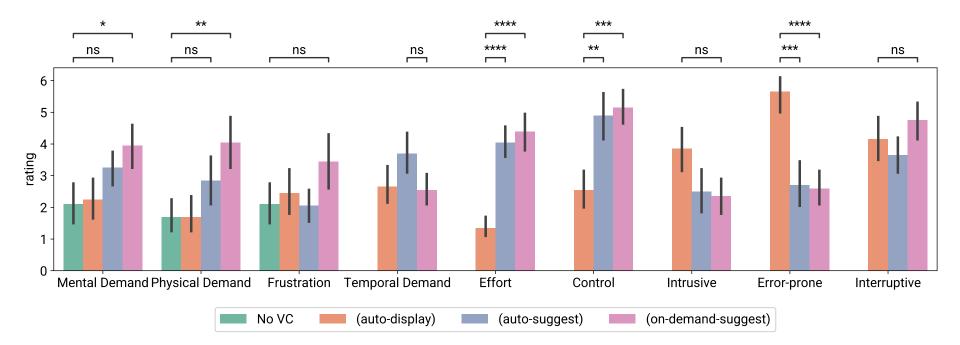
Back in the beginning when we were talking about the Avatar, there are like four or five different versions we might be discussing, the picture helped crystallize it instantly



"

It makes the conversation longer and more interactive.

Diverged Al Proactivity Levels



"Not having to click is huge for me."

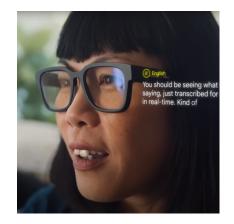
– P7 [Auto-Display]

"I like when these things pop up, so I really know what it is like " – P8 [Auto-Suggest]

"It's less mental overload and distraction because I would only activate it when I want."

- P13 [On-Demand]

Future Work Augmented Communication



Visual Captions for in-person conversations



Personalized Visual Suggestions



Integrating text-to-image models



github.com/google/archat

With ARChat, the HCI / AR / VR / NLP community can make communication more interactive, effective, and accessible with real world impact.

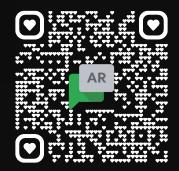
Visual Captions Augmenting Verbal Communication with On-the-fly Visuals



Xingyu "Bruce" Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Alex Olwal, Peggy Chi, Xiang "Anthony" Chen, <u>Ruofei Du</u>

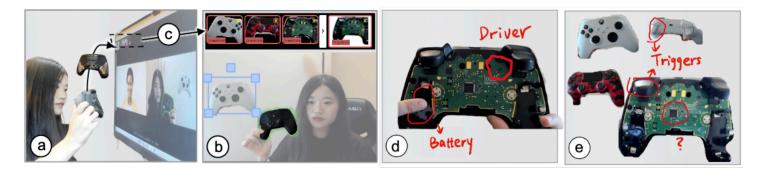
github.com/google/archat





ThingShare

Ad-Hoc Digital Copies of Physical Objects for Sharing Things in Video Meetings



Erzhen Hu, Jens Emil Grønbæk, Wen Ying, Ruofei Du, and Seongkook Heo

ACM CHI 2023











ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Things in Video Meetings

Erzhen Hu¹, Jens Emil Grønbæk², Wen Ying¹, Ruofei Du³, Seongkook Heo¹ University of Virginia¹, Aarhus University², Google Research³





How can Al benefit a broader inclusive community?

ProtoSound: A Personalized and Scalable Sound Recognition System for Deaf and Hard-of-Hearing Users

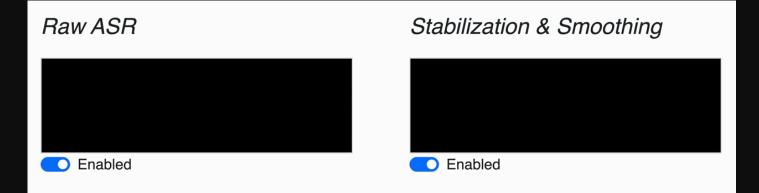
ACM CHI 2012 · Dhruv Jain, Khoa Nguyen, Steven Goodman, Rachel Grossman-Kahn, Hung Ngo, Aditya Kusupati, Ruofei Du, Alex Olwal, Leah Findlater, and Jon Froehlich



Figure 1: ProtoSound is a technique to customize a sound recognition model using very few recordings, enabling the model to scale across contextual variations of sound (*e.g.*, water flowing on a stainless steel *vs*. a porcelain sink) and support new user-specific sound classes (*e.g.*, a piano). Images show some example sound categories that were trained and recognized during our field evaluation using an experimental mobile app built off ProtoSound. See our supplementary video for details.

How can AI + Metaverse improve our life?

Modeling and Improving Text Stability in Live Captions



Xingyu "Bruce" Liu, Jun Zhang, Leonardo Ferrer, Susan Xu, Vikas Bahirwani, Boris Smus, Alex Olwal, and Ruofei Du





Live Transcribe & Notification

Research at Google

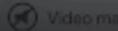
3.6 *





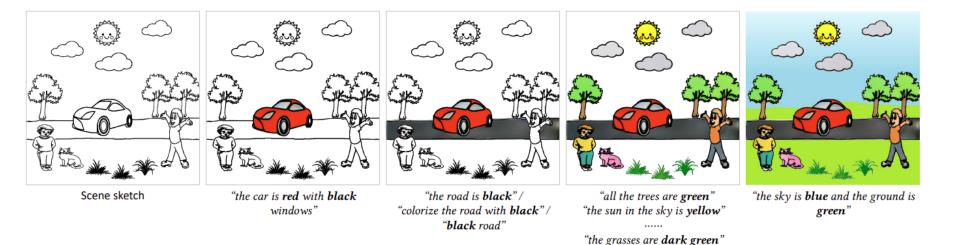


Introduc Live Trans



Language-based Colorization of Scene Sketches

Changqing Zou, Haoran Mo, Chengying Gao, Ruofei Du, and Hongbo Fu (ACM Transaction on Graphics, SIGGRAPH Asia 2019)



DALL E 2

DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language.

Imagen

unprecedented photorealism × deep level of language understanding

Google Research, Brain Team

We present Imagen, a text-to-image diffusion model with an unprecedented degree of photorealism and a deep level of language understanding. Imagen builds on the power of large transformer language models in understanding text and hinges on the strength of diffusion models in high-fidelity image generation. Our key discovery is that generic large language models (e.g. T5), pretrained on text-only corpora, are surprisingly effective at encoding text for image synthesis: increasing the size of the language model in Imagen boosts both sample fidelity and image-text alignment much more than increasing the size of the image diffusion model. Imagen achieves a new state-of-the-art FID score of 7.27 on the COCO dataset, without ever training on COCO, and human raters find Imagen samples to be on par with the COCO data itself in image-text alignment. To assess text-to-image models in greater depth, we introduce DrawBench, a comprehensive and challenging benchmark for text-to-image models. With DrawBench, we compare Imagen with recent methods including VQ-GAN+CLIP, Latent Diffusion Models, and DALL-E 2, and find that human raters prefer Imagen over other models in side-by-side comparisons, both in terms of sample quality and image-text alignment.







Future Directions Fuses Past Events Future Directions With the present

6

(E)

Ì

AT 603

1

00

S



Future Directions

Change the way we communicate in 3D and consume the information





Interactive Perception & Graphics for A Universally Accessible Metaverse



Ruofei Du

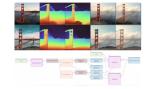
Senior Research Scientist / Manager Google AR www.ruofeidu.com Twitter: DuRuofei@ me@duruofei.com

Self Intro



🛗 🏛 in 🛩 f 💿 🗘 🔯

Featured Publications



Rapsai: Accelerating Machine Learning Prototyping of Multimedia Applications Through Visual Programming

Ruofei Du, Na Li, Jing Jin, Michelle Carney, Scott Miles, Maria Kleiner, Xiuxiu Yuan, Yinda Zhang, Anuva Kulkarni, Xingvu "Bruce" Liu, Ahmed Sabie, Sergio Escolano, Abhishek Kar, Ping Yu, Ram Iyengar, Adarsh Kowdle, and Alex Olwal

pdf, lowres, doi | project, video | cited by, cite



DepthLab: Real-Time 3D Interaction With Depth Maps for Mobile Augmented Reality 14K Installs

Ruofei Du, Eric Turner, Maksym Dzitsiuk, Luca Prasso, Ivo Duarte, Jason Dourgarian, Joao Afonso, Jose Pascoal, Josh Gladstone, Nuno Cruces, Shahram Izadi, Adarsh Kowdle, Konstantine Tsotsos, and David Kim Software and Technology (UIST), 2020.

pdf, doi | website, project, video, slides, code, demo, supp | cited by, cite



Stylization of Multiview Video Textures Microsoft TechFest 2018

Ruofei Du, Ming Chuang, Wayne Chang, Hugues Hoppe, and Amitabh Varshney

pdf, lowres, doi | website, project, video, slides | cited by, cite



Visual Captions: Augmenting Verbal Communication With On-the-Fly Visuals

Xingyu "Bruce" Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Peggy Chi, Xiang "Anthony" Chen, Alex Olwal, and Ruofei Du Proceedings of the 2023 CHI Conference on Human Factors in

Computing Systems (CHI), 2023.

pdf, lowres, doi | project, video | cited by, cite



Geollery: A Mixed Reality Social Media Platform Online Demo of a Metaverse of Mirrored World

Ruofei Du, David Li, and Amitabh Varshney Proceedings of the 2019 CHI Conference on Human Factors in

pdf, doi | website, project, video, slides, demo | cited by, cite



Montage4D: Real-Time Seamless Fusion and





Social Street View: Blending Immersive Street Views With Geo-Tagged Social Media Best Paper Award

Ruofei Du and Amitabh Varshney Technology (Web3D), 2016.

pdf, lowres, doi | website, project, video, slides | cited by, cite

Self Intro Ruofei Du (杜若飞)

Ruofei Du is a Senior Research Scientist at Google and works on creating novel interactive technologies for virtual and augmented realityresearch covers a wide range of topics in VR and AR, including AR interaction (**DepthLab**, **Ad** hoc UI), augmented communication (**CollaboVR**), mixed-reality social platforms (**Geollery**), video-based rendering (**Montage4D**), gaze-based interaction (**GazeChat**, **Kemel Foveated Rendering**), and deep learning in graphics (**3D Representation**, **HumanGPS**, **Sketch Colorization**). His research has been featured by Engadget, The Verge, PC Magazine, VOA News, cnBeta, etc. Du serves as an **Associate Editor** for IEEE Transactions on Circuits and Systems for Video Technology and Frontiers in Virtual Reality. He also served as a committee member in CHI 2021-2023, UIST 2022, and SIGGRAPH Asia 2020 XR. He holds 3 US patents and has published over 30 peer-reviewed publications in top venues of HCI, Computer Graphics, and Computer Vision, including CHI, SIGGRAPH Asia, UIST, TVCG, CVPR, ICCV, ECCV, ISMAR, VR, and I3D. Du holds a Ph.D. and an M.S. in Computer Science from University of Maryland, College Park, and a B.S. from ACM Honored Class, Shanghai Jiao Tong University. Website: https://duruofei.com

Google publications

24 publications

(i)

Personal Website Google Scholar

Ruofei Du

About

 Research
 Image: Human-Computer Interaction and Visualization
 Image: Human-Computer Interaction and Visualizat

Authored publications coog Fiters Sort by: Year ~ Research areas * Research areas * Year * ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Triggs in Video Meetings Ericht Lus and Unit Research areas * Year * ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Trings in Video Meetings Ericht Hu, Jens Emil Grantax, Wen Ying, Budel Bu, Seopkock Heo - Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI), ACM (to appear) Visual Capitions: Augmenting Visuals Year

		Autanate Kai, Filig Tu, Kain iyengai, Autanat Kowute, Alex owai • Frodeedings of the 2023 of the otherence on Human Factors in comparing systems (only, Kow (to appear)				
	+	ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Things in Video Meetings Erzhen Hu, Jens Emil Grønbæk, Wen Ying, <u>Rudfel Du</u> , Seongkook Heo 🔹 Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI), ACM (to appear)	i			
		Visual Captions: Augmenting Verbal Communication with On-the-fly Visuals Xngyu Tance Lu, Viadimir Kirlyik, Xuxiu Yuan, Eeggy.Chi, Xiang Xnthony' Chen, <u>Alex Clivial, Ruofel Du</u> · Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI, ACM (to appear)	i			
		"Slurp" Revisited: Using 'system re-presencing' to look back on, encounter, and design with the history of spatial interactivity and locative media Shengzhi Wu, Darsgh Byrne, <u>Buofel Du</u> , Molly Steenson · ACM Conference on Designing Interactive Systems, ACM (2022)	i			
		OmniSyn: Synthesizing 360 Videos with Wide-baseline Panoramas David Li, Yinda Zhang, Christian Haene, Danhang "Danny" Tang, Amitabh Varshney, <u>Ruofei Du</u> 🔹 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE	i			
		Opportunistic Interfaces for Augmented Reality: Transforming Everyday Objects into Tangible 6DoF Interfaces Using Ad hoc UI Budel Lu, Mathieu Le Go., <u>Blac Ukual</u> , Shenghi Wu, Danhang Danhary Tang, Yinda Zhang, Jun Zhang, David Joseph New Tan, <u>Eddarico Tombar</u> , David Kim · Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, ACM	i			
		PRIF: Primary Ray-based Implicit Function	i			

Self Intro _{Ruofei Du} (杜若飞)

for	Interactive Perception & Graphics Lead / Manager, <u>Google</u> Verified email at google.com - <u>Homepage</u>					
J.J.	VR / AR Interactive Perception Interactive Graphics Human Computer In	teraction Meta	verse	Cited by		VIEW
					All	Since 2
TITLE		CITED BY	YEAR	Citations h-index	870 16	
R Du, EL Turner, M I	time 3D Interaction with Depth Maps for Mobile Augmented Reality Dztisluk, L Prasso, I Duarte, J Dourgarian, J Atonso, edings of the 33rd Annual ACM Symposium on User Interface	104	2020	i10-index	23	
Kernel Foveated Rendering X Meng, R Du, M Zwicker, A Varshney Proceedings of the ACM on Computer Graphics and Interactive Techniques			2018			
Language-based Colorization of Scene Sketches C Zou, H Mo, C Gao, R Du, H Fu # ACM Transactions on Graphics (SIGGRAPH Asia 2019) 38 (6), 1-16			2019		111	
Montage4D: Rea R Du, M Chuang, W	al-time Seamless Fusion and Stylization of Multiview Video Textures Chang, H Hoppe, A Varshney Graphics Techniques (ACM I3D 2018) 8 (1), 1-34	57 *	2019	2016 2017 2018 201	9 2020 2021 202	2 2023
C Zou, Q Yu, R Du,	Ri <mark>chly-Annotated Scene Sketches</mark> H Mo, YZ Song, T Xiang, C Gao, B Chen, H Zhang e on Computer Vision (ECCV), 2018	55	2018	Public access		VIEW
R Du, D Li, A Varshn	d Reality Social Media Platform ey fings of the 2019 CHI Conference on Human Factors in	47	2019	not available Based on funding m	andatas	availa
Z He, R Du, K Perlin	configurable Framework for Creative Collaboration in Virtual Reality	44	2020	Dased on funding in	anuates	
X Meng, R Du, A Va	guided Foveated Rendering rshney n Visualization and Computer Graphics (TVCG) 26 (5), 1972	38	2020	Co-authors	rshney	VIEW
Video Fields: Fu R Du, S Bista, A Var	sing Multiple Surveillance Videos into a Dynamic Virtual Environment	28	2016	Alex Olwal	ge of Computer, Research Scien	
Multiresolution D Z Chen, Y Zhang, K	leep Implicit Functions for 3D Shape Representation Genova, S Fanello, S Bouaziz, C Haene, R Du, mational Conference on Computer Vision (ICCV)	27	2021		f Maryland, Colle	ege
A Log-Rectilinea	r Transformation for Foveated 360-degree Video Streaming CD Brumar, A Varshney s on Visualization and Computer Graphics (TVCG Honorable	26	2021	Danhang Ta Research S Que yinda Zhan Google Res	icientist, Google g	
Evaluating Hapti Text Using Finge	c and Auditory Directional Guidance to Assist Blind People in Reading Printed r-Mounted Cameras	d 26	2016	Adarsh Kov		eer a
ACM Transactions o	Jh, C Jou, L Findlater, DA Ross, JE Freehlich n Accessible Computing (TACCESS) 9 (1), 1-38			David Kim Staff Softwa	are Engineer at G	Google
R Du, A Varshney	w: Blending Immersive Street Views with Geo-tagged Social Media he 21st International Conference on Web3D Technology (Best	26	2016		Allen School of C	omp
Y Jiang, R Du, C Lut	aptive GUI Layout with OR-Constraints teroth, W Stuerzlinger	25	2019	Sean Ryan Research S	Fanello icientist and Man	ager
CHI '19: Proceeding	s of the 2019 CHI Conference on Human Factors in			🕋 Haoran Mo	(莫浩然)	

Follow

GET MY OWN PROFILE

VIEW ALL

Since 2018

821

15

0

VIEW ALL 19 articles available

VIEW ALL

>

>

>

>

>

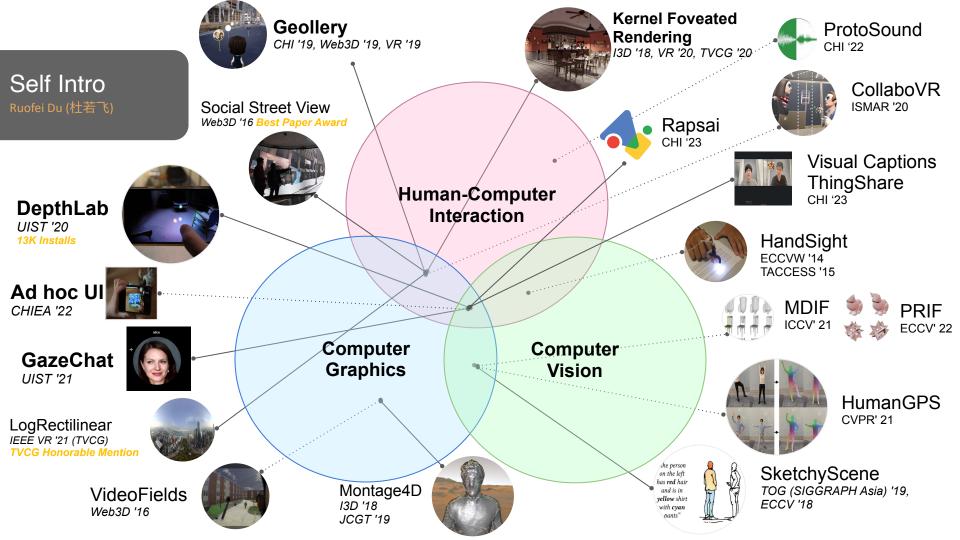
>

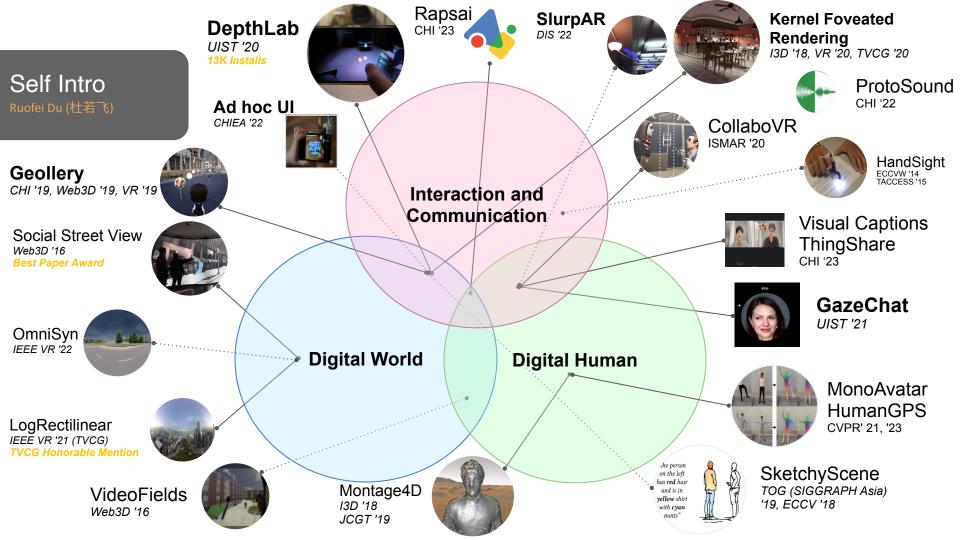
>

>

~

Ruofei Du





Interactive Perception & Graphics for A Universally Accessible Metaverse



Ruofei Du

Senior Research Scientist / Manager Google AR www.ruofeidu.com Twitter: DuRuofei@ me@duruofei.com

Metaverse

How Metaverse is defined by academia and industry?

Snow Cras

Neal Stephenson, 1992.

A STEVEN SPIELBERG FILM

Origin Ready Player One







Co-presence Gaming Accessibility Privacy

Security

<u>Blockchain</u> **Extended Reality (XR)** Things **Economics Digital Twin** NFT **Augmented Reality** of Metaverse nternet **The Future of Internet**

Virtual Reality

VR OS Mirrored World

Avatars Wearable AI Vision

Decentralization

Web 3.0

Neural

What about a non-technical perspective?

Metaverse envisioned a *persistent* digital world,

where people are fully connected as virtual representations.

More importantly, what research directions shall we devote to Metaverse?



metaverse → a medium to *make information more useful and* accessible and help people to live a better physical life Interactive Perception & Graphics for a Universally Accessible Metaverse

Chapter One · Mirrored World & Real-time Rendering

Chapter Two · Computational Interaction: Algorithm & Systems

Chapter Three · Digital Human & Augmented Communication