MoiréPose: Ultra High Precision Camera-to-Screen Pose Estimation based on Moiré Pattern

Jingyi Ning[1], Lei Xie[1*], Yi Li[1], Yingying Chen[2], Yanling Bu[1], Baoliu Ye[1], Sanglu Lu[1]
[1]State Key Laboratory for Novel Software Technology, Nanjing University, China
[2]Wireless Information Network Laboratory, Rutgers University, USA
Email: ningjy@smail.nju.edu.cn, lxie@nju.edu.cn, yili@smail.nju.edu.cn, yingche@scarletmail.rutgers.edu,
yanling@smail.nju.edu.cn, {yebl, sanglu}@nju.edu.cn

ABSTRACT

Camera tracking has become a key technology for various application scenarios, especially for AR-based camera-to-screen interaction. Demand for subtle motion detection in camera tracking makes it essential to explore the six degrees of freedom (6-DoF) pose detection with ultra-high precision. In this paper, we propose a novel sensing method MoiréPose to achieve ultra-high precision on the camera's 6-DoF pose estimation. The purpose of MoiréPose is to derive the camera's 3-DoF position and 3-DoF posture relative to the screen according to the captured moiré pattern, which is produced by the superposition of the camera's Color Filter Array (CFA) and the screen raster projected onto the CFA layer. Based on moiré pattern's high sensitivity to 6-DoF pose movement and robustness to the environmental interference in the frequency domain, we propose a spectrogram-based method to realize the camera's 6-DoF detection with ultra-high precision. Moreover, we propose a thumbnail-based method to effectively extend the working range of MoiréPose, so as to realize pervasive camera-to-screen interaction. We have implemented a prototype system and evaluate the performance in real-world environments. Extensive experiment results show that MoiréPose achieves an average position error of 7.5mm and an overall posture error of 1.66°.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Ubiquitous and mobile computing**;

KEYWORDS

moiré pattern, pose estimation, 6-DoF, frequency analysis

ACM Reference Format:
Jingyi Ning, Lei Xie, Yi Li, Yingying Chen, Yanling Bu, Baoliu Ye, Sanglu Lu. 2022. MoiréPose: Ultra High Precision Camera-to-Screen Pose Estimation based on Moiré Pattern. In *The 28th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '22), October 17–21, 2022, Sydney, NSW, Australia.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3495243.3560537

1 INTRODUCTION

Nowadays, camera tracking has become a key technology in human-computer interaction (HCI) [59], industrial robot positioning [49], post production [31], etc. A typical scenario based on camera tracking is the screen interaction for extensive augmented reality (AR) applications [10, 50, 52] and motion sensing games [12, 36]. The user's headset or handles are usually embedded with cameras, while the laser base stations need to accurately reconstruct the six degrees of freedom (6-DoF) pose of the camera so as to provide an immersive experience [36], especially for aiming, shooting, steering, and other actions that require ultra-high precision. Such demand for subtle motion detection [7, 27] makes it essential to explore the 6-DoF pose detection with higher precision.

Existing technologies for the 6-DoF pose detection of wearable devices mainly include Inertial Measurement Unit (IMU)-based methods [35, 43, 45, 53, 54, 60, 62] and vision-based methods [9, 15, 24, 28, 32, 40, 55]. IMU is an electronic device that detects linear acceleration with accelerometers and rotational rate with gyroscopes. IMU-based methods suffer from integration drift [48]. Even a small measurement error can be accumulated over time and result in significant deviations. For the state-of-the-art vision-based methods, the essence of their positioning mechanism is to analyze the image in the *spatial domain*. Thus, the accuracy of these methods is limited to the *pixel level*. Even though such precision is sufficient to handle conventional interactions, it is far from enough to achieve 6-DoF pose detection for subtle motions.

In this paper, we propose a novel sensing method MoiréPose to achieve ultra-high precision on the camera's 6-DoF pose estimation, including camera's 3-DoF position and 3-DoF posture relative to the screen, as shown in Fig. 1. Specifically, the pose detection result is relative to the intersection point between the camera's optical axis and the screen plane. We define this intersection point as the screen's *virtual origin* O so as to distinguish it from the screen's *physical origin* O' in Fig. 1. The virtual origin is dynamic in different images and it refers to the screen point that is projected to the center of the image. Thus, for each screen image with moiré patterns, MoiréPose can output a camera's 6-DoF pose relative to the screen point projected at the image center. This image-based pose estimation mechanism is like an interface that can be used to implement extensive applications such as continuous 6-DoF tracking.

MoiréPose is based on the following findings. When we use a camera to shoot a screen, irregular splines often appear in the screen area of the image, and the splines change sensitively with the camera pose. These splines, called *moiré pattern*, is produced by the superposition of the camera's Color Filter Array (CFA) [2] and the screen raster projected onto the CFA layer. We call the CFA and the

---

* Lei Xie is the corresponding author.

ACM MobiCom '22, October 17–21, 2022, Sydney, NSW, Australia

ACM ISBN 978-1-4503-9181-8/22/10…$15.00
https://doi.org/10.1145/3495243.3560537

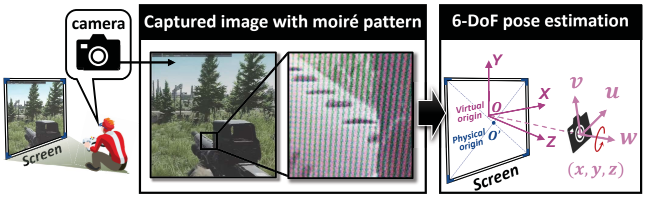


Figure 1: Illustration of MoiréPose: **In the screen coordinate system, the 3-DoF position is defined as the coordinates (x, y, z) of the camera, while the 3-DoF posture is defined as the three space vectors $[\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}]$ of the camera's three axes.**

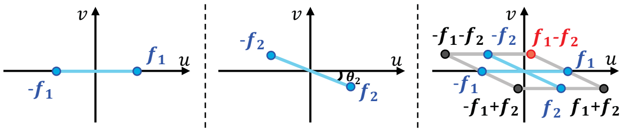
projective screen as *gratings* since they are both formed by regular and periodic pixels. The frequency of CFA and the projective screen raster can reach 500 cycles per millimeter, which are "invisible" for both human visual systems and computers. According to the principle of moiré pattern [6], the frequency of the superimposed moiré pattern is calculated by the *frequency difference* between two gratings with high and close frequencies. Thus, the frequency of moiré pattern can be as low as $10 \sim 50$ cycles per millimeter on the CFA layer, corresponding to the spline width of $10 \sim 50$ pixels in the image, which is clear enough for both human visual systems and computer vision. *Based on the low-frequency property of moiré patterns, the projective screen pixels on the CFA layer can be scaled up from "invisible" micron level to clear spline level for discrimination.* Furthermore, limited feature points, *e.g.*, corner points of the visual marker, used in traditional vision-based methods can be easily affected by external environments, while the splines captured in the moiré-based method repeat abundantly in the spatial domain. *Therefore, the splines' statistical features in the frequency domain can provide more accurate and robust information for pose estimation than feature points used in the vision-based approaches, which only contain limited numbers and are easily affected by the environments.*

Based on the above understanding, the moiré-based mechanism can achieve an ultra-high precision due to moiré pattern's low-frequency property and statistical features in the frequency domain. Accordingly, we adopt frequency analysis to extract moiré features, including the *spatial frequency* and *propagation direction*. To realize the pose estimation, we derive the mathematical relationship between the moiré features and camera's 6-DoF pose, as shown in Fig. 1. For the position estimation, we determine the camera coordinates relative to the screen's virtual origin by developing a new ranging-based model. For the posture estimation, we determine the three posture vectors of the camera in the screen coordinate system by designing a novel roll-angle-based model.

There are three key challenges in this paper. *The first challenge is to ensure both high precision and robustness of the moiré features.* Specifically, recognition of moiré pattern in the spatial domain can be significantly affected by lighting conditions and the screen's pixel geometry [56], including dot, stripes, and triangular arrangement. To address this challenge, we transfer the moiré pattern from the spatial to the frequency domain for analysis. Interestingly, we find that the detailed moiré shape and original screen contents are both transformed into frequency noises, and only the statistical features reflecting the inherent moiré features are highlighted in the spectrogram. Therefore, we propose to adopt frequency analysis to extract robust and accurate moiré features without the interference of spatial noises. *The second challenge is to precisely explore the mathematical relationship between moiré features and the camera's 6-DoF pose relative to the screen.* To estimate the camera position, we find that moiré pattern's frequency in the spatial domain is extremely sensitive to the distance between the camera and the screen. Thus, we propose a ranging-based model to first select several points of interest (POI) on the screen and then calculate the distances (denoted as "distance array") from the camera to each POI on the screen. Based on the distance array, we can deduce the camera position according to an optimized multilateral positioning method. To estimate the camera posture, we find that the rotation of the camera along the optical axis can significantly impact the *propagation direction* of moiré pattern. Thus, we propose a roll-angle-based method to estimate rotation angle along the optical axis $\boldsymbol{w}$ as shown in Fig. 1, and determine the overall posture axes of the camera. *The third challenge is to extend the effective working range of moiré pattern.* The low-frequency moiré pattern can be generated only when the frequencies of the CFA grating and the projective screen grating are close. Once the camera gets much closer or farther to the screen, the frequency difference between the projective screen and CFA, *i.e.*, moiré pattern's frequency, becomes larger. It is hard for the camera to resolve such high frequency since multiple splines are mapped to one camera pixel. To address this challenge, we propose a novel thumbnail-based method to reduce the high frequency of moiré patterns captured at inappropriate (*i.e.*, either too close or too far) distances. Specifically, we sample the high-frequency moiré pattern with an iterative thumbnail ratio (*i.e.*, down-sampling ratio) until we can obtain a clear thumbnail image with low-frequency moiré patterns. Based on our derived thumbnail ratio and the clear moiré patterns in the thumbnail image, we can restore the accurate moiré features of the original high-frequency moiré patterns even at some improper distances. By solving the above three challenges, we are able to realize the camera's 6-DoF detection with ultra-high precision in pervasive camera-to-screen interaction scenarios.

We make three key contributions in this paper. *First, to the best of our knowledge, this is the first work to thoroughly investigate the moiré localization mechanism based on optical devices and apply it to sense the 6-DoF pose of the camera.* We develop an ultra-high precision localization mechanism by extending the precision from conventional *pixel level* to the *spline level* based on the low-frequency property of moiré pattern. Furthermore, we propose a spectrogram-based method by extending limited feature points in the *spatial domain* to the statistical and robust features in the *frequency domain*. It can significantly reduce the noises in the spatial domain caused by ambient light, screen content, and device types. *Second, we derive a mathematical model to associate the 6-DoF camera pose with the extracted moiré features and effectively extend the working range of the moiré-based mechanism.* We derive the impact of the camera pose on the moiré features, and propose a ranging-based model and a roll-angle-based model to determine the camera position and posture, respectively. Furthermore, to ensure the pervasiveness of moiré features, we propose a thumbnail-based method to generate clear moiré patterns and extract accurate moiré features even at some improper distances. *Third, we implement a prototype system of MoiréPose and evaluate the performance in real-world environments. Extensive experiment results show that our system achieves an average position error of* 7.5mm *and an overall posture error of* 1.66°.

(a) Grating g_1 (b) Grating g_2 (c) Grating g_m



(d) Spectrum G_1 (e) Spectrum G_2 (f) Spectrum G_m

Figure 2: Moiré pattern, moiré feature, and spectrum

2 RELATED WORK

Vision-based pose estimation: Existing vision-based technologies for camera's 6-DoF pose estimation mainly include Perspective-n-Point (PnP)-based methods [9, 15, 24, 55] and deep-learning-based methods [28, 32, 40]. PnP-based methods [9, 15, 24, 55] estimate the 6-DoF pose of the camera by finding a scaling and deformed relationship of the feature points from the 3D world to the 2D image. These solutions rely on the priori of markers' location and layout in the 3D world. Once the detection space changes, both the procedures of deployment and measurement must be done again. Deep-learning-based solutions [28, 32, 40] leverage trained neural network models to predict the pose in a given picture, but they suffer from highly complex model training and weak generalization. Furthermore, a critical problem of vision-based solutions is the difficulty in recognizing subtle pose changes, especially when the target is in front of the markers at a far distance. *Whereas, MoiréPose can make it since moiré pattern's low-frequency property has the ability to amplify the subtle pose changes from conventional pixel level into the spline level.*

IMU-based motion tracking: IMU has become an increasingly popular COTS device for motion tracking, including indoor localization [43, 54] and gesture recognition [35, 45, 53, 60, 62]. However, IMU-based methods usually suffer from integration drift since they need to integrate acceleration to calculate velocity and position [48]. Even a small measurement error can be accumulated over time to lead to a significant deviation. In addition, the fatal problem of IMU-based methods is that the very nature of inertial measurement makes accelerometers and gyroscopes insensitive for measuring subtle and slow movement.

Moiré-based solutions: With the popularity of various digital screens, many researchers have focused on eliminating the moiré pattern in screens to provide users with a clear viewing experience [13, 23, 29]. Furthermore, some researchers try to extract meaningful information from the screen's moiré patterns for perception [14, 19, 34] rather than mopping them. Pan *et al.* [34] first exploit the nonlinearity of spatial frequency in camera systems to realize a secure and robust QR code communication. Cheng *et al.* [14] leverage the moiré effect between the CFA and a crafted mID encoding on the screen to realize a new watermark-like technique for digital forensics. These state-of-the-art works provide the possibility of moiré pattern for sensing. However, they can not be applied to camera-to-screen scenarios for deriving pose-related information. In terms of detecting the camera's pose, some state-of-the-art works propose to sense the camera position via moiré mechanism [51, 58]. However, these solutions are not suitable for camera-to-screen scenarios since they both rely on customized moiré-based markers, and they only focus on the camera position estimation without posture detection. Tanaka [51] leverages a customized LentiMark to sense the changes of the camera position in two dimensions. Xiao *et al.* [58] realizes 3-DOF camera position estimation with the aid of marker-based method. It relies on a customized two-layer MoiréBoard, consisting of a 3D printing layer and a display layer. It will be intrusive for the camera-to-screen interaction. *Different from the previous work, MoiréPose is capable of sensing 6-DoF pose of the camera relative to the screen with a non-intrusive manner based on the inherent property of screens and CFA.*

3 PRELIMINARY

Moiré pattern is produced by the superposition of two periodic gratings in two-dimensional space. Consider two simple gratings with a cosine profile function as shown in Fig. 2(a) and Fig. 2(b), which are given by the following function:

$$g_i(x,y)=\frac{1}{2}+\frac{1}{2}\cos(2\pi f_i(x\cos\theta_i+y\sin\theta_i)),\, i\in\{1,2\}, \tag{1}$$

where the function $g(x,y)$ represent the color intensity at the coordinate (x,y) on the 2D image plane. The angle θ indicates the *propagation direction* of the grating's frequency, which is orthogonal to the inherent direction of the grating, as shown in Fig. 2. Particular, the width of the grating in one period is denoted by T, and the *spatial frequency* f of the grating can be calculated by $f=\frac{1}{T}$. To simplify the presentation, we use the vector $\mathbf{s}=[x,y]^{\mathrm{T}}$ to substitute the 2D coordinates and use $\mathbf{f}_i=[f_i\cos\theta_i, f_i\sin\theta_i]^{\mathrm{T}}$, $i\in\{1,2\}$, to represent the frequency vector [8]. According to the moiré principle [6], the superposition of two gratings is equivalent to the multiplication operation. Thus, the superposition $g_m(\mathbf{s})$ illustrated in Fig. 2(c) can be calculated by $g_1(\mathbf{s})\times g_2(\mathbf{s})$ and rewritten as the following equation with newly generated frequency vectors $\mathbf{f}_1-\mathbf{f}_2$ and $\mathbf{f}_1+\mathbf{f}_2$:

$$\begin{aligned} g_m(\mathbf{s}) =& \frac{1}{4}+\frac{1}{4}\cos(2\pi\mathbf{f}_1\cdot\mathbf{s})+\frac{1}{4}\cos(2\pi\mathbf{f}_2\cdot\mathbf{s})+ \\ & \frac{1}{8}\cos(2\pi(\mathbf{f}_1-\mathbf{f}_2)\cdot\mathbf{s})+\frac{1}{8}\cos(2\pi(\mathbf{f}_1+\mathbf{f}_2)\cdot\mathbf{s})\,. \end{aligned} \tag{2}$$

This phenomenon can be easily explained from the frequency domain. Specifically, we show the theoretical Fast Fourier Transform (FFT) results G_1, G_2, and G_m in Fig. 2(d) ∼ Fig. 2(f). We can observe four new frequency vectors in Fig. 2(f) due to the convolution of the frequency vectors $\mathbf{f}_1$ and $\mathbf{f}_2$. Note that the length and direction of the frequency vector $\mathbf{f}_i$ respectively correspond to the *spatial frequency* and *propagation direction* of the grating g_i. Thus, the frequency vector $\mathbf{f}_1-\mathbf{f}_2$ with the lowest spatial frequency exactly corresponds to the most sparse and conspicuous grating in Fig. 2(c), *i.e.*, the moiré pattern. Therefore, the superimposed grating of moiré pattern in Fig. 2(c) can be finally represented as the following expression by reserving the lowest-frequency component and the DC component:

$$\hat{g}_m(\mathbf{s})=\frac{1}{4}+\frac{1}{8}\cos(2\pi(\mathbf{f}_1-\mathbf{f}_2)\cdot\mathbf{s})\,. \tag{3}$$

We use this simple expression to express the moiré pattern gratings in the subsequent sections of this paper.

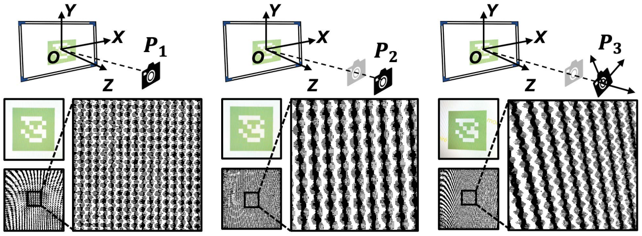


Figure 3: Sensitivity analysis of moiré pattern

4 SENSITIVITY ANALYSIS OF MOIRÉ

4.1 Empirical Study

To demonstrate the sensitivity of the moiré pattern to the relative pose between the screen and camera, we first capture an image of a 24-inch screen displaying a visual maker at a distance of 42cm (denoted as pose "P_1"). After that, we increase the camera-to-screen distance by 5mm ("P_2") and then rotate the camera along its optical axis by 5° ("P_3"). We also capture images for pose P_2 and P_3. The corresponding results are illustrated in Fig. 3. It can be observed that the profile of the visual maker at three pose locations are quite similar, while the extracted moiré pattern changes significantly even for such small changes of the camera pose. Specifically, we can have the following observation:

Observation 1: The *spatial frequency* of moiré pattern changes with the camera-to-screen distance.

Observation 2: The *propagation direction* of moiré pattern changes with the camera posture.

These observations provide us an opportunity to derive the camera pose according to the variation of moiré features.

4.2 Superiority of Moiré Pattern

Next, we conduct theoretical analysis on the sensitivity of moiré patterns. We use T_m, T_c, and T_p to represent the period lengths of the moiré pattern, CFA grating, and screen grating, *i.e.*, the inverse the spatial frequency. Since moiré pattern's spatial frequency f_m is calculated by the difference between the CFA frequency f_c and the projective screen frequency f_p, *i.e.*, $\frac{1}{T_m} = \frac{1}{T_c} - \frac{1}{T_p}$, the period length of moiré pattern T_m compared to the original camera resolution T_c can be expressed as:

$$\frac{T_m}{T_c} = \frac{1}{1 - \frac{T_c}{T_p}}, \tag{4}$$

where $\frac{T_m}{T_c}$ can be denoted as the magnification coefficient $\mathcal{X}_m$ to indicate the magnification ratio of the moiré pattern compared to original camera pixels. Thus, moiré-based method can enlarge the pixel-level feature points used in traditional vision-based methods to $\mathcal{X}_m$ times. According to our empirical experiments, the coefficient $\mathcal{X}_m$ can reach 5.13 ~ 11 times at proper interaction distances (*i.e.*, the period lengths of CFA T_c and projective screen T_p are close), which ensures the ultra-high precision of camera pose estimation even for subtle motions.

4.3 Technical Issues

Based on the above analysis, to realize the ultra-high precision pose estimation via moiré-based mechanism, we still need to deal with three critical technical issues:

1) Pose estimation model: What is the mathematical relationship between the observed moiré features and the 6-DoF pose of the camera relative to the screen?

2) Robustness of moiré features: In different environments with various lighting conditions or device types, how to extract the reliable and stable moiré features to estimate the camera pose?

3) Availability of moiré pattern: Actually, there is an effective distance range to capture clear moiré patterns, *i.e.*, when the frequencies of the CFA and projective screen gratings are close. Once the camera exceeds the distance range, how to ensure the availability of the moiré pattern?

5 MOIRÉ FEATURE EXTRACTION

To derive the camera position (x, y, z) and posture $[\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}]$ relative to the screen, we first need to extract the moiré pattern and derive effective moiré features from the image. However, the moiré pattern in the spatial domain can be easily influenced by the environment and screen types. Specifically, due to the vignetting [5] and distortion [3] of the camera lens, moiré pattern at the image edge is usually blurry and can not be used to calculate moiré features. Besides, different screen pixel geometries [56] lead to various moiré details. To deal with these noises in the spatial domain, in this section, we propose to leverage the center part of the image to derive clear moiré patterns and adopt frequency analysis to extract statistical and robust moiré features without the interference from device types with different pixel geometries.

5.1 Moiré Pattern Extraction

To extract the clear moiré pattern, we leverage a series of image processing methods to eliminate the influence of vignetting effect and enhance the contrast of the moiré pattern, as illustrated in Fig. 4. According to the refraction of the camera lens, the object point with a farther distance to the optical axis gets a more significant distortion in the captured image [3]. Based on our empirical study, the impact of camera lens' vignetting and distortion can be effectively eliminated by setting the side length of the center part to one-third of the short side of the image. Thus, we cut the marginal part with significant distortions and leave the center part with more reliable moiré patterns. Even so, the contrast of the image grayscale is still weak, and the illumination at the four corners is darker than the center area. If we directly perform binarization, some moiré patterns may be submerged in the background, especially the moiré pattern at the four corners. Therefore, we first perform the histogram equalization to enhance the contrast of the moiré pattern and then leverage the multi-scale Gaussian function to adaptively adjust the illumination at the corners based on non-linear Gamma Correction [46]. After that, we perform binarization and use a median filter to derive clear moiré patterns without significant noises and shakes. The final enhanced moiré pattern is shown in Fig. 4(d).

5.2 Feature Calculation

Based on the extracted moiré pattern, it is still difficult to find a unified method to extract moiré features from the spatial domain. It is because: 1) The detailed shape, *e.g.*, stripes, gratings, and dots in Fig. 5, may interfere with the period extraction in the spatial domain. 2) One of the moiré dimensions in the spatial domain may

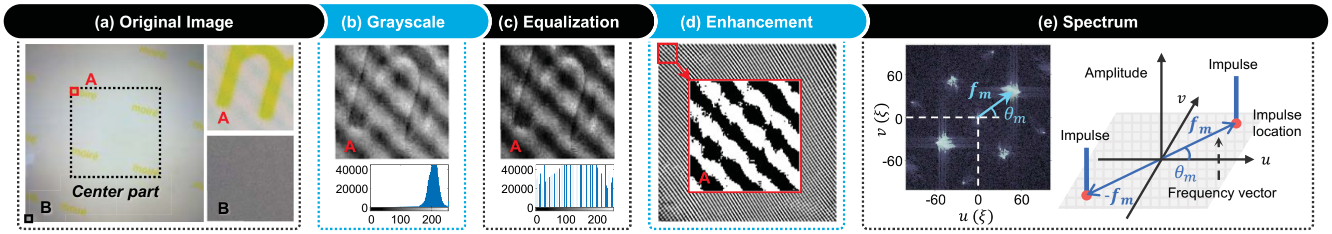


Figure 4: Process of moiré pattern extraction

be weakened by the demosaicing process [21], *i.e.*, the superposition process of CFA and the projective screen grating, as shown in Fig. 5(a). 3) The original content displayed on the screen can also affect the moiré recognition and feature extraction. The intuition to deal with these issues is that *in the frequency domain, the detailed moiré shape and the original screen content are both transformed into frequency noises in the spectrogram. Only the statistical features reflecting the spatial frequency and propagation direction of the moiré pattern can be highlighted in the spectrogram.*

Based on the above understanding, we perform the Fast Fourier Transform (FFT) for the extracted moiré patterns, and the corresponding spectrogram is shown in Fig. 4(e). Without loss of generality, the spectrum of one-dimensional period grating usually consists of three pronounced impulses, including the direct current (DC) impulse at the spectrogram origin and the other two impulses with equal brightness symmetrically distributed about the origin. The DC impulse represents the zero frequency and corresponds to the constant component in the spatial domain, while the target frequency of the moiré pattern we need to consider is the symmetric impulse pair. For each impulse, there exists a *frequency vector* $\mathbf{f}$ to connect the spectrogram origin to the geometric location of the impulse. The *frequency vector* $\mathbf{f}$ in the spectrogram can be expressed by its polar coordinates (f, θ), where θ is the direction of the impulse, *i.e.*, *propagation direction*, and the length f of the frequency vector denotes its frequency in that direction, *i.e.*, the value of the *spatial frequency*. Notably, the specific values of the spatial frequency f and the propagation direction θ can be calculated by the Cartesian coordinates values (u, v) of the impulse as:

$$f = \sqrt{u^2 + v^2}\,,\ \theta = \arctan(\frac{v}{u})\,, \tag{5}$$

where u and v denote the horizontal and vertical components of the gratings in the frequency domain. According to Eq. (5), we can calculate the moiré spatial frequency f_m and the frequency direction θ_m based on the observed value of u and v in Fig. 4(e). Here, the spatial frequency f_m represents the number of cycles within the length of the center part's side. The unit ξ in Fig. 4(e) denotes a constant to convert f_m to the unit of meter. It can be calculated by $\frac{1}{cR_m}$. R_m represents the number of pixels contained in the side length of the center part, and c (unit: meter) denotes the pixel size of the CFA. Note that the features of the weakened dimension in the spatial domain are also contained in the spectrogram. Therefore, we can obtain the moiré features in both two dimensions based on the spectrogram analysis.

6 ESTIMATION OF CAMERA POSITION

According to Observation 1 in Section 4.1, we find that moiré pattern is sensitive to the distance changes. Thus, in this section, we focus on exploring the relationship between the moiré features and the camera-to-screen distance so as to derive the 3-DoF position. Specifically, we build a ranging-based model to calculate the distances from the camera to the selected points of interest (POI) on the screen. Based on the distance array, *i.e.*, the distance from the camera to each POI on the screen, we can deduce the precise coordinates of the camera position relative to the screen's virtual origin O, which is projected to the image center.

6.1 Modeling Camera-to-Screen Distance

In this subsection, we focus on constructing a mathematical relationship between the camera-to-screen distance and the moiré features. As shown in Fig. 6, the moiré pattern is superimposed by the CFA grating and the *projective* screen grating rather than the original screen grating. Based on the observed frequency of the superimposed moiré pattern and the known frequency of the CFA grating, we can deduce the projective screen grating's frequency. Note that the projective screen's frequency contains the camera-to-screen distance information according to the pinhole camera model [38]. Specifically, the pixel width of the projective screen grating is inversely proportional to the distance, *i.e.*, the spatial frequency of the projective screen is proportional to the distance. Thus, we can deduce the camera-to-screen distance based on the pinhole camera model and the moiré features.

Specifically, we use T_c and T_s to respectively denote the width of CFA and the screen grating in one period, and their corresponding *spatial frequencies* can be denoted as $f_c = \frac{1}{T_c}$ and $f_s = \frac{1}{T_s}$. We use f_p to denote the spatial frequency of the projective screen grating. Considering that the grating frequency is directional in the spatial domain, we use vectors $\mathbf{f}_c = [f_c \cos\theta_c, f_c \sin\theta_c]^{\mathrm{T}}$, $\mathbf{f}_s = [f_s \cos\theta_s, f_s \sin\theta_s]^{\mathrm{T}}$, and $\mathbf{f}_p = [f_p \cos\theta_p, f_p \sin\theta_p]^{\mathrm{T}}$ to represent the *frequency vectors* of the CFA, screen, and the projective screen gratings. Their *propagation directions* of the frequency vectors are denoted as angles θ_c, θ_s, and θ_p.

According to the moiré principle [6], moiré pattern's frequency is equal to the frequency difference between two gratings. Thus, the moiré pattern's frequency $\mathbf{f}_m = [f_m \cos\theta_m, f_m \sin\theta_m]^{\mathrm{T}}$ can be expressed as:

$$\mathbf{f}_m = \mathbf{f}_c - \mathbf{f}_p\,. \tag{6}$$

We decompose the frequency vectors into two dimensions as:

$$\begin{bmatrix} f_m \cos\theta_m \\ f_m \sin\theta_m \end{bmatrix} = \begin{bmatrix} f_c \cos\theta_c - f_p \cos\theta_p \\ f_c \sin\theta_c - f_p \sin\theta_p \end{bmatrix}, \tag{7}$$

where f_m and θ_m denote the spatial frequency and the propagation direction of moiré pattern's frequency vector $\mathbf{f}_m$, respectively. Particularly, the spatial frequency f_p of the projective screen contains

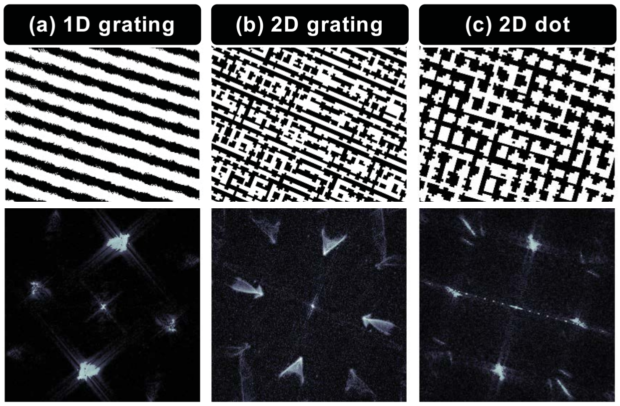


Figure 5: Impact of device types

the distance information according to the pinhole camera model as shown in Fig. 6:

$$f_p = \frac{d f_s}{f}, \tag{8}$$

where f represents the focal length of the camera, and d denotes the distance from the camera's optical center to the screen plane. Since the pixel arrangement of the screen raster is along the X-axis and Y-axis of the screen, the propagation direction θ_s of the screen's frequency vector in Eq. (7) can be set to zero without loss of generality. In addition, the projective screen's propagation direction θ_p is consistent with angle θ_s since projection does not change the direction of the screen grating. Thus, by combining Eq. (7) and Eq. (8), the camera-to-screen distance d can be calculated by:

$$d = \frac{f}{f_s}\left(\sqrt{f_c^2 - (f_m \sin\theta_m)^2} - f_m \cos\theta_m\right), \tag{9}$$

where the spatial frequency f_m and the frequency direction θ_m of moiré pattern can be measured from the captured image according to Section 5. The focal length f, spatial frequency f_c of CFA on the camera, and the spatial frequency f_s of the screen can be determined from the common COTS manuals in advance.

Notably, due to the non-directivity of the moiré pattern's splines, the observed propagation direction θ_m in Eq. (9) has two possible values that are 180 degrees apart, and it will lead to two symmetric solutions of the distance. The ambiguity of the distance can be eliminated according to the effective working range limited by the device parameters. The detailed analysis is shown in Section 9.

6.2 Position Calculation

Based on the above model, given the moiré feature of a certain point on the moiré image, we can obtain the distance from the camera to the certain point on the screen. Therefore, as long as we select appropriate points of interest (POI) on the moiré image to calculate the *distance array*, *i.e.*, distances from the camera to the corresponding POI array on the screen, the camera position can be determined via multilateral positioning algorithm. Note that the splines of moiré pattern repeat abundantly in the image, *i.e.*, the width and direction of each spline are close to those of adjacent splines. Thus, considering the time efficiency, we only need to select a POI array with the maximum position information entropy to obtain the distance array rather than calculating the distance for each pixel in the image.

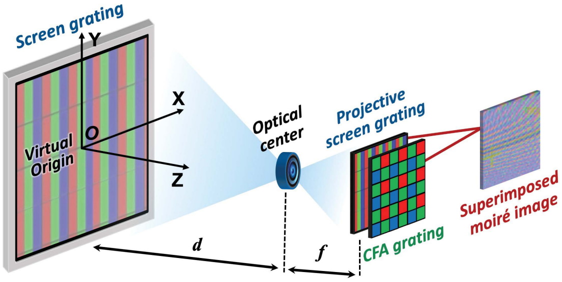


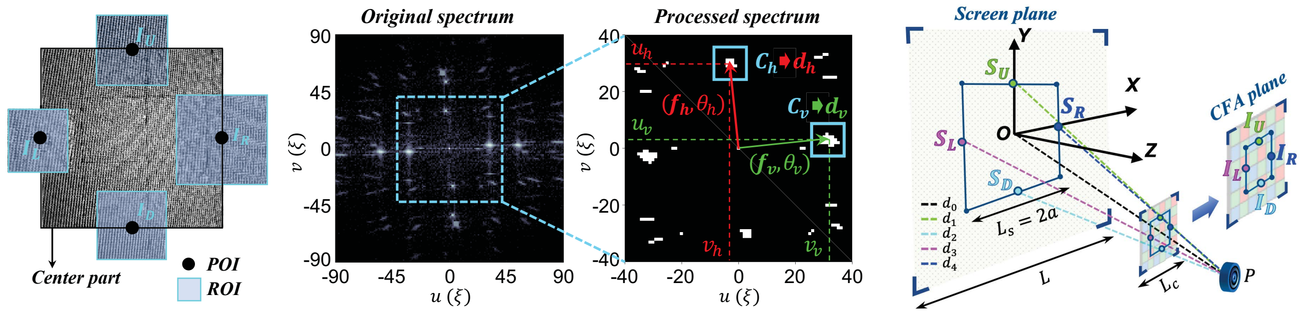
Figure 6: Modeling the distance

6.2.1 *Selection of POI and ROI.*

To ensure that the selected POI array has sufficient position information entropy, it is better to select the POI on center part's boundary dispersedly so that the POI array can cover the perspective information as much as possible. Without loss of generality, we choose points I_U, I_D, I_L, and I_R to form the POI array, as shown in Fig. 7(a). To derive each POI's accurate moiré features, we take each POI as the center and crop a region of interest (ROI) from the original image to calculate the corresponding spectrum. The ROI array is shown in Fig. 7(a). On the one hand, the ROI should be as small as possible so that the moiré feature of each pixel in one ROI can be nearly uniform. On the other hand, the ROI should be big enough to guarantee sufficient resolution of the spectrogram. To make a trade-off, we first tentatively crop the ROI for each POI according to the size of the center part, which is too large for the accurate feature extraction. For each temporal ROI, we calculate the spectrogram and derive the temporal spatial frequency f' according to Section 5.2. According to Nyquist–Shannon sampling theorem [44], to restore the spatial frequency f', the number of pixels R' needs to be greater than $2f'$. To ensure adequate fault tolerance, we set the ROI size R' for each ROI to $4f'$. Note that each ROI's size can be different since it is determined by each ROI's temporal spatial frequency f'. By resizing the ROI array, the feature calculated from each ROI can be more accurate due to the higher consistency among pixels within the smaller ROI.

6.2.2 *Distance Array Calculation via Frequency Analysis.*

Based on the determined ROI, we propose a frequency analysis method to calculate the distance array, *i.e.*, the perspective information that reflects different distances from the camera to the POI array on the screen. Specifically, we perform the FFT for each ROI. Without loss of generality, we take the POI I_R as an example to demonstrate its original and processed spectrograms, as illustrated in Fig. 7(b). The distances of the rest ROIs can be calculated in the same way. To precisely extract the impulse region in the original spectrogram, we perform an adaptive binarization [20] to adaptively isolate the impulse regions, according to the principle that the foreground is brighter than the background. The processed spectrum is plotted in Fig. 7(b). We can observe two pulse clusters C_h and C_v with the propagation directions separated by 90°. Specifically, the cluster C_h and C_v reflect the moiré features of horizontal and vertical splines, respectively. According to the method of feature extraction in Section 5.2, for each impulse cluster, we can derive a pair of



(a) POI and ROI selection (b) Spectrogram of the ROI centered on I_R (c) Ranging-based model

Figure 7: Determining the camera position

moiré features, including the spatial frequency and the propagation direction. It means that for each POI, we can obtain two distances due to the 2D property of moiré patterns. Therefore, a screening mechanism needs to be provided to eliminate the uncertainty of the distance. Ideally, the two distances calculated from 2D moiré patterns are supposed to be the same. However, the perspective angle may lead to different influences on the deformation of the 2D splines, so the distances derived from the two dimensions are different accordingly. *Based on the understanding of perspective transformation, leftwards or rightwards perspective does not affect the frequency of the horizontal splines but significantly affects the vertical splines' frequency.* Thus, for POI I_L and I_R, we leverage the horizontal splines' spatial frequency f_h and propagation direction θ_h of cluster C_h to calculate the distances d_h according to Eq. (9). In the same way, for points I_U and I_D, we leverage the vertical splines' moiré features to calculate the distance d_v according to Eq. (9). Therefore, the distance array from the camera to the POI array on the screen can be determined.

6.2.3 *Position Determination.*

Based on the derived distance array, we can determine the camera's coordinates via the multilateral positioning algorithm. We define the estimated camera position P with the coordinates of $[x, y, z]^T$. The coordinates of each POI S_i on the screen are defined as $[x_i, y_i, z_i]^T$, which can be determined in the next paragraph. Thus, the estimated distance from the position P to each POI S_i can be calculated by $\|P - S_i\|_2$. Besides, we use d_i to denote the measured distance from the position P to each POI S_i. To determine the optimal solution P^* of the position, the difference between the estimated distance array and the measured distance array should be minimum, *i.e.*:

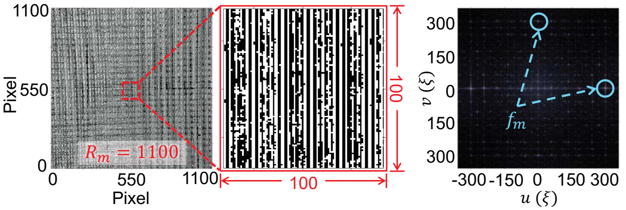
$$P^* = \arg\min_{P} \sum_{i=1}^{n} |\, \|S_i - P\|_2 - d_i \,| \, , n \geq 4 \, . \tag{10}$$

Based on this multilateral positioning method with an optimization model, we can determine the camera position according to the measured distance array and the coordinates of the screen POI array. Thus the critical problem now is to obtain the specific coordinates of the screen's POI array.

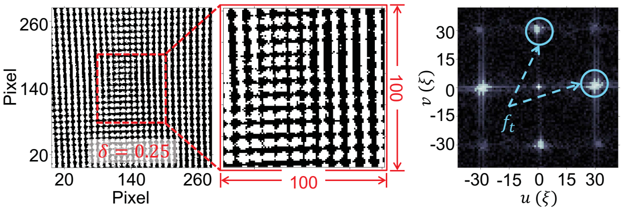
Fig. 7(c) illustrates an imaging process of the camera, and the captured image of the screen is actually on the imaging plane, *i.e.*, the CFA plane. Without loss of generality, we take the points I_U, I_D, I_L, and I_R as the POI array on the imaging plane, and the corresponding POI array on the screen can are denoted by S_U, S_D, S_L, and S_R. To realize the multilateral positioning, we also need to obtain the coordinates of the points S_U, S_D, S_L, and S_R. We assume that there exists a rectangle with the side length of $2a$, and the coordinates of points S_U, S_D, S_L, and S_R can be approximately expressed as $(0, a, 0)$, $(0, -a, 0)$, $(-a, 0, 0)$ and $(a, 0, 0)$, respectively. Next, we need to determine the specific coordinate value of a. As shown in Fig. 7(c), we use L and L_s to denote the length of the full view range on the screen plane and the center part range corresponding to the image center part. Thus, the relationship between the camera's view range L and the component parameters, including the focal length f and the size of the image sensor (approximate to the CFA size L_c) can be expressed as $\frac{L_c}{L} = \frac{f}{d_0}$, where d_0 is the distance from the camera to the screen virtual origin O. Since $L_s = 2a$ and $L = 3L_s$ (according to crop ratio of one third in Section 5.1), the coordinate value a can be estimated by $\frac{d_0 L_c}{6f}$. Therefore, the coordinates value of the POI array on the screen, *i.e.*, S_i in Eq. (10), can be determined so as to derive the optimal camera position P^*.

6.3 Effective Working Range Extension

The low-frequency property of moiré pattern is not always available. Thus, we define the effective working range as the distance range that the moiré pattern captured by the camera can be correctly resolved by MoiréPose. According to Eq. (6), only when the CFA frequency f_c and the projective screen frequency f_p are close, the frequency of the moiré pattern f_m can be small enough, *i.e.*, the period length can be resolved from the spectrogram. According to Nyquist–Shannon sampling theorem [44], to resolve the spatial frequency f_m of the moiré pattern, the resolution R_m of the moiré image needs to be greater than $2f_m$, *i.e.*, $\frac{f_m}{R_m} < 0.5$. However, according to our extensive experiments, we can only resolve the spatial frequency that meets the relationship of $\frac{f_m}{R_m} \in [0.03, 0.15]$ since the moiré splines in real scenarios are not evenly and straight. Once the camera gets much closer or farther to the screen, the moiré pattern's frequency becomes much higher. It is hard for the camera to resolve such high frequency since multiple splines are mapped to one camera pixel. Fig. 8(a) illustrates an example of a high-frequency moiré pattern extracted from the center part of an image, and the corresponding spectrogram is plotted in Fig. 8(b). It is hard to determine which impulse corresponds to the accurate frequency of the moiré pattern. To solve this problem, we propose a novel thumbnail-based method to restore the high spatial frequency



(a) High-frequency moiré pattern (b) Spectrum of moiré



(c) Thumbnail (d) Spectrum of thumbnail

Figure 8: Thumbnail-based method

of the moiré pattern. Specifically, we sample the high-frequency moiré pattern with an iterative thumbnail ratio δ until we can obtain a clear thumbnail image with low-frequency moiré patterns. The thumbnail image with the ratio $\delta = 0.25$ and the spectrogram are plotted in Fig. 8(c) and Fig. 8(d). Here, we leverage Brenner gradient function [11] to assess the clearness of the thumbnail image. We use inexact newton methods [16] to iteratively determine the thumbnail ratio δ, which corresponds to a local optimum value of the Brenner gradient function. Actually, the sampling operation is equivalent to superposing another grating with the spatial frequency of δR_m on the high-frequency moiré pattern. R_m represents the number of pixels contained in the side length of the center part. Thus, the spatial frequency f_t of the generated clear moiré patter in the thumbnail image can be calculated by $f_t = f_m - \delta R_m$. Since the spatial frequency f_t can be easily observed from Fig. 8(d), the high spatial frequency f_m of the original moiré pattern can be restored according to the following equation:

$$f_m = f_t + \delta R_m \,. \tag{11}$$

Based on the thumbnail method, we can extend the resolvable range of moiré pattern's spatial frequency from the relationship $\frac{f_m}{R_m} \in [0.03, 0.15]$ to $\frac{f_m}{R_m} \in [0.03, 0.4]$. It enable MoiréPose to extract the effective features of the moiré pattern in a larger distance range.

7 ESTIMATION OF CAMERA POSTURE

To estimate the 3-DoF posture of the camera, we need to determine the camera's three axes $[\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}]$ in the coordinate system with the virtual origin, as shown in Fig. 9. In Section 6.2.3, we have determined the camera position P relative to the virtual origin O. Since the virtual origin O is the intersection point of the camera's optical axis $\boldsymbol{w}$ and the screen plane, we can directly calculate the expression of the optical axis $\boldsymbol{w}$ as the vector $\overrightarrow{OP}$. Thus, the unit vector $\hat{\boldsymbol{w}}$ of the optical axis $\boldsymbol{w}$ can be calculated by $\hat{\boldsymbol{w}} = \frac{\overrightarrow{OP}}{\|OP\|}$. Here, we use $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$, and $\hat{\boldsymbol{w}}$ to represent the unit vectors of $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ for the sake of subsequent operations. Actually, we only need to determine any one of the axes $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$, and the final axis can be derived by the cross product of the other two axes.

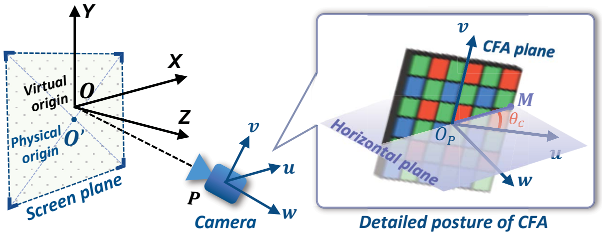


Figure 9: Determining the camera posture

According to Observation 2 in Section 4.1, the propagation direction of the moiré pattern can be easily influenced by the camera's rotation angle along the optical axis, *i.e.*, the relative changes between the frequency directions of CFA and screen gratings. Particularly, the direction of the screen grating in camera-to-screen scenarios is usually certain. It gives us an opportunity to estimate the propagation direction θ_c of the CFA. According to Eq. (7) and Eq. (9), the relationship between moiré pattern's propagation direction θ_m and CFA's propagation direction θ_c can be expressed as:

$$\tan\theta_m = \frac{\sin\theta_c}{\cos\theta_c - \frac{d_0 f_s}{f f_c}} \,, \tag{12}$$

where the propagation directions θ_m of the moiré pattern and the distance d_0 from the camera to the virtual origin O can be derived from Section 5.2 and Section 6.2.2, respectively. By solving Eq. (12), we can obtain the propagation direction θ_c of the CFA, as illustrated in Fig. 9. Here, θ_c denotes the roll angle of the camera around the optical axis, *i.e.*, the angle between axis $\boldsymbol{u}$ and axis-X's projection $\overrightarrow{O_P M}$ on the CFA plane. The target posture vector $\hat{\boldsymbol{u}}$ of the camera can be exactly derived by rotating the unit vector of $\overrightarrow{O_P M}$ around the optical axis $\hat{\boldsymbol{w}}$ by the angle of θ_c. To derive the specific posture, we denote vector $\overrightarrow{O_P M}$ as $[M_x, M_y, M_z]^T$, where M_y can be set to zero since $\overrightarrow{O_P M}$ belongs to the horizontal plane. According to the relationship that vector $\overrightarrow{O_P M}$ in the CFA plane is perpendicular to the known optical axis $\boldsymbol{w} = [w_x, w_y, w_z]^T$, we can solve the ratio between M_x and M_z as $-\frac{w_z}{w_x}$ to further determine the unit vector $\hat{\boldsymbol{m}} = [m_x, m_y, m_z]^T$ of vector $\overrightarrow{O_P M}$. Based on the efficient Rodrigues' rotation formula [39], we can calculate the unit vector $\hat{\boldsymbol{u}}$ by rotating the unit vector $\hat{\boldsymbol{m}}$ around the optical axis $\hat{\boldsymbol{w}}$:

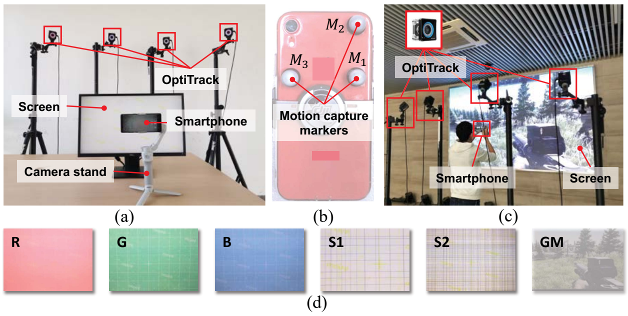
$$\hat{\boldsymbol{u}} = \hat{\boldsymbol{m}}\cos(\theta_c) + \hat{\boldsymbol{w}} \times \hat{\boldsymbol{m}}\sin(\theta_c) + (\hat{\boldsymbol{w}} \cdot \hat{\boldsymbol{m}})\hat{\boldsymbol{w}}(1 - \cos(\theta_c)) \,, \tag{13}$$

where the rotation angle θ_c can be derived from Eq. (12). Eventually, the last posture axis $\hat{\boldsymbol{v}}$ can be determined by $\hat{\boldsymbol{w}} \times \hat{\boldsymbol{u}}$.

8 PERFORMANCE EVALUATION

8.1 Implementation and Methodology

Implementation: We implement and evaluate MoiréPose in both near-field and far-field interaction scenarios to show the performance in accuracy and robustness. The setup is shown in Fig. 10. Our system is built with a screen and a smartphone embedded with cameras. Without loss of generality, we regard the smartphone as the target of interacting with the screen and perform MoiréPose to reconstruct the 6-DoF pose of the smartphone.



(a) (b) (c)

(d)

Figure 10: Experimental setup: (a) Near-field interaction scenario. (b) A smartphone with motion capture marker. (c) Far-field interaction scenario. (d) Screen contents.

Table 1: Parameters of the screen devices

ID	Manuf.	Model	Size	Pixel Size
S1	Dell	E2420H	23.8"	205.8 μm
S2	Apple	MacBook Pro 2017	13.3”	112 μm
S3	Lenovo	ThinkPad T460P	14”	161.4 μm
S4	-	large LED Screen	3.3×2.1 m^2	2.5 mm

Table 2: Parameters of the capturing devices

ID	Manuf.	Model	Focal Length	Pixel Size
C1	Samsung	Galaxy note8	4.3 mm	1.4 μm
C2	Samsung	Galaxy S20 FE	5.4 mm	1.8 μm
C3	Apple	iPhone X	4 mm	1.22 μm
C4	Apple	iPhone XR	4 mm	1.4 μm
C5	Xiaomi	Mi 10S	6.72 mm	1.6 μm
C6	Xiaomi	Redmi 10X	4.74 mm	0.8 μm
C7	Sony	Xperia XZ1	4.4 mm	1.22 μm

Table 3: Working range of camera-screen pairs

	S1	S2	S3	S4
C1	30 ~ 42 cm	12 ~ 23 cm	17 ~ 33 cm	2.7 ~ 3.7 m
C2	28 ~ 42 cm	12 ~ 22 cm	16 ~ 32 cm	2.6 ~ 3.6 m
C3	35 ~ 45 cm	17 ~ 25 cm	25 ~ 35 cm	3.4 ~ 4.0 m
C4	30 ~ 40 cm	13 ~ 22 cm	20 ~ 31 cm	2.9 ~ 3.7 m
C5	39 ~ 58 cm	16 ~ 31 cm	23 ~ 45 cm	3.5 ~ 5.2 m
C6	30 ~ 40 cm	12 ~ 22 cm	20 ~ 32 cm	2.9 ~ 3.6 m
C7	40 ~ 50 cm	18 ~ 27 cm	27 ~ 39 cm	3.7 ~ 4.4 m

Ground Truth: To obtain the ground truth of the camera pose in 3D space, we attach three motion capture markers on the smartphone and leverage OptiTrack [4] to capture the precise coordinates of the three markers, as shown in Fig. 10(b). OptiTrack realizes the tracking precision of $0.2mm$ in our experimental scenarios. We construct OptiTrack's coordinate system based on the virtual origin of the screen. Thus, the ground truth of the camera pose relative to the screen can be obtained according to three markers' coordinates. Specifically, the ground truth of the camera position can be deduced from the topology between the lens and three markers. The ground truth of camera orientation axes $\boldsymbol{u}$ and $\boldsymbol{v}$ can be calculated by marker vectors $\overrightarrow{M_1M_2}$ and $\overrightarrow{M_1M_3}$, and axis $\boldsymbol{w}$ can be derived by $\overrightarrow{M_1M_2} \times \overrightarrow{M_1M_3}$, accordingly.

Setup: 1) *Screen type*: To evaluate the impact of different screen devices, we test three LCD screens, including two laptops and one monitor, and a large LED screen, as shown in Table 1. 2) *Camera type*: To evaluate the impact of different CFA types, we select seven smartphones with different camera parameters, as shown in Table 2. 3) *Distance*: To demonstrate the effectiveness of MoiréPose in pervasive scenarios, we evaluate the system with two typical scenarios, including near-field scenarios with the distance range of 20~50 cm and far-field scenarios with 2.5~4m. Note that each screen-camera pair determines an effective working range for optimal performance. Within the effective working range, the spatial frequency of moiré pattern is relatively low and can be accurately resolved by MoiréPose. We calculate the specific working range for each screen-camera pair and adopt it in our experiment, as shown in Table 3. 4) *Orientation*: We do not strictly fix the camera's posture in our experiment. However, to capture the camera's subtle pose change in front of the screen, the angle difference between the screen plane and the CFA plane, *i.e.*, the smartphone plane, needs to be constrained within $-15^\circ \sim 15^\circ$. Otherwise, the screen may be out of the shooting range. 5) *Brightness*: We evaluate four screen brightness levels in real-world scenarios, including the brightness percentage of 100%, 70%, 40%, and 10%. 6) *Screen content*: We adopt six kinds of screen content, including three color backgrounds and three kinds of complex patterns, as shown in Fig. 10(d). 7) Comparisons: We compare MoiréPose with the checkerboard-based method [61] and the state-of-the-art solution ArUco [33] in both light and dim scenarios.

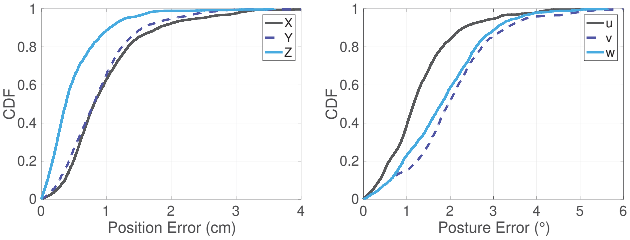


Figure 11: Overall performance of MoiréPose

Metrics: We evaluate the performance with metrics of *position error* and *orientation error*. *Position error* refers to the Euclidean distance between the estimated position and the ground truth position of the camera in 3D space. Besides, we also evaluate the distance deviation along the X, Y, and Z axes, respectively. *Orientation error* refers to the angle difference between the camera's three axes $[\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}]$ and the ground truth. Besides, we use the mean angle error (MAE) to indicate the overall orientation error.

8.2 Overall Performance

MoiréPose realizes ultra-high precision localization with an average position error of $7.5mm$ *and an overall posture error of* 1.66°. Fig. 11(a) and Fig. 11(b) show the Cumulative Distribution Function (CDF) of MoiréPose's position error and posture error with the default devices Dell E2420H display and Samsung Galaxy note8 smartphone in their effective working range.

According to Fig. 11(a), the position error along each axis is controlled within $1cm$ for about 70% of the test images. Notably, more than 93% of the test images achieves a millimeter-level precision along the Z-axis. It is because moiré pattern is extremely sensitive to the camera-to-screen distance. Specifically, since the spacing between the POIs on the screen is much smaller than the distance from the screen to the camera, the values in the distance array get

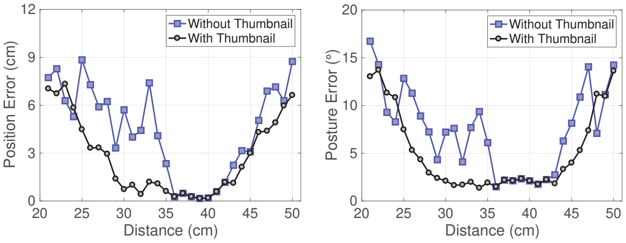


Figure 12: Impact of the distance in near-field scenarios

close. This further leads to the large position errors along the X and Y axes than the Z axis.

According to Fig. 11(b), the posture error along each camera's axis can be controlled within 2.5° for about 70% of the test images. Since the working distance of the default setting is close and the camera's optical axis $\boldsymbol{w}$ is derived from the camera position, even a small position error can lead to a large posture error for the optical axis estimation. Fortunately, the posture estimation of axis $\boldsymbol{u}$ only relies on the moiré features and achieves higher accuracy than other axes. Thus, the roll-angle-based mechanism for the camera's posture estimation can achieve good performance in total.

8.3 Impact of Camera-to-Screen Distance

To demonstrate the effectiveness of MoiréPose in pervasive scenarios, we evaluate the system in both near-field and far-field scenarios as shown in Fig. 10(a) and Fig. 10(c). Notably, different interaction distances can support different applications, and the detailed use case scenarios are discussed in Section 9.

8.3.1 Performance in Near-field Scenarios. MoiréPose can achieve millimeter-level localization precision in near-field scenarios with the thumbnail-based mechanism. According to typical near-field interactions with laptops or displays, we set the distance from 20cm ~ 50cm and select the default devices Dell E2420H display and Samsung Galaxy note8 smartphone. Fig. 12 illustrates the position and posture errors with/without adopting the thumbnail-based method in the near-field scenarios. Note that in the effective working range from $30cm \sim 42cm$, the overall position error is about $6.29mm$, and the average posture error is about 1.92°. When the working range is further getting closer, *i.e.*, $20cm \sim 30cm$, or getting farther, *i.e.*, $42cm \sim 50cm$, the performance of MoiréPose start to decrease significantly. It is because the frequency of moiré pattern at these two distance ranges becomes extremely high. Such high-frequency moiré patterns can neither be captured at far distances due to the limited resolution of the camera nor be captured at close distances due to the camera's limited focusing capacity.

8.3.2 Performance in Far-field Scenarios. MoiréPose can achieve centimeter-level localization precision in the far-field scenarios. Specifically, we set the distance range from 2.5m ~ 4m, and select four kinds of smartphones with different camera types in Table 2 to interact with the large LED screen S4 in Table 1. The detection results of the position and posture are illustrated in Fig. 13. We find that there is an optimal distance range for each camera to achieve optimal performance, and the accuracy can be different for each camera-screen pair. The optimal distance range for each camera is decided by the device parameters, including the focal length and the pixel size, *i.e.*, the frequency of the CFA grating. Specifically, Galaxy note 8 and Galaxy S20 FE perform well at the distance range of 2.7m ~ 3.5m, while Xperia XZ1 is suitable for longer interaction distance, *i.e.*, 3.75m ~ 4m or even farther.

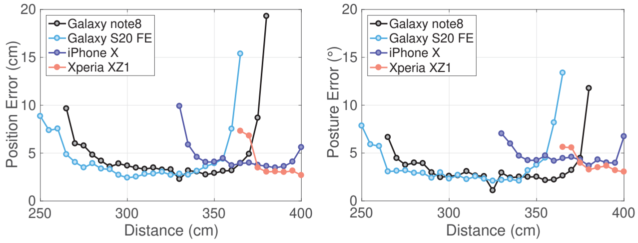


Figure 13: Impact of the distance in far-field scenarios

8.4 Impact of Screen Brightness

MoiréPose achieves fairly good performance with mutable screen brightness. We evaluate the performance in four screen brightness levels, *i.e.*, the brightness percentage is set to 100%, 70%, 40%, and 10%. Fig. 14 shows the corresponding results. Note that lower brightness may increase the instability of the estimation results, *i.e.*, the pose estimation results of 70%, 40%, and 10% brightnesses possess more extensive error ranges than the brightness of 100%. Nevertheless, dark screen brightness does not significantly change the contrast between the captured moiré pattern and the image background. The position and posture errors with 10% brightness can still reach $6.12mm$ and 1.36°, which is even lower than the errors with 100% brightness. This is attributed to the camera's adaptive adjustment for the light source of the screen.

8.5 Impact of Screen Content

Conventional patterns displayed on the screen do not significantly influence the pose estimation result. We evaluate three color backgrounds and three kinds of patterns. The specific screen contents and the experiment results are shown in Fig. 10(d) and Fig. 15. Notably, the pose error derived from the green background is the minimum. It is because green units account for the largest proportion in the CFA with a conventional Bayer filter [1], which effectively increases the sensitivity of green light. Besides, we also evaluate a game background (GM) and two stripes backgrounds (S1 and S2). Note that convention patterns, *e.g.*, the game background, can achieve even higher accuracy than a pure color background, since some curves and patterns exactly make it easier for the camera to focus. However, horizontal and vertical stripes, especially periodic and high-frequency stripes, may cause severe impact on the pose estimation, as shown by pattern "S1" and "S2" in Fig. 10(d).

8.6 Impact of Device Type

To evaluate the generalization of MoiréPose with different devices, we test four kinds of screens in Table 1. For each screen, we test seven cameras with different CFA parameters in Table 2. Notably, since the CFA pixel sizes of different cameras are almost the same order of magnitude, the key factor that determines whether the interaction distance is near-field or far-field is the screen's size. Therefore, the interaction distance from the camera to a laptop or a monitor is near-field, while the distance to the large LED

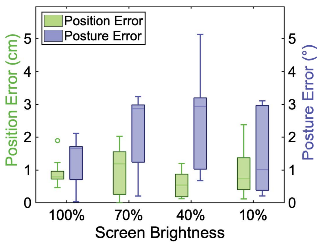


Figure 14: Impact of the screen brightness

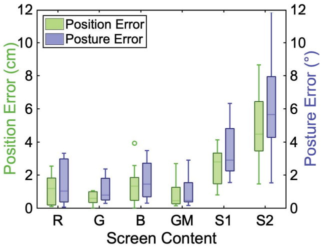


Figure 15: Impact of the screen content

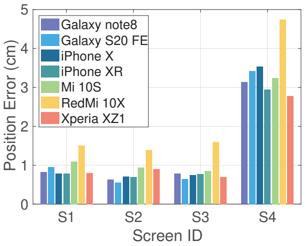


Figure 16: Impact of device types on position

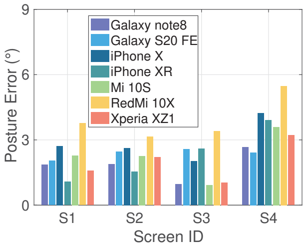


Figure 17: Impact of device types on posture

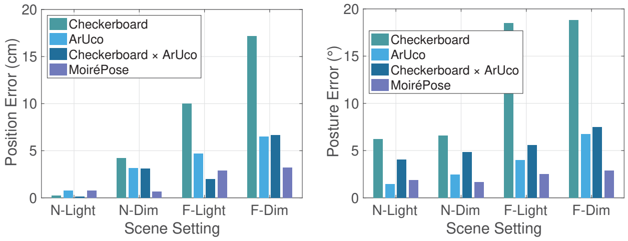


Figure 18: Performance in macro-benchmark

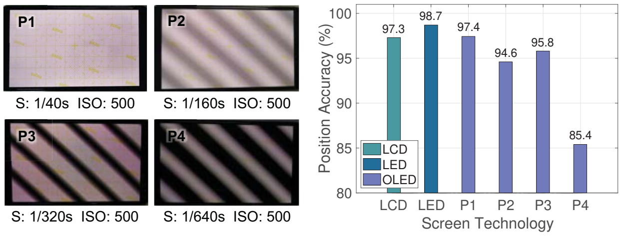


Figure 19: Impact of the screen display technology

screen is far-field. For each device pair, we evaluate the performance within the working range according to Table 3. The position and posture errors are shown in Fig. 16 and Fig. 17, respectively. The average position error in the near-field interaction, *i.e.*, S1, S2, and S3, achieves 8.87 mm, and the average posture error is about 2.14°. The average position error in the far-field interaction, *i.e.*, the large LED screen, reaches about 3.39 cm, and the average posture error is about 3.64°. Although the position error reaches centimeter-level, the ratio between the position bias and the far-filed interaction distance is still minimal.

8.7 Impact of Screen Display Technology

The mainstream screen display technology includes Liquid Crystal Display (LCD) and Organic Light Emitting Diode (OLED). Most LCD screens adopt Direct Current (DC) dimming, while OLED screens use Pulse-Width Modulation (PWM) dimming [47]. Specifically, DC dimming controls the display brightness by adjusting the electric power. This continuous luminous solution does not interfere with the production of moiré pattern. In contrast, PWM dimming varies the duty cycle of a constant current to effectively change the average current, *i.e.*, controlling the brightness by setting the time ratio of the pixel on or off. Thus, when we use a rolling shutter [57] to shoot OLED screens, stroboscope stripes may appear in the image's screen area due to the PWM modulation mode [26, 63], and the width of the stroboscope stripes is related to the modulation frequency, duty cycle, exposure time, etc.

We evaluate the performance of different display technologies, including the LCD screen (Device ID: S1), OLED screen (Device ID: C3), and Light-Emitting Diode (LED) screen (Device ID: S4). Particularly, we adopt the smartphone C3 as the typical OLED screen to evaluate the impact of the PWM dimming mode of OLED technology, which may cause stroboscope stripes in the captured image. The modulation frequency of the smartphone C3 is 120 Hz, so we adjust the shutter speed ("S" in Fig. 19.) to generate the stroboscope stripe with different widths and intensities. The stroboscope stripes and the corresponding detection result of the camera's position are shown in Fig. 19. Particularly, the working ranges of the screens with different display technologies are different. Thus, we use the ratio of the detected camera-to-screen distance to the ground truth to represent the detection accuracy. We can observe that OLED screens can achieve comparable performance to LCD and LED screens at lower shutter speeds. When the shutter speed becomes extremely high, *e.g.*, S=1/640s, the wide stroboscope stripes roll rapidly, causing the camera to lose focus. Therefore, the camera cannot capture clear moiré patterns, and the detection accuracy drops rapidly.

Fortunately, the existing modulation frequency of OLED screens can reach 60 Hz, 120 Hz, or even higher, while the frame rate of the ordinary camera is about 30 fps. The stroboscope stripes can be completely eliminated since the camera's frame rate is much lower than the screen's modulation frequency. Besides, adopting a global shutter is also an alternative solution for eliminating the stroboscope stripes.

8.8 Comparison with Other Solutions

We compare MoiréPose with a widely adopted checkerboard-based method [61] and the state-of-the-art solution ArUco [33]. To evaluate the robustness of MoiréPose and visual-marker-based methods, we adjust the ambient illumination and conduct experiments in well-lit and dimly-lit conditions, respectively. The corresponding performance results are shown in Fig. 18(a) and Fig. 18(b), including both near-filed (denoted as "N") and far-field (denoted as "F") interaction scenarios. For the near-filed interaction with well-lit illumination, MoiréPose achieves performance comparable to the advanced ArUco solution. However, for far-field interaction and dimly-lit scenarios, MoiréPose demonstrates obvious superiority over other methods. It is because long distances and dim illumination can enormously increase the detection error of feature points in the captured 2D image. Besides, when the camera is in front of the visual marker with a far interaction distance, subtle perspective transformation cannot be reconstructed precisely, since a slight change of camera posture does not significantly affect the captured marker's shape in the image. In contrast, for the moiré-based method, the low-frequency property of moiré pattern guarantees good enough performance in the far-field scenarios. The camera's adaptive adjustment for screen light sources also makes MoiréPose robust for the various illumination.

9 LIMITATION AND DISCUSSION

Camera's 6-DoF Tracking: The purpose of MoiréPose is to output a 6-DoF pose estimation result of the camera according to an image, which is consistent with conventional image-based pose estimation tasks. Notably, the screen occupies the main part of the image, and the pose detection result of the camera is relative to the screen's virtual origin. Such an image-based pose estimation mechanism provides an interface to implement extensive applications. For example, to realize the 6-DoF tracking of the camera, we only need to restore the virtual origin's translation between consecutive frames, which can be deduced by the camera's translation. Actually, the investigation of this problem has been done in some state-of-art works [8, 58], *i.e.*, the camera translation can be characterized by the phase information of the moiré pattern. Thus, it is possible to estimate the changes of the screen's virtual origin so as to realize the continuous 6-DoF tracking of the camera.

Bright application perspective: MoiréPose can be applied not only to the camera-to-screen interactions but also to the pervasive camera-to-texture scenarios. The interaction target can be pervasive objects in daily life with dense periodic textures, including stripe clothes, buildings with period textures, or any artificial gratings with high frequency [42]. For example, artificial moiré gratings can help the camera realize much higher localization precision than traditional visual markers, which can be applied to Human-Computer Interaction (HCI) [59] and mobile robot localization [17, 49]. Furthermore, the artificial gratings can also enable the camera to realize the 3D shape reconstruction [22, 25, 37, 41] and the integrated sensing and communication [14, 30, 34]. Thus, MoiréPose has a bright perspective for both enhancing the localization precision and extending the sensing dimensions.

Focus of the camera: The camera's fast movement and the ambient brightness changes can trigger the camera's auto focus. Specifically, smartphone cameras leverage autofocus algorithms and lens motors to achieve the sharpest images. The autofocus algorithms evaluate the sharpness of the image by calculating the contrast value of the image. The lens motor then controls the lens movement until it finds the correct position for the sharpest image. Fortunately, the screens we interact with in real scenarios usually show the interaction content, *e.g.*, the game interface. These complex background patterns can help the autofocus algorithm rapidly converge the contrast calculation so as to realize fast focus. Note that the refocusing phenomenon is not frequent. When using MoiréPose for continuous 6-DoF tracking, the influence of the refocusing can be effectively reduced by leveraging the continuity of the camera pose.

Working range: The working range of MoiréPose is related to the screen's pixel size, which is usually proportional to the screen size. Large screens with wide pixel size usually require a far interaction distance to produce clear moiré patterns since the far distance can help the projective screen grating's frequency approach that of CFA. In contrast, smaller screens with narrower pixels usually require a closer interaction distance to generate clear moiré patterns. Therefore, the effective working range of MoiréPose exactly matches the accustomed usage habits of screens with different sizes. Different working ranges can support different applications. Motion-sensing games with active body stretch usually need an open area, *i.e.*, the far-field interaction scenarios. Whereas implicit interactions, including mouse control [18] and vision care (distance and direction monitoring when using screens), with laptops and displays are suited to near-field scenarios.

Symmetric distance solution: Due to the non-directivity of the moiré pattern's splines, the observed propagation direction θ_m in Eq. (9) can be expressed as:

$$\theta_m = \begin{cases} \theta'_m, & f_c > f_p \\ \theta'_m + \pi, & f_c < f_p \end{cases}, \quad -\frac{\pi}{2} \le \theta'_m < \frac{\pi}{2}, \tag{14}$$

where θ'_m represents the included angle between the line of propagation direction and the positive horizontal axis of the image. When the spatial frequencies of the CFA and the projective screen are equal, *i.e.*, $f_c = f_p$, the frequency of the moiré pattern becomes zero, which corresponds to a critical distance $d_c = \frac{f_c f}{f_s}$. Thus, the distance solutions from the moiré feature lie in two symmetric intervals bounded by d_c. The ambiguity of the distance can be eliminated according to the effective working range limited by the device parameters. Without loss of generality, we choose the closer interval in this paper to ensure the sufficient screen resolution.

Large perspective angle: Since the camera lens is equivalent to a convex lens, the screen grating projected onto the CFA layer has undergone the deformation caused by the convex effect, *i.e.*, moiré pattern's one of the sub-gratings to be superimposed has been deformed. It will cause a significant distortion on the final superimposed moiré pattern, especially for large perspective angles between the screen and camera. Such severe pattern distortion cannot be simply eliminated by image distortion correction. Nevertheless, it is still possible to make a special correction according to the characteristics of the convex lens to restore a more accurate moiré pattern.

10 CONCLUSION

In this paper, we propose a novel sensing method MoiréPose to achieve ultra-high precision on the camera's 6-DoF pose estimation. We extend the localization precision from the conventional *pixel level* to the *spline level* based on the low-frequency property of moiré pattern. We explore a mathematical model to associate the 6-DoF camera pose with the extracted moiré features and effectively extend the working range of moiré-based mechanism. Extensive experiment results show that our system achieves average position error of 7.5 mm and overall posture error of 1.66°.

ACKNOWLEDGMENTS

This work is supported in part by National Natural Science Foundation of China under Grant Nos. 61832008, 61872174, 61832005, 61902175. This work is partially supported by Collaborative Innovation Center of Novel Software Technology and Industrialization. This work is partially supported by the National Science Foundation Grants CNS2120396, CCF1909963, and CCF2211163.

REFERENCES

[1] 2022. Bayer filter. Accessed Mar. 1, 2022.. https://en.wikipedia.org/wiki/Bayer_filter

[2] 2022. Color Filter Array. https://en.wikipedia.org/wiki/Color_filter_array.

[3] 2022. Distortion. Accessed Mar. 1, 2022.. https://en.wikipedia.org/wiki/Distortion

[4] 2022. OptiTrack. Accessed Mar. 1, 2022.. https://optitrack.com

[5] 2022. Vignetting. https://en.wikipedia.org/wiki/Vignetting.

[6] Isaac Amidror. 2009. *The Theory of the Moiré Phenomenon: Volume I: Periodic Layers*. Vol. 38. Springer Science & Business Media.

[7] Pratiksha Andhare and Sayali Rawat. 2016. Pick and place industrial robot controller with computer vision. In *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*. IEEE, 1–4.

[8] Samuel Banks, Richard Green, and Jun Junghyun. 2019. Use of Moiré Patterns in Camera Position Estimation. In *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)*. IEEE, 1–7.

[9] Filippo Bergamasco, Andrea Albarelli, Emanuele Rodola, and Andrea Torsello. 2011. Rune-tag: A high accuracy fiducial marker with strong occlusion resilience. In *CVPR 2011*. IEEE, 113–120.

[10] blippar. 2022. Easily create your own interactive augmented reality experiences. https://www.layar.com/.

[11] John F Brenner, Brock S Dew, J Brian Horton, Thomas King, Peter W Neurath, and William D Selles. 1976. An automated microscope for cytologic research a preliminary evaluation. *Journal of Histochemistry & Cytochemistry* 24, 1 (1976), 100–111.

[12] BroadcastAR. 2022. Explore the unexplorable and touch the untouchable with BroadcastAR. https://www.indestry.com/broadcast-ar.

[13] Xi Cheng, Zhenyong Fu, and Jian Yang. 2021. Improved multi-scale dynamic feature encoding network for image demoiréing. *Pattern Recognition* 116 (2021), 107970.

[14] Yushi Cheng, Xiaoyu Ji, Lixu Wang, Qi Pang, Yi-Chao Chen, and Wenyuan Xu. 2021. {mID}: Tracing screen photos via {Moiré} patterns. In *30th USENIX Security Symposium (USENIX Security 21)*. 2969–2986.

[15] Joseph DeGol, Timothy Bretl, and Derek Hoiem. 2017. Chromatag: A colored marker and fast detection algorithm. In *Proceedings of the IEEE International Conference on Computer Vision*. 1472–1481.

[16] Ron S Dembo, Stanley C Eisenstat, and Trond Steihaug. 1982. Inexact newton methods. *SIAM Journal on Numerical analysis* 19, 2 (1982), 400–408.

[17] Marcos Ferreira, Paulo Costa, Luís Rocha, and A Paulo Moreira. 2016. Stereo-based real-time 6-DoF work tool tracking for robot programing by demonstration. *The International Journal of Advanced Manufacturing Technology* 85, 1 (2016), 57–69.

[18] Yun Fu and Thomas S Huang. 2007. hMouse: Head tracking driven virtual computer mouse. In *2007 IEEE Workshop on Applications of Computer Vision (WACV'07)*. IEEE, 30–30.

[19] Diogo Caetano Garcia and Ricardo L de Queiroz. 2015. Face-spoofing 2D-detection based on Moiré-pattern analysis. *IEEE transactions on information forensics and security* 10, 4 (2015), 778–786.

[20] Basilios Gatos, Ioannis Pratikakis, and Stavros J Perantonis. 2004. An adaptive binarization technique for low quality historical documents. In *International Workshop on Document Analysis Systems*. Springer, 102–113.

[21] Bahadir K Gunturk, John Glotzbach, Yucel Altunbasak, Ronald W Schafer, and Russel M Mersereau. 2005. Demosaicking: color filter array interpolation. *IEEE Signal processing magazine* 22, 1 (2005), 44–54.

[22] Haiyun Guo, Haowen Zhou, and Partha P Banerjee. 2021. Single-shot digital phase-shifting Moiré patterns for 3D topography. *Applied Optics* 60, 4 (2021), A84–A92.

[23] Bin He, Ce Wang, Boxin Shi, and Ling-Yu Duan. 2019. Mop moire patterns using mopnet. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2424–2432.

[24] Hongbo Jiang, Hangcheng Cao, Daibo Liu, Jie Xiong, and Zhichao Cao. 2020. Smileauth: Using dental edge biometrics for user authentication on smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–24.

[25] Hyunki Lee and Hyungsuck Cho. 2007. Stereo Moire Technique: A novel 3-D measurement method using a stereo camera and a digital pattern projector. *International Journal of Optomechatronics* 1, 2 (2007), 209–230.

[26] Yijie Li, Yi-Chao Chen, Xiaoyu Ji, Hao Pan, Lanqing Yang, Guangtao Xue, and Jiadi Yu. 2021. ScreenID: Enhancing QRCode Security by Fingerprinting Screens. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 1–10.

[27] Yuval Litvak, Armin Biess, and Aharon Bar-Hillel. 2019. Learning pose estimation for high-precision robotic assembly using simulated depth images. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 3521–3527.

[28] Jianwei Liu, Jinsong Han, Feng Lin, and Kui Ren. 2020. Adversary Helps: Gradient-based Device-Free Domain-Independent Gesture Recognition. *arXiv preprint arXiv:2004.03961* (2020).

[29] Lin Liu, Jianzhuang Liu, Shanxin Yuan, Gregory Slabaugh, Aleš Leonardis, Wengang Zhou, and Qi Tian. 2020. Wavelet-based dual-branch network for image demoiréing. In *European Conference on Computer Vision*. Springer, 86–102.

[30] Ziwei Liu, Tianyue Zheng, Chao Hu, Yanbing Yang, Yimao Sun, Yi Zhang, Zhe Chen, Liangyin Chen, and Jun Luo. 2022. CORE-Lens: Simultaneous Communication and Object REcognition with Disentangled-GAN Cameras. In *Proc. of the 28th ACM MobiCom*. 1–14.

[31] Andersson Technologies LLC. 2022. What is SynthEyes? https://www.ssontech.com/synovu.html.

[32] Louis-Philippe Morency, Jacob Whitehill, and Javier Movellan. 2008. Generalized adaptive view-based appearance model: Integrated framework for monocular head pose estimation. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE, 1–8.

[33] Rafael Munoz-Salinas. 2012. Aruco: a minimal library for augmented reality applications based on opencv. *Universidad de Córdoba* 386 (2012).

[34] Hao Pan, Yi-Chao Chen, Lanqing Yang, Guangtao Xue, Chuang-Wen You, and Xiaoyu Ji. 2019. mqrcode: Secure qr code using nonlinearity of spatial frequency in light. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–18.

[35] Abhinav Parate, Meng-Chieh Chiu, Chaniel Chadowitz, Deepak Ganesan, and Evangelos Kalogerakis. 2014. Risq: Recognizing smoking gestures with inertial sensors on a wristband. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. 149–161.

[36] PlayStation. 2022. PlayStation Move. https://www.youtube.com/watch?v=7Wuvia1-Mw8.

[37] C Quan, XY He, CF Wang, CJ Tay, and HM Shang. 2001. Shape measurement of small objects using LCD fringe projection with phase shifting. *Optics Communications* 189, 1-3 (2001), 21–29.

[38] Abigail Raney. 2017. Pinhole Camera Theory Summary. https://ourpastimes.com/pinhole-camera-theory-summary-12210465.html.

[39] Olinde Rodrigues. 1840. Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. *J. Math. Pures Appl* 5, 380-400 (1840), 5.

[40] Nataniel Ruiz, Eunji Chong, and James M Rehg. 2018. Fine-grained head pose estimation without keypoints. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2074–2083.

[41] Vladimir Saveljev and Sung-Kyu Kim. 2014. Three-dimensional moiré display. *Journal of the Society for Information Display* 22, 9 (2014), 482–486.

[42] Vladimir Saveljev, Sung-Kyu Kim, and Jaisoon Kim. 2018. Moiré effect in displays: a tutorial. *Optical Engineering* 57, 3 (2018), 030803.

[43] Longfei Shangguan, Zimu Zhou, and Kyle Jamieson. 2017. Enabling gesture-based interactions with objects. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. 239–251.

[44] Claude Elwood Shannon. 1949. Communication in the presence of noise. *Proceedings of the IRE* 37, 1 (1949), 10–21.

[45] Sheng Shen, He Wang, and Romit Roy Choudhury. 2016. I am a smartwatch and i can track my user's arm. In *Proceedings of the 14th annual international conference on Mobile systems, applications, and services*. 85–96.

[46] Yihua Shi, Jinfeng Yang, and Renbiao Wu. 2007. Reducing illumination based on nonlinear gamma correction. In *2007 IEEE International Conference on Image Processing*, Vol. 1. IEEE, I–529.

[47] Sinocrystal. 2022. LCD VS OLED. https://displaysino.com/technologyDetails/LCD-VS-OLED.html.

[48] Isaac Skog and Peter Händel. 2006. Calibration of a MEMS inertial measurement unit. In *XVII IMEKO world congress*. Citeseer, 1–6.

[49] Kai-Tai Song, Cheng-Hei Wu, and Sin-Yi Jiang. 2017. CAD-based pose estimation design for random bin picking using a RGB-D camera. *Journal of Intelligent & Robotic Systems* 87, 3 (2017), 455–470.

[50] Maki Sugimoto, Georges Kagotani, Minoru Kojima, Hideaki Nii, Akihiro Nakamura, and Masahiko Inami. 2005. Augmented coliseum: display-based computing for augmented reality inspiration computing robot. In *ACM SIGGRAPH 2005 Emerging technologies*. 1–es.

[51] Hideyuki Tanaka. 2020. Ultra-High-Accuracy Visual Marker for Indoor Precise Positioning. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2338–2343.

[52] Chung-Hsien Tsai and Jiung-Yao Huang. 2018. Augmented reality display based on user behavior. *Computer Standards & Interfaces* 55 (2018), 171–181.

[53] He Wang, Ted Tsung-Te Lai, and Romit Roy Choudhury. 2015. Mole: Motion leaks through smartwatch sensors. In *Proceedings of the 21st annual international conference on mobile computing and networking*. 155–166.

[54] He Wang, Souvik Sen, Ahmed Elgohary, Moustafa Farid, Moustafa Youssef, and Romit Roy Choudhury. 2012. No need to war-drive: Unsupervised indoor localization. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*. 197–210.

[55] John Wang and Edwin Olson. 2016. AprilTag 2: Efficient and robust fiducial detection. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 4193–4198.

[56] Wikipedia. 2022. Pixel Geometry. https://en.wikipedia.org/wiki/Pixel_geometry.

[57] Wikipedia. 2022. Rolling shutter. https://en.wikipedia.org/wiki/Rolling_shutter.

[58] Chang Xiao and Changxi Zheng. 2021. MoiréBoard: A Stable, Accurate and Low-cost Camera Tracking Method. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 881–893.

[59] Robert Xiao, Chris Harrison, Karl DD Willis, Ivan Poupyrev, and Scott E Hudson. 2013. Lumitrack: low cost, high precision, high speed tracking with projected m-sequences. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 3–12.

[60] Hanbin Zhang, Chenhan Xu, Huining Li, Aditya Singh Rathore, Chen Song, Zhisheng Yan, Dongmei Li, Feng Lin, Kun Wang, and Wenyao Xu. 2019. Pdmove: Towards passive medication adherence monitoring of parkinson's disease using smartphone-based gait assessment. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 3 (2019), 1–23.

[61] Zhengyou Zhang. 2000. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence* 22, 11 (2000), 1330–1334.

[62] Pengfei Zhou, Mo Li, and Guobin Shen. 2014. Use it free: Instantly knowing your phone attitude. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. 605–616.

[63] Shilin Zhu, Chi Zhang, and Xinyu Zhang. 2017. Automating visual privacy protection using a smart led. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. 329–342.