Multi-Camera Lighting Estimation for Photorealistic Front-Facing Mobile Augmented Reality

Yiqin Zhao*
Worcester Polytechnic Institute
Worcester, MA, USA
yzhao11@wpi.edu

Sean Fanello
Google
Mountain View, CA, USA
seanfa@google.com

Tian Guo
Worcester Polytechnic Institute
Worcester, MA, USA
tian@wpi.edu

ABSTRACT

Lighting understanding plays an important role in virtual object composition, including mobile augmented reality (AR) applications. Prior work often targets recovering lighting from the physical environment to support photorealistic AR rendering. Because the common workflow is to use a back-facing camera to capture the physical world for overlaying virtual objects, we refer to this usage pattern as back-facing AR. However, existing methods often fall short in supporting emerging front-facing mobile AR applications, e.g., virtual try-on where a user leverages a front-facing camera to explore the effect of various products (e.g., glasses or hats) of different styles. This lack of support can be attributed to the unique challenges of obtaining 360° HDR environment maps, an ideal format of lighting representation, from the front-facing camera and existing techniques. In this paper, we propose to leverage dual-camera streaming to generate a high-quality environment map by combining multi-view lighting reconstruction and parametric directional lighting estimation. Our preliminary results show improved rendering quality using a dual-camera setup for front-facing AR compared to a commercial solution.

CCS CONCEPTS

• **Computing methodologies → Mixed / augmented reality**; • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; • **Computer systems organization → Distributed architectures**.

KEYWORDS

Mobile AR; lighting estimation; multi-camera system

*Work completed while part-time interning at Google.

arXiv:2301.06143v1 [cs.CV] 15 Jan 2023

1 INTRODUCTION

Lighting estimation is an important and long-standing task in computer vision and graphics communities [6, 14, 21]. Over the past few years, we have observed a wide adoption of lighting estimation methods for end-user-facing applications. For example, scene lighting estimation is essential to render visually-coherent virtual objects in mobile AR applications [21, 26]; in computational photography, lighting estimation is useful because photos can be post-edited to have different lighting conditions [14]. In this work, we explore the key research questions in lighting understanding for an emerging application domain, *front-facing mobile AR*, particularly try-on applications where end users leverage handheld mobile devices such as smartphones to overlay products of interest on their faces. Mobile AR try-on apps promote online shopping and often require photorealism to provide user experiences on par with physical try-on.

Front-Facing mobile AR try-on shares many practical challenges with other lighting estimation apps, e.g., limited field-of-views (FoVs), but also has its own unique challenges and opportunities. For instance, even though commercial AR SDKs support the world-space device pose tracking via back-facing cameras, they do not currently support such tracking via the front-facing camera. Adding such support is not a mere integration of front-facing camera streams to existing algorithms but requires solving challenges that are rooted in moving objects (face in this case) and limited overlaps between images. On the other hand, front-facing mobile AR try-on is less sensitive to spatial variance because of the proximity of the observation (the phone camera) and the rendering (the face) positions [25]. This insensitivity allows us to operate in the 2D image space—rather than the more computation-intensive 3D space [25]—when generating the proper lighting information.

Providing a photorealistic and visually-coherent try-on experience can be boiled down to supporting reflection, shadow, and correct color tone [26]. All three features can be achieved by utilizing a 360° HDR environment map [4]. However, obtaining such an environment map on mobile devices is difficult for key reasons, including small FoVs and limited support for HDR streaming. Even in cases when high-end phones could capture HDR videos, they do not support *simultaneous multi-camera streaming*. Furthermore, front-facing AR, compared to back-facing AR, has an even more limited FoV, and the camera often fails to observe important environmental information that is useful for reflective rendering. In short, it is challenging to obtain high-quality environment maps even on modern mobile devices.

This paper outlines a research roadmap that leverages *multi-camera* to deliver high-quality lighting information for front-facing
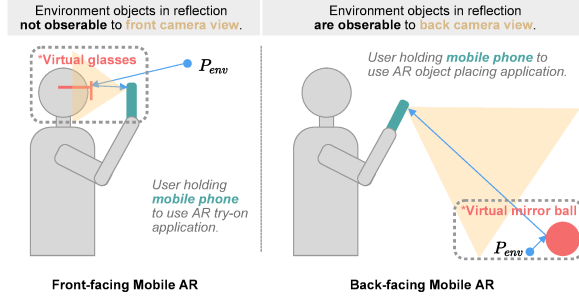
Figure 1: The challenge of reflective rendering. Mobile phones and example light rays are illustrated in green and blue, respectively.

mobile AR. State-of-the-art approaches either only support a subset of the features [26] or generate incoherent reflection [21]. To overcome the limitations of prior work, our key idea is to generate a high-quality environment map by *combining multi-view lighting reconstruction and parametric directional lighting estimation*. Our system design uses LDR multi-camera video streaming as input because streaming dual-camera LDR images requires less energy, compared to HDR streaming, and is better supported across recent mobile devices. Our proposed three directions are grounded on our preliminary investigation of dual-camera streaming (§3). To efficiently increase mobile device observation, we propose to regulate the back-facing camera usage by *time and mobility* to conserve energy. We propose a simple and lightweight face-tracking component for stitching camera views to generate an environment map. Lastly, we propose further improving the completeness and color tone of the environment map and supporting shadow casting by extending our parametric directional lighting estimation model with the back-facing camera streaming.

In summary, we make the following key contributions:

- We characterize the challenges of supporting photorealistic front-facing mobile AR, an important and emerging application domain of the lighting estimation task.
- We demonstrate that using a dual-camera setup can improve the mobile AR try-on experience with more visually-coherent rendering, especially for reflective objects.
- We describe key questions and promising designs for achieving a visually-coherent virtual try-on experience in a dual-camera AR system, including reflection, accurate color tone, and shadow casting.

2 CHALLENGES

Front-Facing mobile AR try-on, an emerging application of lighting estimation, exhibits new and unique challenges compared to other mobile AR shopping apps, e.g., IKEA Place[1]. Because these AR apps often use the back-facing camera, in this paper, we refer to them as *back-facing AR*. Figure 1 illustrates the key usage difference between these two types of AR apps and the implication on reflective rendering. In the back-facing AR, the user often points the mobile camera toward a rendering position; the environment information

Table 1: Comparisons of AR framework lighting estimation support for front-facing mobile AR apps.

Framework	Ambient	Directional	Spherical Harmonics	Env. Map
ARKit[2]	Yes	Yes	Yes	No
ARCore[3]	Yes	No	No	No
Proposed	Yes	Yes	Yes	Yes

required for rendering is often included in the corresponding camera views. In contrast, users often hold the mobile camera relatively closer to their faces when they use front-facing AR apps. Because of the capturing direction of the front-facing camera, it often misses the relevant environment information, making reflection rendering more challenging.

Existing mobile AR-based try-on apps leverage depth estimation, 3D face modeling [18], and real-time rendering to improve object fitting and pose tracking. However, due to the usage of reflective materials in many product designs, visual incoherence can occur if the 3D product objects are not rendered with the correct environmental lighting. At a high level, environment lighting requires a complex data structure representing omnidirectional environment observations and directional intensity variations. Traditionally, high-quality environment lighting can be obtained using specialized hardware like physical light probes or 360° cameras, which are cumbersome to be integrated into mobile AR apps. Table 1 shows that existing mobile AR frameworks only provide partial lighting support and cannot effectively leverage the new features on modern mobile phones like multi-camera streaming. In contrast, our proposed framework will provide a comprehensive set of lighting-related features for front-facing mobile AR.

3 THE PROMISE OF MULTI-CAMERA

We show that dual-camera streaming, a recent hardware advancement supported by major phone vendors, can improve the visual results for front-facing AR-based try-on. Furthermore, we show that EnvMapNet [21], a state-of-the-art lighting estimation model, cannot fully utilize the dual-camera inputs. Additional setup details are described below, and the research challenges and proposed solutions are presented in §4.

Setup. We used a tripod to fix the mobile phone position and then captured the front-facing and back-facing camera images, for three indoor scenes. Specifically, we used the front-facing, wide, and ultra-wide back-facing cameras, with 70°, 75°, and 120° FoVs, on iPhone 14 Pro. Next, we imported the captured images into `Blender` and generated virtual sunglasses on top of the user's face using the front-facing camera image. We tested the energy usage of multi-camera streaming using Apple's sample dual-camera capturing code[4]. We measured the device battery usage by running the dual-camera streaming app for 10-minute sessions on an iPhone 11 Pro. We set the screen brightness value to 50% and disabled all background tasks and networking, and measured the idle device battery consumption. After each measurement, we recharged the battery to 90%. We tested 6 configurations by keeping the front-facing camera always on and turning on/off the back-facing camera programmatically.

---

[1]https://tinyurl.com/yrruy65r

[2]https://developer.apple.com/augmented-reality/

[3]https://developers.google.com/ar
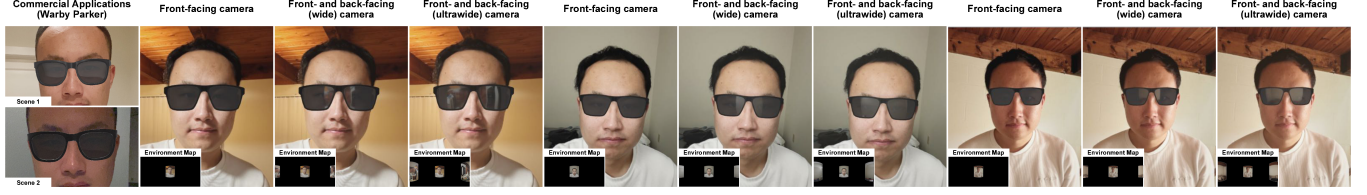
[4]https://tinyurl.com/mr32e7ab

Figure 2: Reflection rendering comparisons of different rendering setups. Warby Parker uses a static environment map which leads to incorrect reflections. Switching from front-facing to dual-camera streaming results in a more complete environment map and improved reflection rendering.

We repeated the measurement 3 times for each configuration and reported the average value.

3.1 Rendering Result Analysis

Figure 2 shows the rendering visual results using different camera setups and a commercial try-on app Warby Parker [5]. We observe that combining front-facing and back-facing camera views improves the environment map's quality and, consequently, the rendering quality. Specifically, the rendered sunglasses in the camera setup with the largest FoV, i.e., *front-facing and (ultra-wide) back-facing cameras*, are more visually realistic and coherent than the other two renderings.

Capturing high-quality ground truth lighting information is often challenging and requires using specialized hardware [20]. Prior work used a sphere metal ball to recover lighting information. Still, the resulting environment map can often show distortion and lower quality (compared to direct camera capture) [26]. In this work, we use the rendering results from the *front-facing and (ultra-wide) back-facing cameras* as the basis to calculate the PSNR. To avoid the impact of non-virtual objects on the calculation, we only calculate the PSNR of rendered virtual sunglasses between different camera setups. For front-only and front+wide back-facing camera setups, we calculated the average PSNR of 25.70db and 27.05db, respectively. The results align with the visual appearances and indicate the lower qualities of the front-facing camera setup.

To quantify the environment map's completeness, we used a metric from our prior work that calculates the environment observation coverage over uniformly sampled directions [25]. Assume that the front-facing and back-facing wide cameras share the same origin and the image ratio of 4:3. This setup covers 22.81% of the observation directions, 2.14x of using the front-facing camera only.Switching from the wide to the ultra-wide camera leads to 38.28% direction coverage, 3.6x of using front-facing camera alone.

3.2 Evaluating the Performance of EnvMapNet

We evaluated EnvMapNet's ability to generate HDR environment maps for dual-camera setup (dual camera mask) by training the model following the hyper-parameters and training setting described in the paper [21]. The masks are generated assuming a 70° front-camera FoV and a 120° back-camera FoV. We constructed a dataset using 450 online free HDR environment maps and 2100 data items from the Laval HDR dataset [5]. We observe that the dual-camera setup improves the average PSNR of predicted environment

maps from 10.3db/11.7db (original/single camera mask) to 12.1db. Figure 3 shows an example of visual comparisons. As we can see, even though dual-camera improves the prediction accuracy, the environment map details and light source directionalities are still inaccurate.

3.3 Energy Consumption Measurement

Table 2 shows the relationship between various dual-camera streaming configurations and battery usage. For a 10-minute session, simultaneously streaming front-facing and back-facing cameras takes on average 5% of total battery, 2.5x of capturing video using the front-facing camera alone. When reducing the back-facing camera capturing time to 3 minutes, the total battery usage drops to 3%. This suggests that selectively turning on and off the back-facing camera when it is not needed can conserve battery. This observation provides an important insight for our adaptive dual-camera capturing policy design, as will be described in §4.1. Further, our testing app consumes 62.7MB of memory in the front-facing-only mode while using 78MB of memory when streaming with both cameras, a 24.4% increase. Our empirical observation aligns with a recent work that leverages two back-facing cameras to remove face motion blur [13].

4 RESEARCH ROADMAP

This section presents three key questions and promising designs for achieving a visually-coherent virtual try-on experience in *a dual-camera AR system*, including reflection, accurate color tone, and shadow casting. Figure 4 illustrates how solutions to these questions could improve the existing front-facing mobile AR workflow.

- **Q1:** How to efficiently increase the mobile device observation? The ideal solution should fully utilize the existing mobile hardware in an energy-conserving way.
- **Q2:** How to effectively fuse camera observations to generate an environment map? The ideal solutions should handle temporal visual data in a noisy environment.

Table 2: Battery usage for dual-camera streaming.

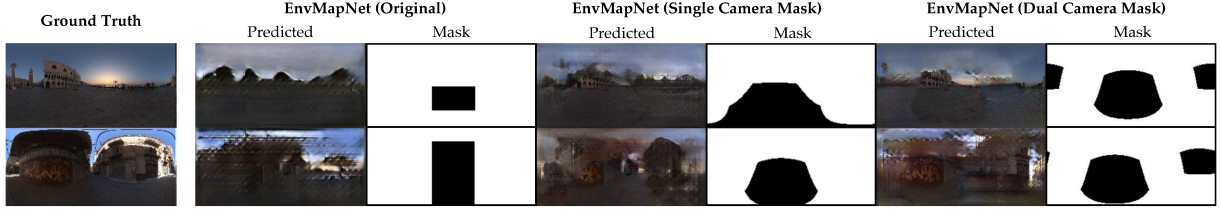| Front-Facing (mins)	Back-Facing (mins)	Battery Usage
0	0	<1%
10	0	2%
10	1	2%
10	3	3%
10	5	4%
10	7	5%
10	10	5%

---

[5]https://www.warbyparker.com

Figure 3: Visual comparisons of EnvMapNet [21] with different masks. We apply different masks to simulate different camera view setups. In addition to EnvMapNet [21]'s original mask inputs, we provide two mask setups with 70° front-camera FoV and a 120° back-camera FoV. We observe that even though dual-camera improves the prediction accuracy, the environment map details and light source directionalities are still inaccurate.
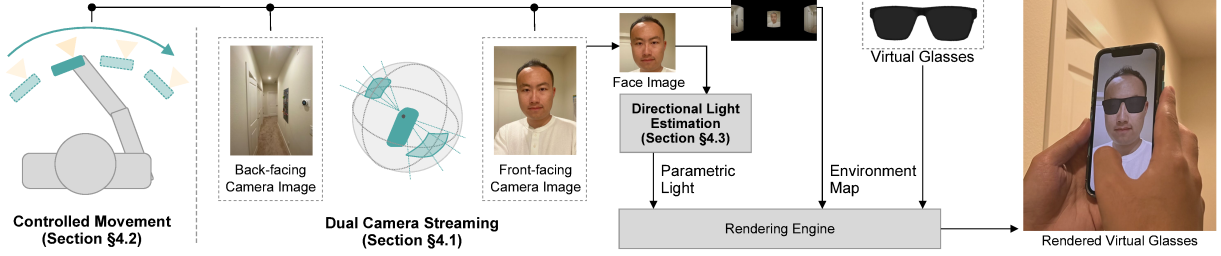

Figure 4: Example workflow. Our proposed system fuses dual-camera images to generate high-quality environment maps (§4.1) by leveraging a controlled movement pattern (§4.2) to stitch multi-view images and a parametric directional lighting model to improve the environment map's completeness and support correct color tone and shadow (§4.3).

- **Q3:** How to leverage learning-based models to estimate out-of-camera views and support shadow casting? The ideal solution should work nicely with dual-camera LDR image inputs.

4.1 Dual-Camera Streaming Policies

Based on our preliminary investigation that dual-camera streaming can increase the observation, but at the expense of mobile energy (§3), we propose to regulate the back-facing camera usage by *time and mobility* to conserve energy. Our proposed two policies are extended from our recent work LitAR's noise-tolerant data capturing policies for back-facing AR [26] and are inspired by the adaptive policy from Lai et al. for deblurring [13].

First, we propose to use a simple timer-based policy to turn on and off the back-facing camera to capture potential environmental changes. The timer starts at a larger interval w, e.g., more than the rendering interval of 33 ms, and will be adjusted based on observed changes. Once the back-facing camera is on, we will stream for w before turning it off to mitigate the tail energy phenomenon [10]. We will decrease the timer interval by half if we observe sufficient environmental changes, indicating a more dynamic scene, over the capturing window. We could effectively determine the change between two consecutive images with metrics such as SSIM or PSNR. If little change is observed, we will gradually increase the timer interval. This policy is similar to the additive increase/multiplicative decrease algorithm.

Second, we propose to leverage the inertial measurement unit (IMU) sensor data to trigger *early capturing* when significant device movement is detected. To avoid capturing images with an undesirable visual artifact, e.g., motion blur, the back-facing camera will only start streaming once the mobile device is stable. Every motion-triggered capturing will reset the timer.

4.2 LDR Environment Map Generation

Obtaining 360° HDR images for visually coherent lighting rendering is a challenging task for mobile AR. Even with the dual-camera streaming strategies described above, obtaining the 360° HDR environment map is still challenging because modern mobile cameras only have limited pixel intensity range support. Consequently, we propose to use LDR environment maps and parametric directional lighting model (§4.3) to approximate the HDR environment maps. This section describes our proposal to leverage guided user movement to combine multi-view images (both front-facing and back-facing) to generate LDR environment maps.

Image-Based Stitching Via Face Tracking. We propose a simple and lightweight component to exploit device mobility and the dual-camera streaming feature to obtain high-quality device pose tracking data and generate high-quality environment maps. The key idea is to use a controlled movement pattern and leverage the user's faces in front-facing camera images to establish *a shared anchor point*, which can then be used as the reference for tracking device poses across frames.

Our design is driven by the advancement of 3D face modeling research. Using face geometry and camera parameters estimation models, e.g. MediaPipe[6] face mesh [7], we could obtain the relative pose of the camera to the user's face. To combine multi-frame camera images, we propose to use the user's face as a shared anchor point by asking the user to fix their head pose during capturing.

---

[6]https://mediapipe.dev

To ease the user experiences, we propose to design an onscreen instruction for guiding the users to keep their head static while moving the mobile device along the trajectory of the great circle (with the radius of the user's arm length)[7]. Figure 4 illustrates the idea of controlled movement. We only ask the user to perform this guided movement at the app launch time and when significant movement is detected while using the app.

Stitching Image Views for Environment Maps. To leverage the additional environment observations for improving reflection rendering quality, we first stitch the front and back camera images into a partial 360° environment map. Specifically, we assume the front-facing and back-facing cameras share the same origin and map the camera image pixels to an environment map based on camera physical parameters and the pinhole camera model [22]. To bootstrap the user experiences when the AR application starts, we initialize the environment map data from the guided control movement at the beginning of the try-on app. During application usage, we progressively update the environment map using image feature matching methods, e.g., [9], and merge new back-facing camera images into the reconstructed environment map. For front-facing camera image, as the user's head pose and facial expressions might change during the app usage, we propose to only use the current frame for environment map reconstruction.

Analysis of Environment Observation. We present an analysis of the environment observation increase when combing multi-frame images. To test the observation increase, we manually created a setup in Blender with 3 back-facing camera images, each with 120° FoV. We set the image pose along a horizontal movement trajectory to simulate a ±30° movement. We used the same observation coverage measurement metric as in §3 and found that the environment observation can be increased to cover 50% directions, 1.3x of using one back-facing camera image alone.

4.3 Parametric Directional Light Estimation

HDR environment maps are commonly used for rendering shadows of virtual objects. However, directly reconstructing 360° HDR environment maps requires prolonged user engagement for scanning and can be prone to quality issues. Furthermore, our prior work shows that using learning-based models to inpaint camera views to 360° HDR environment maps can be highly inaccurate for complex real-world indoor scenes [26]. To address these limitations, we propose to approximate the HDR environment maps by combining reconstructed LDR environment maps and estimated parametric directional lights. At a high level, our proposed directional light estimation model will work in tandem with the environment map reconstruction pipeline described in previous sections. The estimation model will improve the prediction of the color tone and the completeness of the environment map, as well as supporting shadow rendering.

Specifically, we propose to extend the estimation model [14] to use the back-facing camera images to improve the lighting color estimation accuracy. We expect that this design will improve the estimation accuracy of the environment color even when unseen

objects or unnatural environmental lighting appear in the image. This is because back-facing camera images greatly enriches the environment lighting information in the input data, leading to more accurate light color estimation. Besides parametric directional lights, our estimation model also outputs a low-resolution environment map. We plan to use the estimated environment map to fill unseen regions on the reconstructed environment map. This design will further improve the reflection rendering and overall color tone accuracy. As the reconstructed environment map is refined progressively during AR app usage, we expect the completeness of the environment map and directional light color to also be improved.

5 RELATED WORK

Lighting Estimation. Recent deep learning-based work has investigated recovering lighting information in both outdoor and indoor scenes [6], leverages images of arbitrary scenes, portraits [14], and shadows [16]. Although DL-based approaches can often deliver good visual appearances, they require training on large datasets with the lighting ground truth and can have a longer than hundreds of milliseconds inference time [24]. Prior mobile-oriented work, including our own [25, 26], addressed the above two issues by either forgoing the DL route [19, 26] or exploring the advancement in the mobile camera system for the back-facing mobile AR applications [2, 21, 25]. Our proposed work combines a promising set of existing lighting estimation techniques to deliver tailored experiences for mobile AR try-on applications.

Virtual Try-On Applications. Recent work on virtual try-on has demonstrated good progress in overlaying clothes on a human body with little deformation and correct occlusion [17]. Earlier work on image-based try-on takes two inputs, images of a person and the desired clothes, and generates a new image where the same person is now wearing the desired clothes [8]. Other fashion items, such as shoes, cosmetics, and glasses, have recently received more attention [12]. For example, ARShoe proposed a real-time mobile AR-based shoe try-on system [1] that improved upon PIVTONS [3]. Despite the growing interest in virtual try-on applications, very few works address the visual coherency of the rendered fashion items, which is the focus of our work.

Multi-Camera Mobile System. To address the limitation of low FoV associated with a single camera, several works have proposed to fuse the inputs from multiple cameras to improve the quality of lighting estimation. GLEAM supported an optional feature that allows different mobile devices (hence cameras) to share lighting information [19]. Li et al. proposed to use a 360° panoramic stereo installed in the targeted indoor scene to estimate the scene properties, including lighting [15]. Multi-camera systems have also been proposed for other tasks, such as removing face blurs [13] and estimating depth [23]. Our proposed work shares the key idea of leveraging multiple cameras to increase the scene observations but tackles three unsolved challenges for front-facing mobile AR.

6 CONCLUSION

Lighting estimation has attracted a lot of attention over the past few years to improve the realism and visual details of AR apps [5,

---

[7]Guided movement is a common design for mobile applications such as taking panorama [11].

6, 24, 25]. In this paper, we aim to bring photorealism to front-facing mobile AR try-on apps, a promising method for empowering consumers with a try-before-purchase experience in e-commerce. Our proposed techniques take inspiration from our prior work in lighting estimation [14, 15, 24, 25] and will address the unique challenges in front-facing mobile AR. Two baselines we plan to compare against include dual-camera LitAR and EnvMapnet, variants based on their single-camera designs [21, 26]. We expect the first baseline to generate lower-quality environment maps and does not support shadow and the second baseline to output environment maps that do not match the physical environments.

We believe the key idea of employing multi-camera can open many interesting directions for mobile AR applications. This paper focuses on a specific case of front-facing and back-facing dual-camera streaming to improve lighting understanding. Our proposed techniques could also benefit other deployment scenarios, such as dual back-facing cameras and edge-based 360° cameras. As the number of cameras increases, obtaining high-quality environment maps will become easier via coordinated streaming. However, questions regarding the energy and quality trade-offs remain unclear and need to be studied further.

ACKNOWLEDGMENTS

REFERENCES

[1] Shan An, Guangfu Che, Jinghao Guo, Haogang Zhu, Junjie Ye, Fangru Zhou, Zhaoqi Zhu, Dong Wei, Aishan Liu, and Wei Zhang. 2021. ARShoe: Real-Time Augmented Reality Shoe Try-on System on Smartphones. In *Proceedings of the 29th ACM International Conference on Multimedia*. 1111–1119.

[2] Dachuan Cheng, Jian Shi, Yanyun Chen, Xiaoming Deng, and Xiaopeng Zhang. 2018. Learning scene illumination by pairwise photos from rear and front mobile cameras. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 213–221.

[3] Chao-Te Chou, Cheng-Han Lee, Kaipeng Zhang, Hu-Cheng Lee, and Winston H Hsu. 2018. Pivtons: Pose invariant virtual try-on shoe with conditional image completion. In *Asian Conference on Computer Vision*. Springer, 654–668.

[4] Paul Debevec. 2006. Image-based lighting. In *ACM SIGGRAPH 2006 Courses*. 4–es.

[5] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. 2017. Learning to Predict Indoor Illumination from a Single Image. *ACM Transactions on Graphics* (2017).

[6] Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde. 2019. Fast Spatially-Varying Indoor Lighting Estimation. *CVPR* (2019).

[7] Ivan Grishchenko, Artsiom Ablavatski, Yury Kartynnik, Karthik Raveendran, and Matthias Grundmann. 2020. Attention mesh: High-fidelity face mesh prediction in real-time. *arXiv preprint arXiv:2006.10962* (2020).

[8] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S Davis. 2018. Viton: An image-based virtual try-on network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7543–7552.

[9] Chris Harris, Mike Stephens, et al. 1988. A combined corner and edge detector. In *Alvey vision conference*, Vol. 15. Citeseer, 10–5244.

[10] Junxian Huang, Feng Qian, Alexandre Gerber, Z Morley Mao, Subhabrata Sen, and Oliver Spatscheck. 2012. A close examination of performance and power characteristics of 4G LTE networks. In *Proceedings of the 10th international conference on Mobile systems, applications, and services (MobiSys '12)*. Association for Computing Machinery, New York, NY, USA, 225–238.

[11] Peter Kán. 2015. Interactive HDR Environment Map Capturing on Mobile Devices.. In *Eurographics (Short Papers)*. 29–32.

[12] Jiyeon Kim and Sandra Forsythe. 2008. Adoption of virtual try-on technology for online apparel shopping. *Journal of interactive marketing* 22, 2 (2008), 45–59.

[13] Wei-Sheng Lai, Yichang Shih, Lun-Cheng Chu, Xiaotong Wu, Sung-Fang Tsai, Michael Krainin, Deqing Sun, and Chia-Kai Liang. 2022. Face Deblurring Using Dual Camera Fusion on Mobile Phones. *ACM Trans. Graph.* 41, 4, Article 148 (jul 2022), 16 pages. https://doi.org/10.1145/3528223.3530131

[14] Chloe LeGendre, Wan-Chun Ma, Rohit Pandey, Sean Fanello, Christoph Rhemann, Jason Dourgarian, Jay Busch, and Paul Debevec. 2020. Learning Illumination from Diverse Portraits. In *SIGGRAPH Asia 2020 Technical Communications* (Virtual Event, Republic of Korea) *(SA '20)*. Association for Computing Machinery, New York, NY, USA, Article 7, 4 pages. https://doi.org/10.1145/3410700.3425432

[15] Junxuan Li, Hongdong Li, and Yasuyuki Matsushita. 2021. Lighting, reflectance and geometry estimation from 360 panoramic stereo. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 10586–10595.

[16] Daquan Liu, Chengjiang Long, Hongpan Zhang, Hanning Yu, Xinzhi Dong, and Chunxia Xiao. 2020. Arshadowgan: Shadow generative adversarial network for augmented reality in single light scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8139–8148.

[17] Yuwei Meng, Pik Yin Mok, and Xiaogang Jin. 2010. Interactive virtual try-on clothing design systems. *Computer-Aided Design* 42, 4 (2010), 310–321.

[18] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. 2009. A 3D Face Model for Pose and Illumination Invariant Face Recognition. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. 296–301. https://doi.org/10.1109/AVSS.2009.58

[19] Siddhant Prakash, Alireza Bahremand, Linda D Nguyen, and Robert LiKamWa. 2019. Gleam: An illumination estimation framework for real-time photorealistic augmented reality on mobile devices. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 142–154.

[20] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. 2010. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann.

[21] Gowri Somanath and Daniel Kurz. 2021. HDR Environment Map Estimation for Real-Time Augmented Reality. https://arxiv.org/pdf/2011.10687.pdf

[22] Wikipedia. 2022. *Pinhole Camera*. Retrieved Oct 12, 2022 from https://en.wikipedia.org/wiki/Pinhole_camera

[23] Yinda Zhang, Neal Wadhwa, Sergio Orts-Escolano, Christian Häne, Sean Fanello, and Rahul Garg. 2020. Du 2 net: Learning depth estimation from dual-cameras and dual-pixels. In *European Conference on Computer Vision*. Springer, 582–598.

[24] Yiqin Zhao and Tian Guo. 2020. Pointar: Efficient lighting estimation for mobile augmented reality. In *European Conference on Computer Vision*. Springer, 678–693.

[25] Yiqin Zhao and Tian Guo. 2021. Xihe: A 3D Vision-Based Lighting Estimation Framework for Mobile Augmented Reality. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '21)*. 28–40.

[26] Yiqin Zhao, Chongyang Ma, Haibin Huang, and Tian Guo. 2022. LITAR: Visually Coherent Lighting for Mobile Augmented Reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–29.