Opto-Electronic Advances

CN 51–1781/TN ISSN 2096–4579 (Print) ISSN 2097–3993 (Online)

Scale-invariant 3D face recognition using computer-generated holograms and the Mellin transform

Yongwei Yao, Yaping Zhang, Huanrong He, Xianfeng David Gu, Daping Chu and Ting-Chung Poon

Citation: Yao YW, Zhang YP, He HR, et al. Scale-invariant 3D face recognition using computer-generated holograms and the Mellin transform. *Opto-Electron Adv* **8**, 250095(2025).

https://doi.org/10.29026/oea.2025.250095

Received: 3 May 2025; Accepted: 27 August 2025; Published online: 28 October 2025

Opto-Electronic Advances

Article

November 2025, Vol. 8, No. 11

DOI: 10.29026/oea.2025.250095 CSTR: 32247.14.oea.2025.250095

Scale-invariant 3D face recognition using computer-generated holograms and the Mellin transform

Yongwei Yao[1], Yaping Zhang[1,2]*, Huanrong He[1], Xianfeng David Gu[3], Daping Chu[4] and Ting-Chung Poon[5]

We present a novel method for scale-invariant 3D face recognition by integrating computer-generated holography with the Mellin transform. This approach leverages the scale-invariance property of the Mellin transform to address challenges related to variations in 3D facial sizes during recognition. By applying the Mellin transform to computer-generated holograms and performing correlation between them, which, to the best of our knowledge, is being done for the first time, we have developed a robust recognition framework capable of managing significant scale variations without compromising recognition accuracy. Digital holograms of 3D faces are generated from a face database, and the Mellin transform is employed to enable robust recognition across scale factors ranging from 0.4 to 2.0. Within this range, the method achieves 100% recognition accuracy, as confirmed by both simulation-based and hybrid optical/digital experimental validations. Numerical calculations demonstrate that our method significantly enhances the accuracy and reliability of 3D face recognition, as evidenced by the sharp correlation peaks and higher peak-to-noise ratio (PNR) values than that of using conventional holograms without the Mellin transform. Additionally, the hybrid optical/digital joint transform correlation hardware further validates the method's effectiveness, demonstrating its capability to accurately identify and distinguish 3D faces at various scales. This work provides a promising solution for advanced biometric systems, especially for those which require 3D scale-invariant recognition.

Keywords: 3D face recognition; computer-generate holography; Mellin transform; scale invariance; biometrics

Yao YW, Zhang YP, He HR et al. Scale-invariant 3D face recognition using computer-generated holograms and the Mellin transform. *Opto-Electron Adv* **8**, 250095 (2025).

Introduction

Face recognition technology, particularly in 3D, is increasingly becoming indispensable in the realm of biometric authentication due to its superior accuracy and robustness when compared to traditional two-dimensional (2D) methods[1,2]. As the demand for more reliable and robust recognition systems grows, researchers have focused on overcoming the inherent challenges associated with 3D facial recognition. One major challenge is the variability in scale, which arises from differences in the distance between the subject and the imaging device, posing a significant obstacle[3]. This issue is especially

[1]Yunnan Provincial Key Laboratory of Modern Information Optics (LMIO), Kunming University of Science and Technology, Kunming 650500, China; [2]Cambridge Digital Humanities (CDH), University of Cambridge, Cambridge CB2 1RX, UK; [3]Computer Science Department, SUNY at Stony Brook, Stony Brook, New York 11794, USA; [4]Centre for Photonic Devices and Sensors, University of Cambridge, 9 JJ Thomson Avenue, Cambridge CB3 0FA, UK; [5]Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061, USA.

*Correspondence: YP Zhang, E-mail: yaping.zhang@gmail.com

Received: 3 May 2025; Accepted: 27 August 2025; Published online: 28 October 2025

critical in real-world applications where subjects are not consistently positioned at a fixed distance from the camera, potentially degrading the recognition performance. Addressing this scale variability is essential to enhance the reliability and applicability of 3D face recognition systems in security, surveillance, and identity verification scenarios.

In recent years, various methods for 3D face recognition have emerged[4]: Keypoint-based methods[5] and local surface feature techniques[6–8] often extract local descriptors such as curvature, texture, or point clouds, but these features are inherently sensitive to scale variations unless explicitly normalized, which introduces additional computational complexity. Geometric model approaches, such as the triangular geometrical model by Ali et al.[9], achieve scale invariance by comparing triangles in 3D space based on similarity criteria. However, these methods simplify the 3D data into sparse geometric representations, which limits the ability to preserve holistic facial information and makes reconstruction of the original 3D face impossible from the processed data. Recent advances in deep learning[10] offer new approaches for 3D face recognition with some robustness to scaling, but learning-based solutions[11] inevitably depend on substantial computer hardware resources and extensive large-scale 3D face training datasets[1].

The Mellin transform is an integral transform that converts functions into a form that is particularly useful for analyzing scale changes[12]. In this study, we present a method that incorporates the Mellin transformation along with computer-generated holography for scale-invariant 3D face recognition. Holography, invented by Gabor, is a technique that allows 3D images to be recorded on a 2D medium, such as film or a CCD camera, creating what is known as a hologram[13,14]. Once we have recorded a hologram, we can reconstruct the 3D object using optical or digital technique. Zhou et al. proposed a 3D face recognition method utilizing computer-generated holograms[15], which enhanced recognition efficiency by converting 3D feature points into 2D holograms and employing digital correlation for matching. However, the method did not demonstrate robustness with respect to scale invariance. Our approach leverages the Mellin transform's capability to handle scale variations by normalizing the scale changes in the holographic domain, thus ensuring that the recognition process remains effective regardless of the scale of a 3D face in the 2D image. 3D shape- matching has been simplified to 2D image-matching using conformal geometry[16]. At the same time, there is a technique to enhance state-of-the-art 2D face recognition approaches with 3D features[17]. In our computer-generated holographic technique, we create computer-generated holograms of 3D surfaces using triangular meshes, which are derived from 3D face scans. Since holography preserves holistic information of a 3D face[13,14], recognition will be highly accurate.

In this paper, we propose a novel method for achieving scale-invariant 3D face recognition by integrating computer-generated holography (CGH) with the Mellin transform. This approach effectively addresses a fundamental challenge in 3D biometric recognition—namely, the variability in object scale caused by differences in acquisition distance. By encoding 3D facial data into optical scanning holography (OSH)-inspired mesh-layer-based holograms and applying a coordinate transformation in the Mellin domain, we construct a robust recognition framework capable of maintaining consistent performance across a wide range of scale variations. The proposed method preserves the holistic 3D structure of facial models with high fidelity and theoretically guarantees scale invariance, which is further validated through extensive numerical simulations and hybrid optical/digital joint transform correlation (JTC) experiments.

Methodology

Digital generation of 3D face holograms inspired by optical scanning holography (OSH)

We use a 3D face database[18] from the publicly available database of the 3D Scanning Laboratory at Stony Brook University, in which the data are captured by a phase-shifting structured light-ranging system with approximately 80 K 3D points [camera: Dalsa CA-D6-0512W with resolution 532×516 pixels of pixel size 10 μm, fringe projector: Kodak DP900 projector with available wavelengths of 360–700 nm]. The point cloud model is then converted to a triangular-mesh model by using triangulation along with computer graphics. Figure 1(a) and 1(c) display the captured face of real persons named "Sophie" and "Alex", respectively. Their polygon-mesh representations are shown in Fig. 1(b) and Fig. 1(d).

There are three prominent techniques to generate computer-generated holograms: point-based[19], layer-based[20] and polygon-based (or mesh-based) methods[21]. Most recently, Yao et al. have proposed a hybrid layer-mesh-based technique for fast generation of computer-

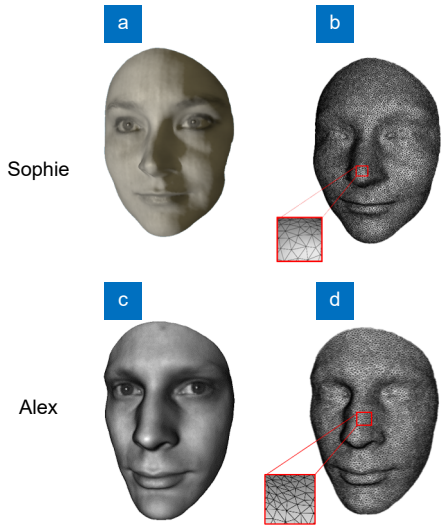

Fig. 1 | Captured face of real persons. (**a**) Picture of Sophia. (**b**) Sophia consisting of polygon meshes. (**c**) Picture of Alex. (**d**) Alex consisting of polygon meshes. The 3D face models 'Sophie' and 'Alex' are sourced from the publicly available database of the 3D Scanning Laboratory at Stony Brook University and are used under academic license.

generated holograms[22]. We shall use such technique to represent the 3D face as a series of discrete layers from the polygon-mesh 3D face. Briefly, in the technique, for a given 3D mesh, we calculate the barycentric coordinates for each triangular mesh. Each barycenter, indicated by blue or yellow dots, is mapped to the nearest layer, thereby preserving depth information, as shown in Fig. 2(a). Once layering is completed, we obtain mesh-layer-based 3D face data, $\sum_{i=1}^{n} I_0(x,y;z_i)$, where $I_0(x,y;z_i)$ represents the intensity distribution function of the i-th image plane, located at an axial distance z_i from the hologram plane. The 3D face is divided into n image planes that are uniformly spaced at Δz and parallel to the hologram.

Each layer $I_0(x,y;z_i)$ is subsequently convolved with digitally generated cosine Fresnel zone plates (FZP), $\frac{1}{\lambda z_i}\cos\left[\frac{\pi}{\lambda z_i}(x^2+y^2)\right]$, and sine Fresnel zone plates (FZP), $\frac{1}{\lambda z_i}\sin\left[\frac{\pi}{\lambda z_i}(x^2+y^2)\right]$, resulting in the corresponding cosine hologram, $H_{\cos}(x,y)$, and sine hologram, $H_{\sin}(x,y)$, for the 3D face:

$$H_{\cos}(x,y)=\sum_{i=1}^{n} I_0(x,y;z_i) * \frac{1}{\lambda z_i}\cos\left[\frac{\pi}{\lambda z_i}(x^2+y^2)\right], \tag{1}$$

and

$$H_{\sin}(x,y)=\sum_{i=1}^{n} I_0(x,y;z_i) * \frac{1}{\lambda z_i}\sin\left[\frac{\pi}{\lambda z_i}(x^2+y^2)\right], \tag{2}$$

where * denotes 2-D convolution, and λ represents the wavelength of the laser used to generate the hologram. Equations (1) and (2) are derived based on principles inspired by physical phenomena and represent the actual holograms produced through optical scanning holography (OSH)[14,23,24]. Therefore, Eq. (1) and Eq. (2) can be referred to as OSH-based computer-generated holograms (CGHs) of a layered 3D object. From the above two holograms, we obtain a complex hologram by performing a digital complex operation as follows[14,23,24]:

$$H_c(x,y)=H_{\cos}(x,y)+\mathrm{j}H_{\sin}(x,y), \tag{3}$$

where $\mathrm{j}=\sqrt{-1}$. Figure 2(b) illustrates the process of

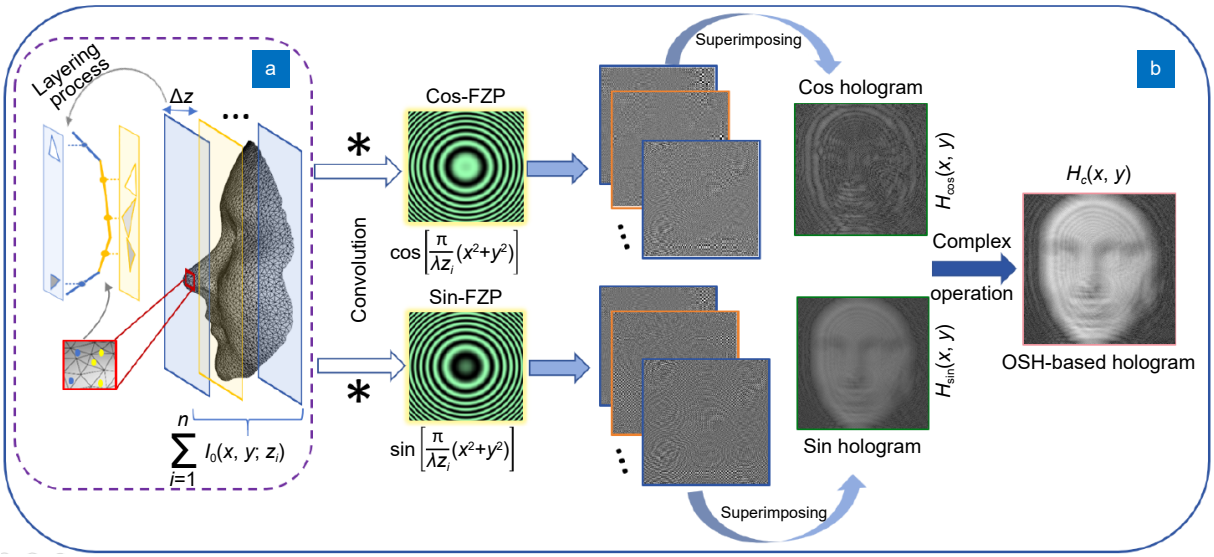

Fig. 2 | (**a**) Layering of 3D meshed face. (**b**) Digital generation of complex hologram $H_c(x, y)$ using OSH-based holograms.

Eq. (1) to Eq. (3). By leveraging the layering approach and applying the principles of OSH, we ensure that our digital holograms maintain high fidelity in representing the 3D characteristics of the facial models. This process is crucial for achieving reliable and robust 3D face recognition, particularly under various scale conditions.

Mellin transform holograms and their correlations

Once the 3D information of the face is encoded in OSH-based digital holograms, the Mellin transform can be performed on the hologram to achieve scale-invariant 3D face recognition. The Mellin transform of the 3D face hologram $H_c(x,y)$ can be represented as

$$H_{\mathrm{Mc}}(\omega,\gamma)=\mathcal{M}\{H_c(x,y)\}=\iint_0^{\infty}H_c(x,y)\,x^{\mathrm{j}\omega-1}y^{\mathrm{j}\gamma-1}\mathrm{d}x\mathrm{d}y, \tag{4}$$

where $\mathcal{M}\{\cdot\}$ represents the Mellin transform operation. (ω,γ) are the coordinates of the output plane of the Mellin transform. Performing coordinate transformations to the input coordinates $x=\mathrm{e}^t$, $y=\mathrm{e}^v$, i.e., $\ln(x)=t$, $\ln(y)=v$, Eq. (4) is manipulated to become

$$\begin{aligned}H_{\mathrm{Mc}}(\omega,\gamma)=&M\{H_c(x,y)\}=\iint_{-\infty}^{\infty}H_c(\mathrm{e}^t,\mathrm{e}^v)\,\mathrm{e}^{\mathrm{j}\omega t}\mathrm{e}^{\mathrm{j}\gamma v}\mathrm{d}t\mathrm{d}v\\=&F\{H_c(\mathrm{e}^t,\mathrm{e}^v)\},\end{aligned} \tag{5}$$

where $\mathcal{F}\{\cdot\}$ represents the two-dimensional Fourier transform (FT). Equation (5) indicates that the Mellin transformation can alternatively be performed by first applying the Mellin coordinate transformation to obtain $H_c(\mathrm{e}^t,\mathrm{e}^v)$, followed by Fourier transform.

After scaling the 3D facial model and generating the scaled complex hologram $H_c(ax,by)$, as outlined in Eq. (4), the Mellin transform of $H_c(ax,by)$ is given by

$$\begin{aligned}M\{H_c(ax,by)\}=&a^{-\mathrm{j}\omega}b^{-\mathrm{j}\gamma}M\{H_c(x,y)\}\\=&\mathrm{e}^{-\mathrm{j}\omega\ln(a)}\mathrm{e}^{-\mathrm{j}\gamma\ln(b)}M\{H_c(x,y)\},\end{aligned} \tag{6}$$

where a and b represent the scale factors in the x and y directions. Equation (6) demonstrates that the Mellin transform of a scaled complex hologram is identical to the Mellin transform of the original complex hologram, except for some phase factors that depend on the scale factors. From Eq. (6), we have

$$|\mathcal{M}\{H_c(ax,by)\}|=|\mathcal{M}\{H_c(x,y)\}|. \tag{7}$$

Furthermore, we can show that

$$\begin{aligned}&\mathcal{F}^{-1}\{\mathcal{M}\{H_c(x,y)\}\mathcal{M}^*\{H_c(ax,by)\}\}\\&=[H_c(\mathrm{e}^t,\mathrm{e}^v)\otimes H_c(\mathrm{e}^t,\mathrm{e}^v)]*\delta(\omega-\ln(a),\gamma-\ln(b)),\end{aligned} \tag{8}$$

where $\otimes$ represents correlation operation. Equation (8) represents the correlation result in the Mellin-transformed domain between the original complex hologram $H_c(x,y)$ and a scaled version of it $H_c(ax,by)$, where a and b are scale factors in the x and y directions, respectively. Physically, this means that scale changes in the original 3D face holograms do not affect the correlation result, but only shift the peak position, making the system inherently scale-invariant. In other words, this formulation allows us to match scaled versions of a 3D face hologram by transforming the problem into a domain where scale becomes a shift, making it easier to recognize faces regardless of their size.

Results and discussion

Scale-invariant 3D face recognition

Based on Eq. (8), Fig. 3 presents the overall framework for achieving scale-invariant 3D face recognition using Mellin transformed complex holograms. The process begins with slicing the 3D face mesh model. Two realistic 3D face models, 'Sophie' and 'Alex,' are used for simulations. They are sourced from the publicly accessible archives of Stony Brook University's 3D Scanning Laboratory[18]. The 3D face model Sophie, consisting of 14254 polygons, is scaled using factors a and b ranging from 0.4 to 2.0, resulting in 17 different sized versions of Sophie as the 3D face to be recognized. Another 3D face model, Alex, consists of 14007 triangular polygons, is used as a comparison in face verification experiments with Sophie.

The 3D face models are segmented along the z-axis into 60 distinct layers, with each layer separated by 150 μm. The segmentation specifies layers at intervals $z=z_i$, for $i=1,2,\cdots,60$, extending from $z_1=260$ mm to $z_{60}=269$ mm. Specifically, the number of layers (60) and the layer spacing (150 μm) were chosen to balance human eye resolution and computational efficiency. The layering interval and number of layers were empirically optimized based on preliminary experiments to ensure high fidelity in representing the 3D characteristics of the facial models. In addition, we carried out the same number of layers and set the same layer spacing parameters for all 3D face models, so as to avoid the influence of the setting of layer number and layer spacing on the recognition results. Subsequently, OSH-based holograms of size

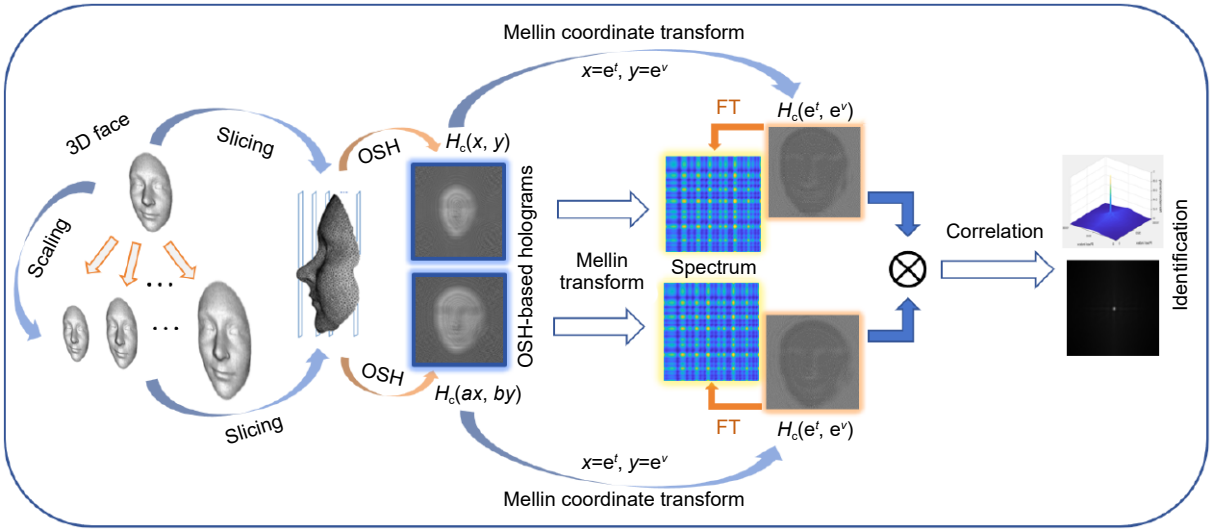

Fig. 3 | Overall block diagram of scale-invariant 3D face recognition: ⊗ denotes correlation.

6 mm × 6 mm and resolution of 256 × 256 pixels were generated. We have used wavelength of 532 nm as a laser source. Figure 4(a) depicts the magnitude of the complex hologram of Sophie without any scaling ($a = b = 1$). Figure 4(b) plots the magnitude of the Mellin transform calculated according to Eq. (4), serving as a reference for further correlation operation.

The dimensions of the Sophie 3D facial model are varied systematically with scale factors a and b adjusted from 0.4 to 2.0 in increments of 0.1, excluding the standard scale ($a = b = 1.0$), resulting in 16 distinctly scaled models. The scale range, beginning at 0.4, was selected to ensure that facial features remain sufficiently resolvable and recognition accuracy is preserved, taking into account the spatial bandwidth of the spatial light modulator (SLM) and the pixel resolution of the 3D face models employed in our experiments. The 16 complex holograms are displayed in Fig. 5(a), mapping in 4 × 4 cells. Figure 5(b) and 5(c) further display the magnitude of the Mellin transform of the complex holograms for Sophie with scale factors $a = b = 0.5$ and $a = b = 2.0$, respectively. Comparing Fig. 4(b), Fig. 5(b) and Fig. 5(c) reveals that the magnitude of the Mellin transform remains the same for different scale factors, thereby demonstrating the validity of Eq. (7) for Mellin transformed complex holograms.

To validate the effectiveness of our scale-invariant 3D face recognition methodology, cross-correlation computations were conducted between the reference complex hologram shown in Fig. 4(a) and the array of 16 complex holograms detailed in Fig. 5(a). With Mellin transform, i.e., calculations were performed according to Eq. (8),

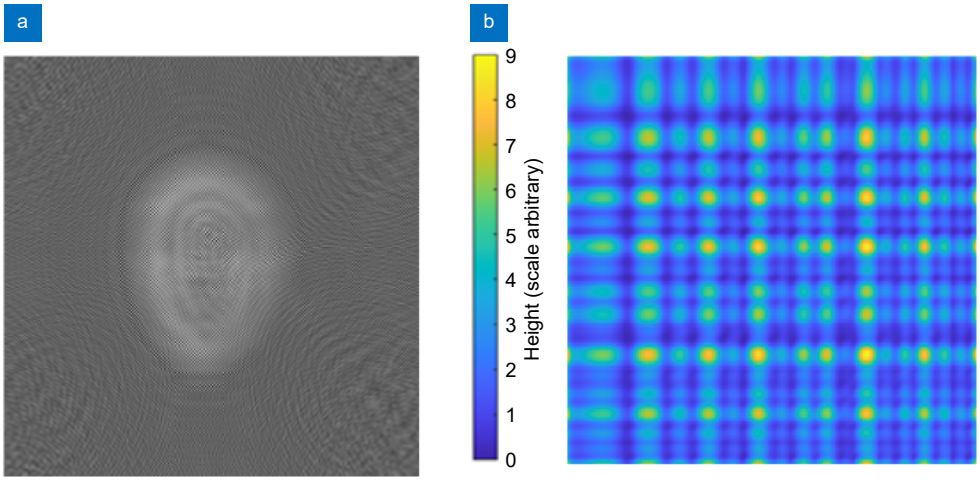

Fig. 4 | (**a**) Complex hologram of Sophie (plotting magnitude). (**b**) Magnitude of Mellin transform of (a).

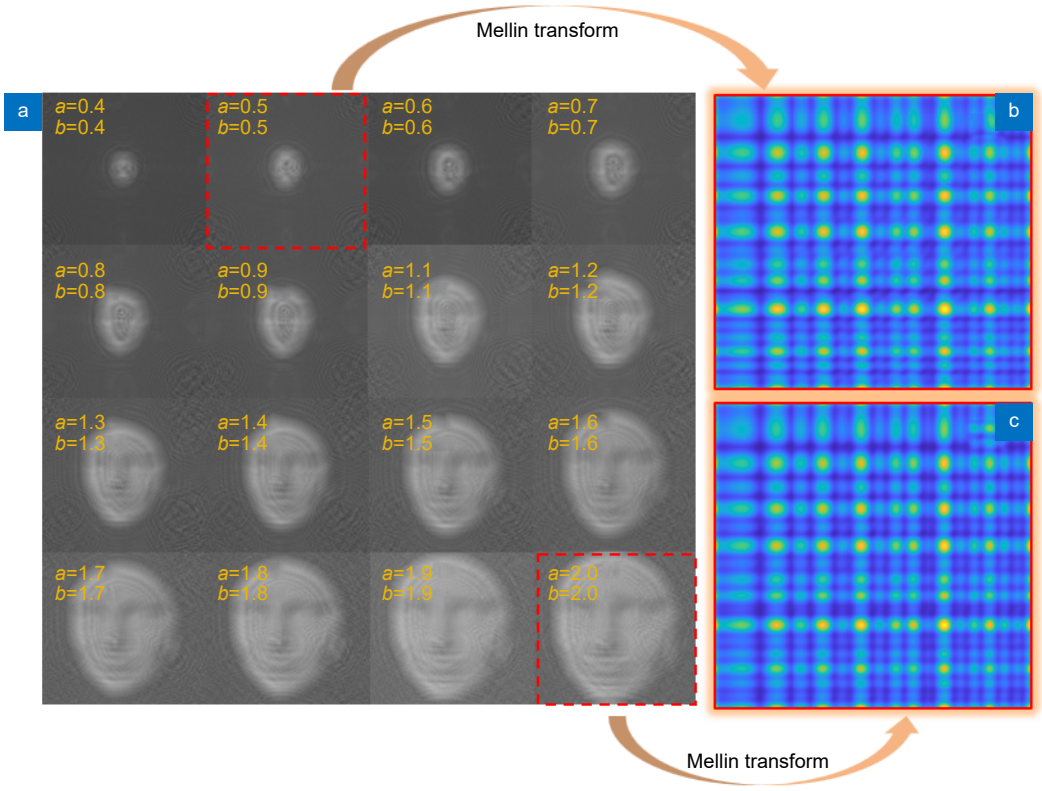

Fig. 5 | (**a**) Complex holograms for 16 scaled faces, mapping in 4 × 4 cells. (**b**) Magnitude of Mellin transforms with scale factors *a*=*b*=0.5. (**c**) Magnitude of Mellin transforms with scale factors *a*=*b*=2.0.

Fig. 6(a) and Fig. 6(b) depict the normalized cross-correlation patterns at scale factors $a = b = 0.5$ and $a = b = 2.0$, respectively, with reference complex hologram. Figure 6(c) shows the autocorrelation of the reference complex hologram, i.e., with no scaling, i.e., $a = b = 1.0$. These visualizations highlight the capability of the Mellin transformation to consistently identify 3D facial features across a range of scale factors,

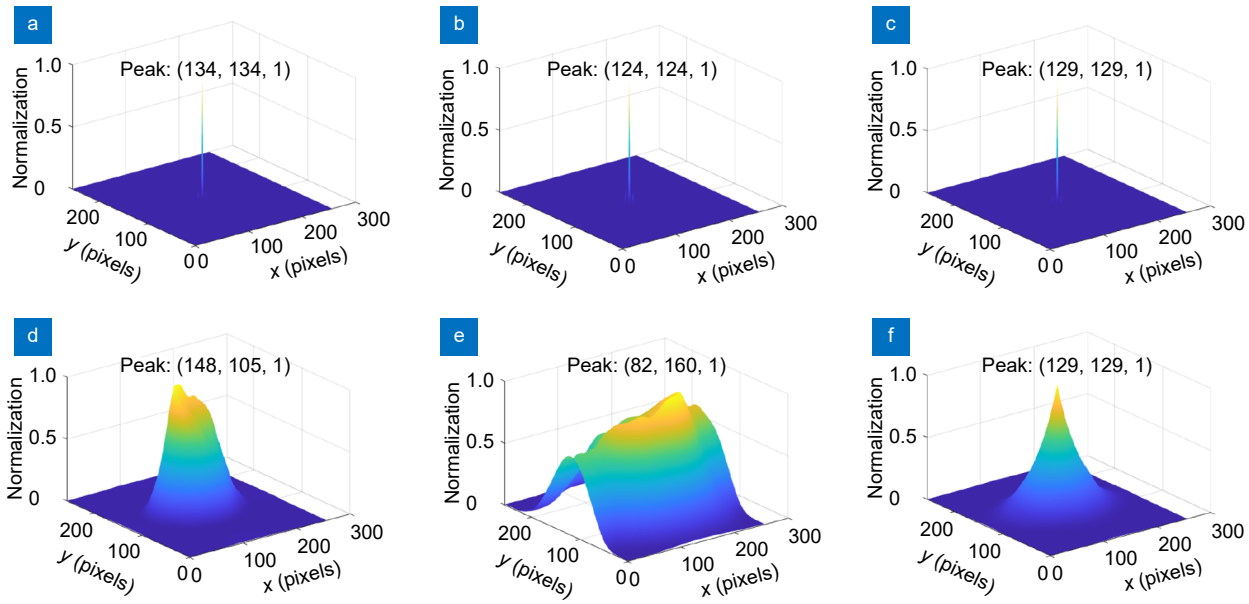

Fig. 6 | Comparison of OSH-based holograms correlation results with and without Mellin transform at different scale factors for the same 3D face model Sophie. (**a**) Normalized correlation results with Mellin transform at scale factors *a*=*b*=0.5. (**b**) *a*=*b*=2.0. (**c**) *a*=*b*=1.0. (**d**) Normalized correlation results without Mellin transform at scale factors *a*=*b*=0.5. (**e**) *a*=*b*=2.0. (**f**) *a*=*b*=1.0.

emphasizing its effectiveness in scale-invariant 3D face recognition.

When there is no Mellin transform involved, the normalized cross-correlation results for scale factors $a = b = 0.5$ and $a = b = 2.0$ are depicted in Fig. 6(d) and Fig. 6(e), respectively. Figure 6(f) depicts the autocorrelation pattern of the reference complex hologram, i.e., with scale factors $a = b = 1.0$. We recognize that there are no discernible peaks in Fig. 6(d) and Fig. 6(e), and the identification is failed. Figure 6(f) shows a peak because of autocorrelation of the reference complex hologram. It is interesting to note that by comparing Fig. 6(c) and Fig. 6(f), when the complex hologram is coordinate transformed with the Mellin transform [see Eq. (8)], the autocorrelation peak is very sharp as illustrated in Fig. 6(c). Hence, the use of Mellin coordinate transformation has preprocessing capability that enhances autocorrelation result, which is an unexpected benefit.

Correlation performance evaluation

We use the Peak-to-Noise Ratio (PNR) metric to quantitatively evaluate the performance of the proposed scale-invariant 3D face recognition method. The PNR is defined as the correlation peak value P_{corre} over the mean energy $\bar{E}$ over the correlation plane, indicating the correlation peak value relative to the mean noise on the correlation plane. The calculation of PNR is calculated as follows:

$$PNR = \frac{P_{\text{corre}}}{\bar{E}} = \frac{P_{\text{corre}}}{\frac{1}{M \cdot N}\sum_{m=1}^{M}\sum_{n=1}^{N}|C(m,n)|^2}, \quad (9)$$

where P_{corre} denotes the maximum value of the normalized correlation peak, $\bar{E}$ is the mean background energy calculated over the correlation plane including the main peak area, $C(m,n)$ is the correlation function, and M and N are the number of pixels in the x and y coordinates over the correlation plane.

Figure 7 shows the PNR values for the recognition based on correlation of coordinate-transformed complex holograms, as the scale factor varies from 0.4 to 2.0. As shown in the figure, the PNR values with the Mellin transform remain a constant value of log10(PNR)=4.5 throughout the scale change. For a reference, we show a blue dot in the figure, which corresponds to the case based on correlation of complex holograms only for a=b=1.0. The PNR value is log10(PNR)=1.2. The proposed method shows an improvement of at least two orders of magnitude in PNR values.

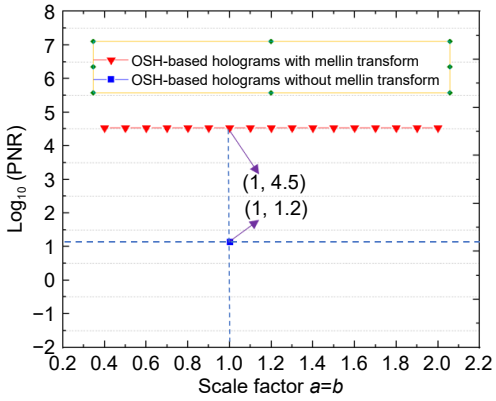

Fig. 7 | Curves of PNR of the correlation peaks at different scaling factors.

Hybrid optical/digital joint transformation correlation experiments

After coordinate transformation of complex holograms, correlation in Eq. (8) can be performed digitally using FFT. However, as the size of the complex hologram increases, the computation time will become prohibitively high. Since an optical lens can perform Fourier transform, it is advantageous to investigate a hybrid (optical/digital) approach to perform the calculation of Eq. (8). Briefly, if $f(x,y)$ is a pattern in the front focal plane of the lens with focal length f, on the back focal plane of the lens we have, besides some constant,

$$\mathcal{F}\{f(x,y)\}\big|_{k_x=\frac{k_0 x}{f},k_y=\frac{k_0 y}{f}} = \iint_{-\infty}^{\infty} f(x,y)\, e^{jk_x x + jk_y y} dxdy \Big|_{k_x=\frac{2\pi x}{\lambda f},k_y=\frac{2\pi y}{\lambda f}}, \quad (10)$$

where k_x and k_y are spatial frequencies corresponding to the x and y variables, respectively. It is noted that after the Fourier transform, the spatial frequency variables k_x and k_y are repalced by $\frac{2\pi x}{\lambda f}$ and $\frac{2\pi y}{\lambda f}$ to denote that the resulting function is still a function of spatial variables x and y. Hence, we refer to such a lens as a Fourier transform lens (FL) in optics[14]. Fourier transformation is processed in parallel in the optics domain, irrespective of the object's size, with the processing time being constant and equal to twice the focal length divided by the speed of light.

We have chosen the well-known joint transform correlation system in optics to investigate the hybrid implementation of Eq. (8)[14,25]. The hybrid optical/digital joint transform is shown in Fig. 8. A pair consisting of a

phase-only reference complex hologram $H_{\rm cr}$ and a target complex hologram $H_{\rm ct}$ is loaded onto a 2D spatial light modulator (SLM) from a computer. A 2-D SLM is a device with which one can imprint a 2-D pattern on a laser beam by passing the laser beam through it or by reflecting the laser beam off the device[14,23]. In Fig. 8, we show a reflection-type phase-only SLM[26] [We can only acquire phase-only or amplitude-only SLMs from commercial sources]. Hence, the pair of the pattern on the SLM is $H_{\rm cr}(x-x_0,y)+H_{\rm ct}(x+x_0,y)$. The two patterns are displayed side by side with a separation of $2x_0$. Upon reflection from the SLM, the reflected light is projected onto the focal plane of the Fourier transform lens (FL) onto the camera, giving the so-called joint transform spectrum (JTS) intensity pattern $J_{\rm M}$ to be detected by the camera:

$$\begin{aligned} J_{\rm M}(x,y) &= \left|F\left\{H_{\rm cr}(x+x_0,y)+H_{\rm ct}(x-x_0,y)\right\}\right|^2 \\ &= \left|F\left\{H_{\rm cr}\right\}\right|^2+\left|F\left\{H_{\rm ct}\right\}\right|^2 \\ &\quad + F^*\left\{H_{\rm cr}\right\}F\left\{H_{\rm ct}\right\}\exp\left(\mathrm{j}2k_xx_0\right) \\ &\quad + F^*\left\{H_{\rm ct}\right\}F\left\{H_{\rm cr}\right\}\exp\left(-\mathrm{j}2k_xx_0\right), \end{aligned} \tag{11}$$

again with the understanding that k_x and k_y are repalced by $\frac{2\pi x}{\lambda f}$ and $\frac{2\pi y}{\lambda f}$ to get back to the spatial domain. The absolute operation signifies that the camera detects intensity physically[14]. Once the camera detects the intensity pattern, the camera sends its output to a computer to perform the final digital Fourier transform (such as the use of FFT) to obtain correlation outputs:

$$\begin{aligned} &\mathcal{F}\left\{J_M(x,y)\right\}_{k_x=\frac{2\pi x}{\lambda f},k_y=\frac{2\pi y}{\lambda f}} \\ &\propto \left[H_{\rm rr}(-x,-y)+H_{\rm tt}(-x,-y)\right. \\ &\quad \left.+H_{\rm rt}(-x-2x_0,-y)+H_{\rm tr}(-x+2x_0,-y)\right], \end{aligned} \tag{12}$$

where $H_{\rm rt}(x,y)=H_{\rm cr}(x,y)\otimes H_{\rm ct}(x,y)$. $H_{\rm rr}$ and $H_{\rm tt}$ are the autocorrelations of the reference and the target complex holograms, entered at the origin of the correlation plane. $H_{\rm rt}$ and $H_{\rm tr}$ are the two cross-correlations, centered at $x=\pm 2x_0$. If $r=t$, $H_{\rm rr}$ and $H_{\rm tt}$ will give two additional strong auto-correlation peaks beside the centered peak. Equations (11) and (12) refer to the discussion of two complex holograms. If we employ the Millen transform, the correlations in Eq. (12) in principle is replaced by $H_{\rm cr}(\mathrm{e}^t,\mathrm{e}^v)\otimes H_{\rm ct}(\mathrm{e}^t,\mathrm{e}^v)$ if we neglect the values of a and b as $\omega\ln(a)\approx 0$ and $\gamma\ln(b)\approx 0$ [see Eq. (8)].

In the actual experiments as shown in Fig. 8, a green semiconductor laser with a wavelength of $\lambda=532$ nm is expanded by the beam expander (BE) and directed onto the phase-only spatial light modulator (SLM) (HOLOEYE PLUTO (NIR-011)), which has a resolution of 1920 × 1080 and a pixel size of 8 μm. The SLM is loaded with joint inputs formed by $H_{\rm cr}(\mathrm{e}^t,\mathrm{e}^v)$ and $H_{\rm ct}(\mathrm{e}^t,\mathrm{e}^v)$ with scale factors $a=b=1.0$ and $a=b=2.0$ for complex holograms. The light reflected from the SLM is reflected by a beam splitter (BS) and then passes through a Fourier transform lens (FL) with a focal length of $f=300$ mm before being received by a camera (MMRY UC900C Charge-coupled Device). The camera is positioned at the back focal plane of the Fourier transform lens FL to record the JTS intensity pattern $J_{\rm M}$. Once the JTS is recorded, we can digitally perform Fourier transform operation on a computer to obtain the various correlation

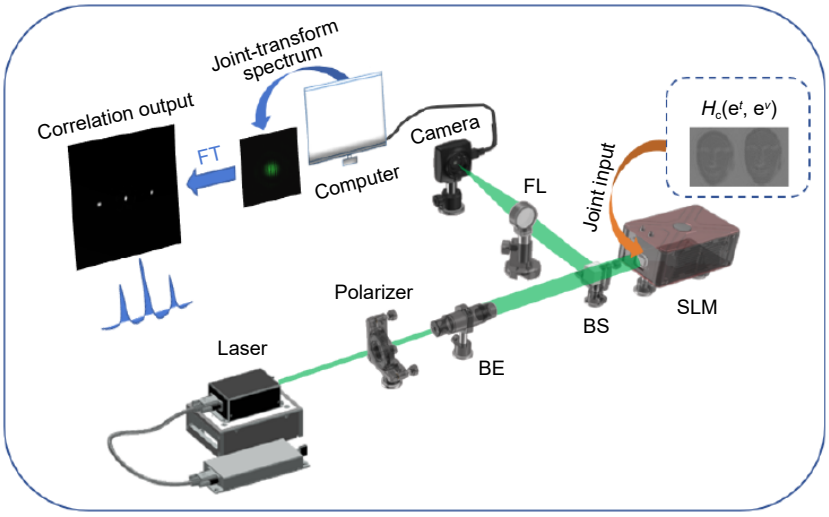

Fig. 8 | Hybrid optical/digital joint transform correlator. BE, beam expander; BS, beam splitter; SLM, spatial light modulator; FL: Fourier transform lens.

outputs. If the complex holograms of 3D face are identical regardless of any scales, the correlation output will contain three strong correlation peaks: the central peak and the ones on either side being the auto-correlation peaks for matched patterns.

Figure 9 shows the joint inputs loaded onto the SLM, which consists of the OSH-based complex holograms of different 3D faces without and with Mellin coordinate transform. Figure 9(a) and 9(b) display pairs of OSH-based phase-only holograms of the 3D face of Sophie, with scale factors $(a = b = 1.0)$ and $(a = b = 2.0)$, whereas Fig. 9(c) and 9(d) present pairs of OSH-based phase-only hologram of the 3D face of Sophie and Alex after encoding with the Mellin coordinates (ω, γ), again with scale factors $(a = b = 1.0)$ and $(a = b = 2.0)$ as indicated in the figure.

Figure 10 presents the results of the hybrid optical/digital joint transformation correlation experiments, where the inputs are from Fig. 9(a) and 9(b). The aim of these experiments is to verify through optical testing that accurate matching occurs only when the input faces are of the same identity and have the same scale, in the absence of the Mellin transform. Figure 10(a) and 10(d) are the JTS intensity patterns captured by the camera corresponding to the joint inputs in Figs. 9(a) and 9(b), respectively. Figure 10(b) and 10(e) are the results from performing the Fourier transform on Fig. 10(a) and Fig. 10(d) in the computer, i.e., giving us the correlation outputs. Figure 10(c) and 10(f) show the normalized intensity along the red dashed lines in Fig. 10(b) and 10(e), respectively. Figure 10(c) shows three strong correlation peaks, indicating successful verification of the input 3D face of Sophie. The correlation outputs in Fig. 10(e) shows only one bright spot (the central peak), indicating a failure to recognize the input 3D face of Sophie under different scaling factors. Therefore, the conventional method without Mellin transform cannot achieve scale-invariant 3D face recognition.

Figure 11 shows the results of the hybrid optical/digital joint transformation correlation experiments with inputs from Fig. 9(c) and 9(d). These experiments demonstrate that, after applying the Mellin transform, accurate matching is still achieved even when the input faces belong to the same identity but are at different scales. In Fig. 9(d), we intentionally introduced two uncontrolled variables, identity and scale, in order to contrast with Fig. 9(c). This comparison was made to show that, when the input faces are of different identities, even with the Mellin transform, the matching does not succeed. Figure 11(a) and 11(d) display the JTS intensity patterns captured by the camera corresponding to the joint inputs in Fig. 9(c) and 9(d), respectively. Figure 11(b) and 11(e) present the FFT outputs of Fig. 11(a) and 11(d), respectively. Figure 11(c) and 11(f) illustrate the normalized intensity along the red dashed lines in Fig. 11(b) and 11(e), respectively. Again, the illustration of 3-peak output in Fig. 11(c) denotes a match of the 3D face of Sophie even with scale factor of 2. On the other hand, a single peak output in Fig. 11 represent no match between Sophie and Alex.

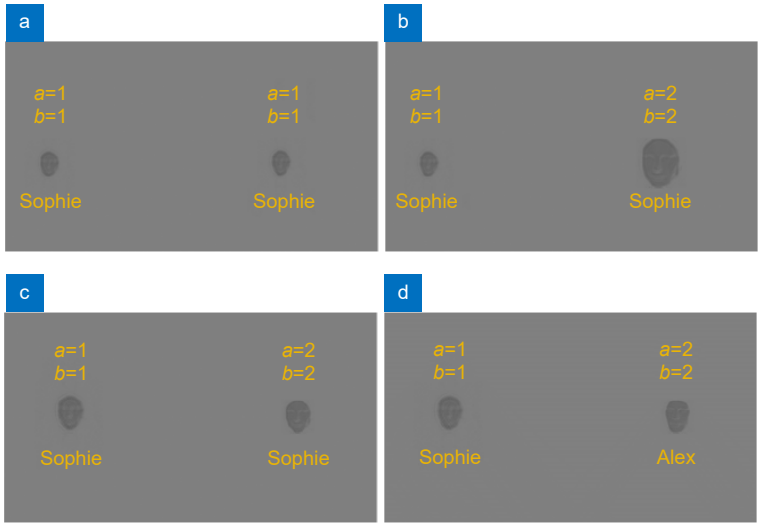

Fig. 9 | Joint inputs of OSH-based holograms without and with encoding of the Mellin coordinates in the hybrid joint transform correlator experiment. (**a**) Joint input of Sophie (*a*=*b*=1.0) vs. Sophie (*a*=*b*=1.0) without encoding of the Mellin coordinates. (**b**) Sophie (*a*=*b*=1.0) vs. Sophie (*a*=*b*=2.0) without encoding of the Mellin coordinates. (**c**) Sophie (*a*=*b*=1.0) vs. Sophie (*a*=*b*=2.0) with encoding of the Mellin coordinates. (**d**) Sophie (*a*=*b*=1.0) vs. Alex (*a*=*b*=2.0) with encoding of the Mellin coordinates.

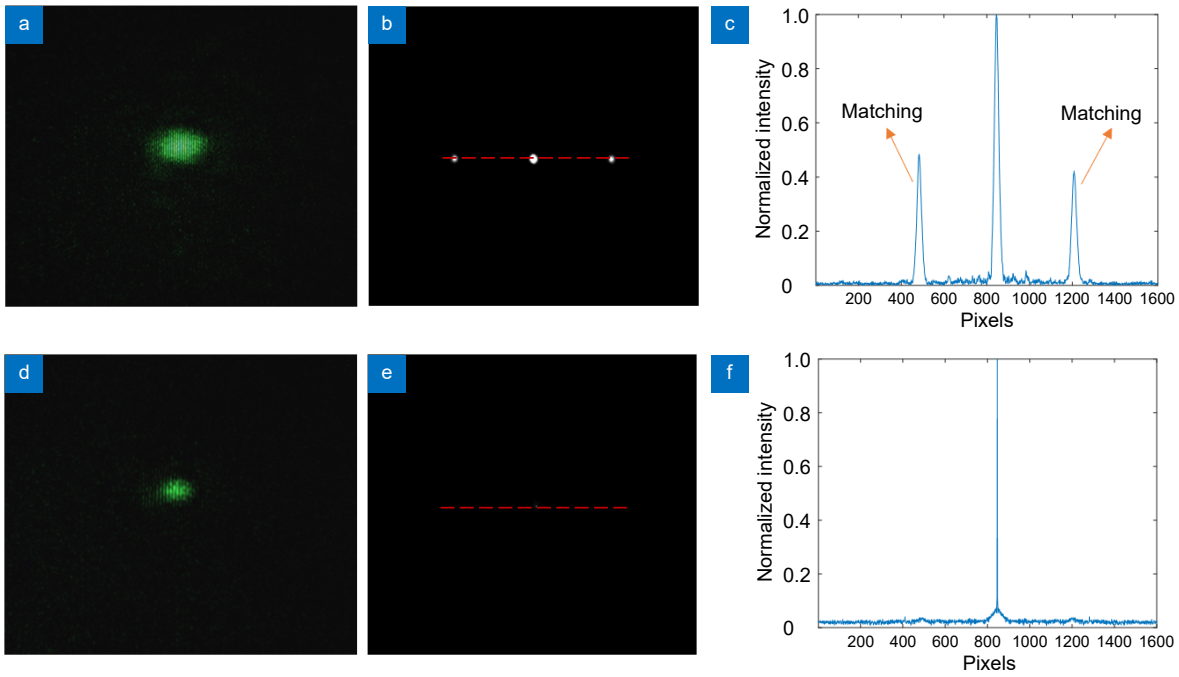

Fig. 10 | (**a**) JTS intensity patterns detected by camera corresponding to the joint input in Fig. 9(a). (**b**) FFT output of (a). (**c**) Normalized intensity plot along the red dashed lines in 10(b). (**d**) JTS intensity patterns detected by camera corresponding to the joint input in Fig. 9(b). (**e**) FFT output of 10(d). (**f**) Normalized intensity plot along the red dashed lines in (e).

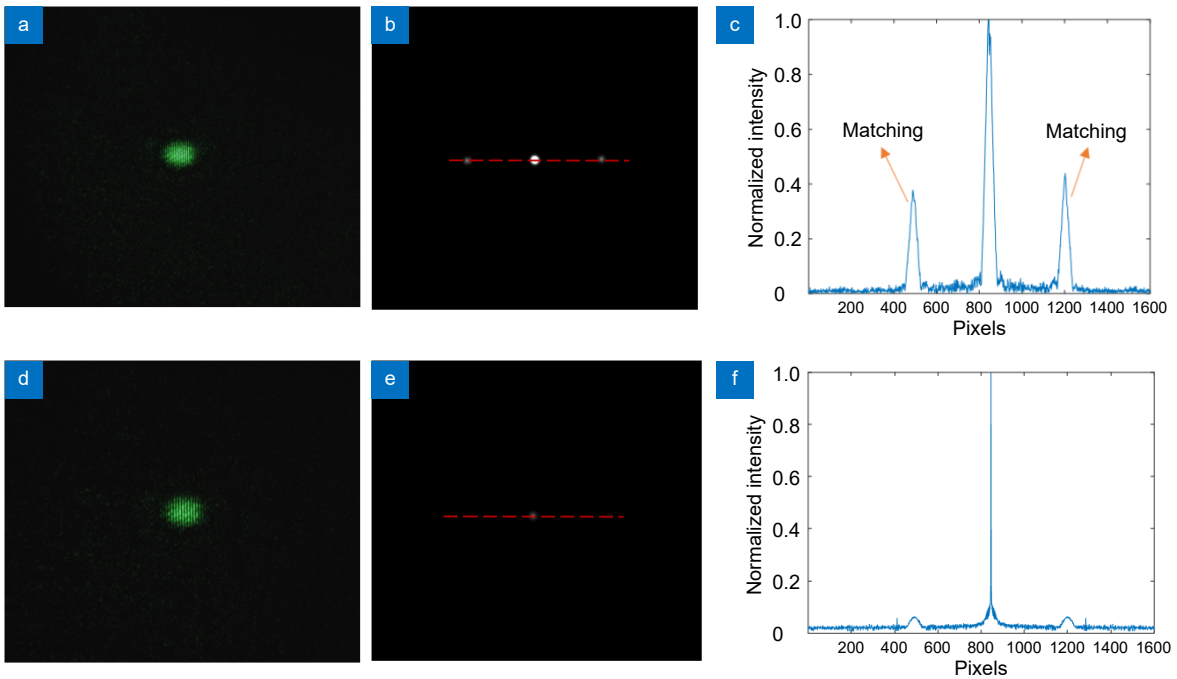

Fig. 11 | (**a**) JTS intensity patterns detected by camera corresponding to the joint input in Fig. 9(c). (**b**) FFT output of (a). (**c**) Normalized intensity plot along the red dashed lines in (b). (**d**) JTS intensity patterns detected by camera corresponding to the joint input in Fig. 9(d). (**e**) FFT output of (d). (**f**) Normalized intensity plot along the red dashed lines in (e).

Arbitrary cross-scale matching experiments

To further validate the robustness of the proposed method in real-world scenarios, where a fixed-scale reference is not always available, we conducted cross-scale matching experiments between arbitrary scale pairs. Specifically, for the 3D model "Sophie," we performed correlation tests between randomly selected scale combinations. In these experiments, the reference hologram

uses scale factor $a = b = \alpha$, and the query hologram uses scale factor $a = b = \beta$, with α and β chosen from the set $\{0.4, 0.5, 0.6, \cdots, 2.0\}$, including cases where $\alpha \neq \beta$.

Figure 12 shows representative correlation outputs for three different arbitrary scale pair combinations. In Fig. 12(a), the correlation is performed between the 3D face model at 0.4× scale (reference) and 0.7× scale (query), resulting in a clear and sharp peak at position (125, 125) with a normalized value of 1. Figure 12(b) presents the result of correlating the 0.5× scale reference with the 1.2× scale query, where the peak is located at (123, 123), also normalized to 1, demonstrating successful matching despite the scale difference. In Fig. 12(c), the correlation between the 1.5× scale reference and the 2.0× scale query yields a peak at (127, 127), again with a normalized peak value of 1, confirming robust matching at larger scale factors.

The consistent appearance of a single dominant peak in all cases confirms that the proposed method can reliably identify and match 3D faces across arbitrary scale pairs without relying on a standard template at $a = b = 1.0$. The slight positional shifts in the correlation peak (e.g., ±2 pixels) are expected and consistent with the theoretical prediction in Eq. (8), where the shift corresponds to the logarithmic scale factors used in the Mellin domain correlation.

Figure 13 presents the hybrid optical/digital joint transform correlation (JTC) experimental results corresponding to the cross-scale pair with scale factors 1.5× vs. 2.0×. In Fig. 13(a), the captured image of the JTS intensity pattern shows the result recorded by the camera after passing through the Fourier transform lens. The SLM displays two phase-only holograms of the same 3D face model at different scales, specifically at 1.5× and 2.0×, placed side by side as joint inputs. Figure 13(b) shows the Fourier transform of the captured JTS, revealing the matched correlation outputs. The two bright spots located on either side of the origin represent the two autocorrelation terms between the scaled holograms, with the red dashed line indicating the horizontal axis along which the intensity profile is extracted. In Fig. 13(c), the normalized intensity profile along the red dashed line is shown, revealing two distinct autocorrelation peaks symmetrically located on either side of the origin. These peaks correspond to the successful correlation between the 1.5× and 2.0× scaled holograms, and the sharpness and intensity of the peaks confirm the robustness of the matching despite the scale difference.

These results demonstrate that the proposed system achieves true scale-invariant 3D face recognition, maintaining high correlation contrast even when the reference and query data are at completely different scales.

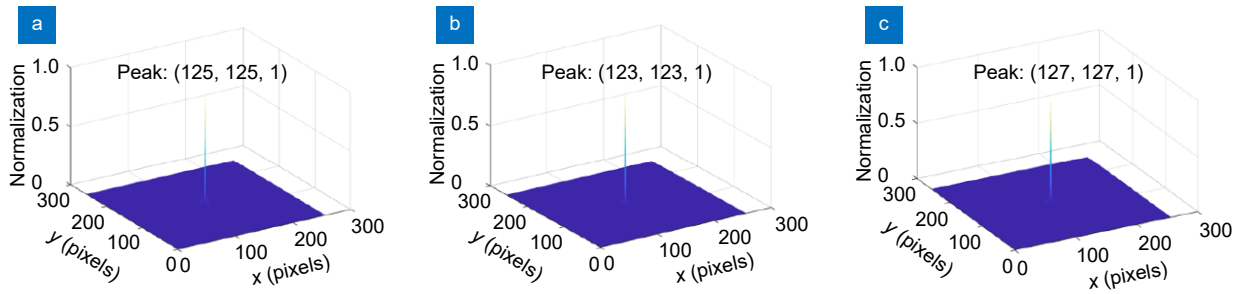

Fig. 12 | (**a**) Correlation result at scale 0.4× (reference) vs. 0.7× scale (query). (**b**) 0.5× (reference) vs. 1.2× scale (query). (**c**) 1.5× (reference) vs. 2.0× scale (query).

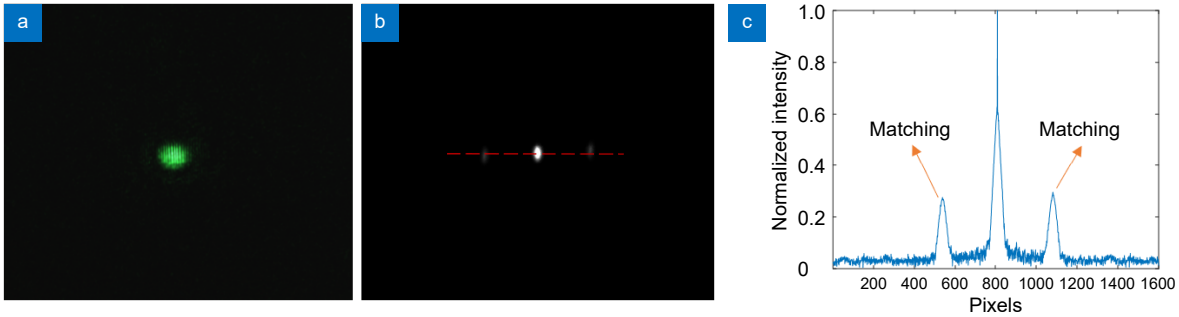

Fig. 13 | (**a**) JTS intensity patterns detected by camera corresponding to the joint input with scale factors 1.5× vs. 2.0×. (**b**) FFT output of (a). (**c**) Normalized intensity plot along the red dashed lines in (b).

This property is crucial for practical deployment, where subjects may appear at variable distances in uncontrolled environments.

Conclusions

In this study, we have demonstrated the effectiveness of scale-invariant 3D face recognition by leveraging computer-generated holograms and the Mellin transformation. Our method preserves the scale-invariant characteristics of 3D face holograms, ensuring consistent recognition performance regardless of different scale factors. Testing results, both purely computational and hybrid, confirm that the proposed technique of computer-generated holograms with the Mellin transform exhibits robust correlation peaks which are unaffected by the changes in scale of a 3D face. Through our investigations, we have discovered an unexpected benefit: applying the Mellin transform for correlation serves as a crucial and important preprocessing step, significantly enhancing the correlation peak, as evidenced by the comparison between Fig. 6(c) and Fig. 6(f). These findings highlight the potential of our approach for accurate and reliable 3D face recognition, even when the input faces are subject to scale variations. The outcomes of this research provide a solid foundation for further advancements in scale-invariant biometric recognition systems, emphasizing the practicality and efficiency of integrating computer-generated holography with the Mellin transform in real-world applications.

We would like to note that although we only used triangular meshes for 3D surfaces in this work, our approach is also applicable when the 3D face is decomposed with different methods such as isogeodesic strips[27]. Additionally, while we have utilized holograms (holistic information of a 3D face) for correlation, RGB-D images can also provide precise 3D information when combined with the Mellin transform[28]. We have not used any specific features for 3D recognition as our approach leverages holistic information from 3D faces in the form of complex holograms generated from its 3D meshes, which is a unique aspect of the technique proposed here. Future work will focus on addressing practical challenges—such as depth-induced defocus, spot size variation, and system imperfections—through the integration of advanced preprocessing techniques and exploration of hardware enhancements. Additionally, evaluations will be extended to larger 3D face datasets using standardized biometric performance metrics, aiming to comprehensively validate the robustness, scalability, and real-world applicability of the proposed method.

References

1. Guo YL, Wang HY, Wang LG et al. 3D face recognition: two decades of progress and prospects. *ACM Comput Surv* **56**, 54 (2024).
2. Bowyer KW, Chang K, Flynn P. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Comput Vis Image Underst* **101**, 1–15 (2006).
3. Ming X, Wei FY, Zhang T et al. Group sampling for scale invariant face detection. *IEEE Trans Pattern Anal Mach Intell* **44**, 985–1001 (2022).
4. Soltanpour SM, Wu QMJ. Weighted extreme sparse classifier and local derivative pattern for 3D face recognition. *IEEE Trans Image Process* **28**, 3020–3033 (2019).
5. Mian AS, Bennamoun M, Owens R. Keypoint detection and local feature matching for textured 3D face recognition. *Int J Comput Vis* **79**, 1–12 (2008).
6. Samir C, Srivastava A, Daoudi M. Three-dimensional face recognition using shapes of facial curves. *IEEE Trans Pattern Anal Mach Intell* **28**, 1858–1863 (2006).
7. Huang D, Ardabilian M, Wang YH et al. 3-D face recognition using eLBP-based facial description and local feature hybrid matching. *IEEE Trans Inf Forensics Secur* **7**, 1551–1565 (2012).
8. Hariri W, Tabia H, Farah N et al. 3D face recognition using covariance based descriptors. *Pattern Recognit Lett* **78**, 1–7 (2016).
9. Ali ASO, Asirvadam VS, Malik AS. Scale- invariant face recognition using triangular geometrical model. In *2013 IEEE International Conference on Signal and Image Processing Applications* 396–401 (IEEE, 2013). http://doi.org/10.1109/ICSIPA.2013.6708039.
10. Li ZS, Sun JS, Fan Y, Jin YB, Shen Q et al. Deep learning assisted variational Hilbert quantitative phase imaging. *Opto-Electron Sci* **2**, 220023 (2023).
11. Jabberi M, Wali A, Neji B et al. Face ShapeNets for 3D face recognition. *IEEE Access* **11**, 46240–46256 (2023).
12. Casasent D, Psaltis D. Position, rotation, and scale invariant optical correlation. *Appl Opt* **15**, 1795–1799 (1976).
13. Gabor D. A new microscopic principle. *Nature* **161**, 777–778 (1948).
14. Zhang YP, Poon TC. *Modern Information Optics with MATLAB* (Cambridge University Press, Cambridge, 2023).
15. Zhou HW, Sui XM, Cao LC et al. Digital correlation of computer-generated holograms for 3D face recognition. *Appl Opt* **58**, G177–G186 (2019).
16. Wang S, Wang Y, Jin M et al. Conformal geometry and its applications on 3D shape matching, recognition, and stitching. *IEEE Trans Pattern Anal Mach Intell* **29**, 1209–1220 (2007).
17. Martino JMD, Suzacq F, Delbracio M et al. Differential 3D facial recognition: adding 3D to your state-of-the-art 2D method. *IEEE Trans Pattern Anal Mach Intell* **42**, 1582–1593 (2020).
18. X. D. Gu, "Xianfeng David Gu's homepage," https://www3.cs.stonybrook.edu/~gu/ (2023). Accessed: 2025-03.
19. Tsang PWM, Poon TC, Wu YM. Review of fast methods for point-based computer-generated holography [invited]. *Photonics Res* **6**, 837–846 (2018).

20. Chen JS, Chu DP. Improved layer-based method for rapid hologram generation and real-time interactive holographic display applications. *Opt Express* **23**, 18143–18155 (2015).
21. Zhang YP, Fan HX, Wang F et al. Polygon-based computer-generated holography: a review of fundamentals and recent progress [invited]. *Appl Opt* **61**, B363–B374 (2022).
22. Yao YW, Zhang YP, Poon TC. Layer-mesh-based holograms for fast generation and high-quality reconstruction. *Opt Lasers Eng* **175**, 108027 (2024).
23. Poon TC, Wu MH, Shinoda K et al. Optical scanning holography. *Proc IEEE* **84**, 753–764 (1996).
24. Poon TC. *Optical Scanning Holography with MATLAB®* (Springer, New York, 2007).
25. Weaver CS, Goodman JW. A technique for optically convolving two functions. *Appl Opt* **5**, 1248–1249 (1966).
26. Yang YQ, Forbes A, Cao LC. A review of liquid crystal spatial light modulators: devices and applications. *Opto-Electron Sci* **2**, 230026 (2023).
27. Berretti S, Del Bimbo A, Pala P. 3D face recognition using isogeodesic stripes. *IEEE Trans Pattern Anal Mach Intell* **32**, 2162–2177 (2010).
28. Zhu XY, Yu C, Huang D et al. Beyond 3DMM: learning to capture high-fidelity 3D face shape. *IEEE Trans Pattern Anal Mach Intell* **45**, 1442–1457 (2023).

Acknowledgements

We are grateful for financial supports from the National Natural Science Foundation of China (Grant No. 62275113; 62405124).

Author contributions

Y.W. Yao developed the initial concept for the manuscript under the guidance of Professors Y.P. Zhang, D.P. Chu and T.-C. Poon. Y.P. Zhang supervised the project. Y.W. Yao and H.R. He performed experiments. Y.W. Yao and David X.F. Gu analyzed the data. Y.P. Zhang, D.P. Chu and T.-C. Poon reviewed and edited the manuscript. All authors commented on the manuscript.

Competing interests

The authors declare no competing financial interests.