Computed stereo lensless X-ray imaging

J. Duarte[1], R. Cassin[1], J. Huijts[1], B. Iwan[1,2], M. Kholodtsova[1], F. Fortuna[3], L. Delbecq[3], H. Chapman[4], M. Fajardo[5], M. Kovacev[2], W. Boutu[1] and H. Merdji[1*]

[1]LIDYL, CEA, CNRS and Université Paris-Saclay, CEA Saclay 91191 Gif-sur-Yvette, France.
[2]Leibniz Universität Hannover, Institut für Quantenoptik, Hannover, Germany
[3] CSNSM, CNRS/IN2P3 and Université Paris-Saclay, 91405 Orsay, France
[4] Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany.
[5] Instituto de Plasmas e Fusão Nuclear, IST Lisboa, Portugal

[*]correspondence to: hamed.merdji@cea.fr

Recovering 3D properties of artificial or biological systems at low X-ray dose is critical as most techniques are based on computing hundreds of 2D projections. This prevents single-shot 3D imaging using ultrafast X-ray sources. Here we show that computed stereo-vision concepts can be transposed to X-rays. Stereo vision is of great importance in the field of machine vision and robotics. Here, two X-ray stereo views are reconstructed from coherent diffraction patterns and a nanoscale 3D representation of the sample is computed from disparity maps. Similarly to the brain perception, computed stereo vision algorithms use constraints. We demonstrate that phase contrast images relax the disparity constraints and allow revealing occulted features. We also show that label nanoparticles further extend the applicability of the technique to complex samples. Computed stereo X-ray imaging will find applications at X-ray free electron lasers, synchrotrons, laser-based sources as well as for industrial and medical 3D diagnosis.

In nature, most objects possess complex three-dimensional dynamical structures, whose deep understanding is crucial in several fields of study. The large development of ultrafast coherent X-ray sources allows 2D single-shot imaging on nanometre-femtosecond scale using lensless imaging techniques, widely developed on small-to-large-scale facilities[1-6]. The extension to ultrafast 3D imaging is, however, challenging. Nowadays, 3D nanometre-scale imaging techniques are mainly based on computed tomography. The sample is rotated with respect to the illumination source, allowing for a full set of 2D projections, which are recombined to form a 3D image[7,8,9]. However, such achievement requires hundreds of views, making it incompatible with imaging of ultrafast processes or dose-sensitive samples[10]. To allow imaging extended objects, ptycho-tomography has been proposed[11,12]. While leading to impressive 3D resolutions, this technique is also extremely demanding in terms of number of required projections[13]. Imaging before destruction of single particles, as proposed on X-ray FELs, overcomes the radiation dose problem[14]. Nevertheless, this technique requires a huge number of identical samples and generates an extremely large amount of data that needs to be sorted, classified and combined to provide a full set of consistent 3D data[15]. There is an intensive work on decreasing the number of orientations, an extreme solution being stereo imaging. Although X-ray stereoscopy was discovered in the end of the 19th century[16], it didn't find wide scientific applications immediately. Recently, however, electron stereopsis microscopy has shown to produce unprecedented 3D perception of nanometre-scale details[17]. The main drawback about this approach is that the 3D effect is purely physiological. Indeed, the human brain can get a fast 3D perception of the sample by processing binocular disparities in the cortex region, but without quantitative depth information. Moreover, to allow the cognitive 3D reconstruction, the angle between the two views has to be small, limiting the gain in structural information. Several experiments have taken place at synchrotron beamlines using stereo imaging, but none have achieved a 3D reconstruction stemming from a single shot pair of images[18–20]. In 2008, Schmidt *et al.*[21] proposed a theoretical study of a method dividing an X-ray FEL beam into two sub-beams using a crystal. In 2014, Gallagher-Jones *et al.*[22] probed the 3D structure of an RNAi microsponge by combining coherent diffractive imaging (CDI) reconstructions from successive single-shot diffraction patterns from an X-ray FEL, and from X-ray diffraction from synchrotron. However, this method requires several acquisitions at multiple angles. Techniques to retrieve the 3D structure from a single diffraction pattern have also been proposed, but they work under limited circumstances and heavily rely on sample *a priori* knowledge[23–27]. To date, it has not been possible to obtain a 3D reconstruction from a single X-ray acquisition. Still, stereoscopic coherent imaging has been proposed as a future and promising technique for nanoscale fast-time-frame 3D imaging at X-FELs[28]. Here we propose to extend the *Computer Stereo Vision*[29] concept from the visible to X-rays. Instead of constructing a stereo anaglyph with only qualitative 3D perception, our approach retrieves depth information by computing disparity maps from two CDI stereo views. Here, we illustrate through different examples the applicability and the limitations of the technique.

Femtosecond X-ray stereo imaging

First, we demonstrate single shot stereo imaging using a soft X-ray optical setup based on a high harmonics (HH) beam separated into two coherent twin beams, which illuminate a sample with a controllable angle (see Fig. 1). The setup enables recording in a single acquisition two stereo diffraction patterns, reaching nanometre-transverse resolution, on a femtosecond timescale, without *a priori* knowledge of the sample. Details on the HH beamline can be found in the Methods section. To generate the two sub-beams, we insert a grazing incidence prism between the off-axis parabola and the sample. Two silicon mirrors are adjusted such that the two beam foci overlap on the sample, with a controllable angle. In this versatile setup, the angle between the two beams can be easily changed, by tilting and moving the plane mirrors, allowing for an adaptive geometry to the sample under study. An additional advantage of this setup is the possibility to control the temporal overlap between the two beams, with attosecond accuracy, enabling 3D X-ray pump / X-ray probe experiments. In our work, the two X-ray beams are diffracted by the sample and the far field patterns are recorded simultaneously using a single CCD camera. A typical set of stereo diffraction patterns acquired at a separation angle of 19° is shown in Fig. 2a. The two patterns exhibit an overlap in the high-spatial frequency regions, which does not affect the reconstructions as the useful diffraction information is extracted from a smaller area. The number of useful photons in each diffraction pattern is roughly equivalent (few 10^7 photons per shot).

Each diffraction pattern of Fig. 2a is isolated and inverted independently using a *difference map* algorithm[30], a generalization of the hybrid input-output algorithm. Figures 2b-c show the amplitude reconstructions of the diffraction patterns corresponding to the left and right views, respectively, of the sample of Fig. 1b. They represent the coherent average of 45 independent runs of the phase retrieval algorithm. The spatial resolution of each view is estimated by the 10-90% criteria to be 127 nm. Differences between the two views are clear: all the edges of the cross-lid structure are visible in Fig. 2c, whereas some of them are hidden in Fig. 2b.

Qualitative 2D structural and spatial information from two observation angles is achieved in a single femtosecond acquisition. However, it is possible to go further and recover depth information from those images. Indeed, from the pair of reconstructed views of the sample, one can compute the disparity map. Disparity refers to the distance between two corresponding points in the two images of a stereo pair. By matching each pixel from one image to the other, one can calculate the distance between them to produce an image where each pixel represents the disparity value for that pixel with respect to the corresponding image. The disparity map can then be converted into depth information by a simple equation, given the geometry of our setup:

$$z(P,\theta) = \frac{d(P)}{tan\theta_1 + \ tan\theta_2}\ . \ (1)$$

In equation (1), z is the relative depth value of the point $P(x,y,z)$ of the object, $d(P)$ is its disparity value and θ_1 , θ_2 are the angles between the line perpendicular to the CCD and each stereo beam, respectively. From eq. (1) one can notice that the voxel size on the depth axis decreases with the angle between the two illuminating beams. However, there is an upper limit for this angle, which is not straightforward to determine as it depends on the sample structure. Indeed, a strong constraint of the computed stereo algorithm imposes that identical features can be identified in both views, to be able to calculate the corresponding disparity values. The achieved pixel matching allows then to project the 2D information into 3D voxels.

In order to minimize the artefacts due to the pure amplitude nature of the sample transmission, Figs. 2b-c were converted into binary images. The disparity calculations were thus limited to the edges of the sample, which allow non-ambiguous pixel correspondences (see Methods for details). The intermediate depth values were then retrieved from 3D interpolation routines. Figures 3a-b show, respectively, the experimental disparity map and the 3D stereo reconstruction obtained from the image reconstructions shown in Figs. 2b-c. Our geometry leads to a voxel size of 49x49x146 nm^3 and an estimated spatial resolution of 127x127x379nm^3 (see details in Methods).

Analysing the 3D reconstruction of Fig. 3b, one can note that the cross-shaped structure is clearly visible, however it presents some artefact connections to the membrane in the areas where the different planes superimpose. This stems from the fact that the sample is a pure-amplitude transmission object. Indeed, this induces shadow effects as a result of occulted areas in the projection geometry, where surface orientation and edges are under-determined. Occluding contour artefact is a standard problem in vision science, which can be solved by adding additional 2D views as constraint to surface orientation determination. Multi-view stereo imaging is able to reconstruct detailed 3D information of a scene and is fully exploited in reflective stereo photometry[31]. However, in the context of X-ray vision, and due to optical indexes in this spectral domain, the reflectivity of a scene can be quite poor. As an alternative to multi-view approach, we propose to use phase contrast images that exploit the transparency of a sample to X-rays, available for example at XFELs.

Computed X-ray phase contrast stereo imaging

We have performed simulations using the same "cross" sample but adding non-zero transmission to the probe beams. To simulate the different phase shifts, distinct absorption values are attributed to the central cross and to the membrane, resulting in stereo views composed by different grey tones. Since the images are not obtained by a parallel camera system, an image rectification step is necessary (see Supplementary Section 1.2). Rectifying a pair of stereo images requires a set of point correspondences between the two views, which is often accomplished by combining feature detection and feature matching algorithms (see methods). After the rectification process, the disparity maps are computed, employing the same method used for the experimental data (see Fig. 3c). Compared to Fig. 3a, which shows large discontinuities in the disparity values, Fig. 3c shows a continuous sampling of the disparity along the object contour. 2D projections often superimpose objects at different depths on each other and subtle differences in the projections may be invisible or completely lost. Here, two phase-contrast images are registered and allow recovering "behind scene" information with a high fidelity. A 3D phase stereo reconstruction can be computed and is reported in Fig. 3d (see also supplementary movie 1). The information on the phase shifts unveils the existence of superimposed planes, which allows retrieving 3D features otherwise hidden by the membrane. Overall, we obtain a satisfying 3D rendering with accurate details about the structure of the sample. Phase contrast stereo imaging can be further exploited when sparse 2D information is obtained from the sample. Indeed, amplitude and phase views from the sample generate redundant information that can be exploited cooperatively to satisfy stereo algorithm constraints.

Computed X-ray stereo imaging using nanoparticle labels

We then extend the computed stereo vision concept to harder X-rays and more complex samples. Our demonstration exploits two views from tomographic data published and made available by Chapman et al.[9].

The sample is composed by an arrangement of 50-nm-diameter gold spheres spread on a micron height Si_3N_4 pyramid (Figs. 4a-b). A complete set of 3D tomographic coherent diffractive imaging data were measured at a photon energy of 750 eV. In our work, we exploit only two 2D CDI reconstructions recorded at a separation angle of 7^{o} (see Figures 4c-d). The spatial resolution of each view is estimated to be 42 nm. Figure 4e displays the reconstructed experimental disparity map. From this input we then compute the 3D stereo depth-map shown in Fig. 4f and supplementary movie 2. Both the disparity map and the 3D calculations are obtained without any *a priori* knowledge. This step was facilitated by the high signal-to-noise ratio of the reconstructed views provided by the high-scattering factors of the gold nanoparticles. The stereo reconstruction reproduces well the height of the pyramid and the 3D distribution of nanoparticles. However individual ones cannot be distinguished. Indeed, the stereo reconstruction shown in Fig. 4f is achieved with a 3D voxel size of 15x15x120 nm^3 and an estimated spatial resolution of 42x42x343 nm^3, which is larger in depth than the size of a nanoparticle. However, the dashed red circles shown in Fig. 4a and 4f highlight a feature corresponding to a depth of δz = 105 nm, which is retrieved even though representing dimensions below the estimated resolution value. Nanoparticles offer high contrast to X-rays and can be used to label the 3D surface of structures, or, in a dense-distribution configuration, their volume. Here, the uniqueness constraint of the computed stereo technique is achieved by the nanoparticles spatial localisation, which provides a solid convergence of the stereo-matching algorithm.

Conclusions

The ability to directly measure and evaluate 3D shapes with unprecedented time resolution and reliability extends our knowledge about the nature and function of complex systems. In human vision, 3D perception is generated from a stereopsis reconstruction performed by our brain. In computer stereo vision, the depth information is extracted after several operations: image rectification, disparity map estimation and 3D point cloud projection. Here, we have extended the computer stereo vision concept to X-rays and performed 3D reconstructions at nanometre scale, using single acquisition. Obtaining accurate and realistic disparity maps requires that the two stereo angles have a good overlap between the same details on the object. This can be difficult to obtain with pure amplitude objects observed in transmission. Indeed, to retrieve the stereo disparity, computer stereo vision relies on computational constraints, which are based on similarity, uniqueness and continuity assumptions[29]. We demonstrated in this article that X-ray phase contrast images allow to fulfil these constraints and that occulted scenes can be recovered by using phase contrast views as stereo pairs. However, the method shows difficulties to trace the depth of structures with spherical topographies and smooth composition gradients, which are ambiguous in transmission geometry. Nevertheless, using labels, more complex 3D structures can be retrieved without the need of pre-processing images or including any *a priori* knowledge. Here gold nanoparticles were used to sample the disparity and a final 3D representation of a nano-pyramid was achieved. For future applications such nanoparticles, similarly to labels in fluorescence microscopy[32-34], can be used to follow complex soft features found, for instance, in biological systems. Overall, the proposed stereo method exploiting only two views considerably lowers the acquisition time and the dose delivered to the sample. Furthermore, associated to lensless imaging techniques, it allows following *in vivo* ultrafast 3D motions at a nanometre scale. For real-time 3D vision at high repetition X-ray sources, such as X-FELs, the image processing can be further optimized by using adaptive compressed stereo images, based on redundancies on the two views[35]. Note that the realisation of synchronized hard X-ray stereo beams is proposed at the European XFEL[36] and a similar beamline is under design at SACLA XFEL, Japan[37]. Computed stereo imaging can have a tremendous impact in 3D structural imaging of single macromolecules for example. Indeed, using hard X-ray coherent diffractive imaging to recover the stereo views with atomic resolution, specific atoms can be used as local-3D probes to fully match the stereo algorithm constraint. With such potential, we foresee outstanding applications in biology, materials science, medicine or even in the industry.

Methods

Experimental setup. The experiment was performed at the SLIC laser centre in CEA Saclay. The high-order harmonic beam setup is described in detail elsewhere[4]. We generate the harmonic beam using an amplified Ti:Sa laser system, which delivers 30 mJ, 60 fs pulses, using a loose focusing geometry with a focal length lens of 5.65m. The XUV beam propagates collinearly with the driving IR laser, which is attenuated using IR antireflective silica plates in grazing incidence. After optimization, we reach 4.10^9 photons/pulse for harmonic 33 in neon with a spectral bandwidth, $\lambda/\Delta\lambda$, of 150 and 20 fs pulse duration. A 22.5° off-axis-parabola of 20 cm focal length focuses the harmonic beam to a 5x7 μm^2 focal spot (FWHM) and selects harmonic 33 (λ= 24 nm) thanks to a multilayer coating deposited on its surface. The sample is positioned at the parabola's focus, and the CCD detector (2048x2048 pixels, pixel size 13.5 µm) is located 26 mm away. Using the sharp edge of the prism, the HH beam is split into two half-beams. Each one is reflected back towards the sample by a pure silicon plate. The prism and silicon plates setup (Fig. 1a) is inserted after the parabola in order to increase the angle between the two sub-beams, otherwise limited by the parabola aperture. The focus of each stereo spot is then enlarged (compared to the direct focusing) to 10x7 μm^2. Note that the whole setup could be placed before the focusing optics, provided that the angle between the two focused beams is large enough. The positions and tilts of the two plates are remotely controlled by vacuum compatible motors, offering the possibility to vary the angle between the two beams. The XUV transmission of the apparatus was estimated to about 75% at a 24 nm wavelength.

Sample preparation. The sample (Fig. 1b) is a 6.9x6.1 μm^2 cross, drilled on a membrane (75 nm of Si_3N_4 with a 150 nm Au layer and 4 nm of Cr for adhesion) using a focused ion beam. We first patterned the outer edges of the cross with a low gallium ion current. The soft patterning allows controlling the attachment of the inner cross to the edges. Then, electrostatic forces prevented the lid from falling and "attached" it permanently to the membrane at two opposite contact points.

Data acquisition and reconstruction. Although the experiment is single-shot compatible, we compensate the low dynamic range of the CCD and increase the available photon flux by using the *high dynamic range* (HDR) technique[38]. Therefore, two sets of diffraction patterns are recorded with integration times of 30 s and 240 s. First, we crop the overlapped part between the two diffraction patterns in order to isolate each projection. Then the central part of the short acquisition pattern is recombined with the higher frequencies of the one with large integration time, using a *Gaussian* filter (σ=2) on the edges to smooth the transition. A ratio, extracted from both corresponding non-saturated regions, is applied to appropriately rescale the intensity values. Both diffraction patterns are then reconstructed using a difference map algorithm[31]. For each projection, we launch 45 independent runs of the algorithm and select for each run the iteration that minimizes the error function $\epsilon_n = \sqrt{\sum_{i=1}^{N} |d_i(x_n)|^2}$, where ϵ_n is the distance between the two successive estimates x_{n+1} and x_n. In our case the algorithm converges after roughly a hundred iterations. The DFT

(Discrete FT) image registration algorithm[39] is then used to superpose the 45 best reconstructions and average them.

Pre-processing of the CDI stereo views. Giving that the computed stereo imaging technique is based on pixel matching, a fundamental requirement is that similar illumination conditions apply to both views. This is, however, challenging with transmission geometry, as the pixel amplitudes depend on the viewing axis. Indeed, magnitude variations can arise either from crossing different materials with various refractive indexes or from differences in material thickness along the imaging axis. Moreover, experimental issues such as beam's non-uniform intensity profile or partial coherence can also have an impact, as well as algorithm-related artefacts, especially in low-photon-flux regimes. To avoid depth calculation errors due to all these phenomena, our approach relies on identifying non-ambiguous features to retrieve the depth information. In the cases where these features are too sparse as, for instance, objects with spherical structure or smooth gradients, the intermediate depth values are retrieved through 3D interpolation methods, upon crossing the information of the achieved 3D shape with the amplitude/phase stereo views.

Consequently, depending of the quality of the data and the shape of the features present in the sample, additional pre-processing may be mandatory. For the computed 3D reconstruction of the cross-shaped structure, because of the limited SNR in the reconstructed stereo pair, a first step of binary conversion is required. Therefore, first, the images are resized by a factor 4, with a *bicubic* interpolation in the intermediate pixels. A Gaussian low pass filter is then applied to both images to reduce the effect of the noise (σ= 1.9). After filtering, the images are turned binary by defining suited binary thresholds (threshold values in a 0-to-1 scale: 0.40 and 0.25, in left and right images, respectively). Finally, to avoid errors in the binary conversion, the isolated regions constituted by less than 400 aggregated pixels, with no correspondence in the pair image, are removed.

The nanoparticles labelled sample only required background filtering. Hence, a binary mask including populated regions of the pyramid structure is multiplied by the stereo views of Figs. 4c-d, in order to avoid calculating random disparity values over the background.

Computed 3D reconstruction of the cross-shaped structure. For the HHG setup, since both views are recorded in the same CCD camera, no image rectification is needed (see Supplementary Section 1.2). The disparity maps are calculated employing a simple block matching routine. The images are divided into blocks of 3x3 pixels and, for each picture block, a scan is made over blocks of the same size in the pair picture. The scan is allowed 65 pixels to the left (negative disparity) and to the same amount to the right (positive disparity) of the block central pixel position. A simple sum of absolute differences, added to a less weighted pixel proximity term, is employed as a cost function, to select the best match from the set of candidate blocks (see Supplementary Section 1.3). A second-order sub-pixel interpolation is realized over each disparity value which results from the sorting process. Note that the disparity values are only retrieved over the edges of our 3D structure since it is ambiguous to find matching blocks in the uniform black/white regions. Using this method, two disparity maps are generated, representing the disparity of the left image with respect to the right one (left map, Fig. 3a) and vice versa.

The 3D information is extracted from the disparity maps by employing equation (1), deduced from the camera geometry. For matching the information of the two disparities, a coordinate correction is required. Hence, the x_1 and x_2 coordinates of the left and right disparity maps, respectively, are converted to the object coordinate x. This conversion is obtained from the relations $x\,(x_1, z,\, \theta_1) = \,x_1 - z\, tan\, \theta_1$ (on the left map) and $x\,(x_2, z,\, \theta_2) = \,x_2 - z\, tan\, \theta_2$ (right map).

After retrieving all the coordinates of the points in the 3D space, the stereo consistency of the two disparity maps is evaluated. In this step, the 3D points extracted from each disparity map whose coordinates do not have a match for the second map are discarded (see Supplementary Section 1.4).

From the remaining data, a point cloud is created and the outlier points are removed. A point is considered an outlier if the average distance to its *k*-nearest neighbours is above a specified threshold *t*. For all datasets, *k* = 80 points and *t* = 0.1. This threshold *t* specifies the number of standard deviations away from the estimated average distance. From the remaining points, the 3D shape of the sample is already visible, with the edged structures completely reconstructed.

Next we apply a process in which the information achieved from the 3D reconstruction and the direct stereo views are crossed to achieve the final 3D representation of the sample (see Supplementary Section 1.4). By fitting a 3D plane in the cross-shape cut of the membrane, a 3D plane surface is computed and a square frame with three points of length is added to the extremities of the point cloud, defining the limits of our 3D reconstruction. A 3D interpolation is, then, realized over the resultant point cloud to infer the intermediate values. The routine employed for the purpose is based on a *Natural Neighbour Interpolation* to the scattered sample points[40]. Crossing the information on the white regions of our stereo views (object transmission function equal to 1), an extra stereo point cloud is computed, composed by stacks of planar points, which we know to correspond to the empty volume of our sample. Excluding from the interpolated 3D mesh the neighbours (0.2 micrometre precision) of the empty region cloud, we reach a final 3D reconstruction of the sample (Fig. 3b).

Computed phase stereo imaging process. The two stereo views have a separation angle of 12° (see Supplementary Section 2.1). In the simulated stereo images, sixteen edge points are manually selected (see Supplementary Section 2.2). Using all the selected points, the fundamental matrix between both views is computed employing the *Normalized Eight-Point Algorithm*[41]. The images are then re-projected, in order to make all the matching points lay in the same horizontal lines[42].

Note that besides the two disparity maps obtained from the direct rectified views, two more are calculated. These intend to target specifically the edge areas, allowing for pixel matches in regions which show superposition of different structures in different views. Therefore, a directional gradient along the x-axis[43] – direction of disparities - is applied to the rectified stereo views and two new stereo views are generated with a clear highlight on the edges. Two new disparity maps are then extracted from these views, possessing additional information on the structures (see Supplementary Section 2.3). After discarding the inconsistent points between the right and left disparity maps for both cases, the resultant point clouds are merged.

The next step consists in crossing the information of the 3D point cloud and the direct phase images (see Supplementary Section 2.4). Due to the reduced number of views and the existence of superimposed structures, it is necessary to identify isolated sample components and address each component individually. For this step, image segmentation tools and gradient calculations can be used to automatize the process. After identifying the structures, one should fit surfaces in each structure, according with its phase variations and use these surfaces to detect the outlier points. If necessary, some surface points can be added in the active point cloud to help with the 3D interpolation. Note that if the 3D interpolation is made directly for the full point cloud, one can have wrong connections between structures, due to the lack of information in the internal regions.

In the specific case of the simulated sample, since the phase is flat, two planes are fitted to the achieved point cloud. Two new point clouds are then generated, each being constituted by the inlier points of a fitted plane. A 3D scattered interpolation is, then, performed to each cloud and the respective empty volumes removed. Note that all these steps follow the same lines explained for the experimental case. The final reconstruction of the sample is achieved from the assembly of the two 3D structures (Fig. 3d).

Computed 3D reconstruction of the metallic nanoparticles structure. For the gold nanoparticles pyramid, the images provided were already rectified. The disparity calculations were applied in the same terms as the previous case, but using blocks of 23x23 pixels and a disparity range of 30 pixels. The large window size allowed reducing the matching errors, while providing a smoother structure of the arrangement of the nanoparticles. Indeed, the aim was not to reconstruct individual nanoparticles, but the overall structure they were lining in. The same mask used for the pre-processing step was multiplied by the retrieved disparity maps, in order to clean the errors of the disparity over the edges, precisely allowing for the choice of these larger windows in the matching process. The 3D information was extracted employing equation (1). The stereo consistency of the two disparity maps was evaluated and the final point cloud was created after removal of the outlier points. All these steps followed the same process as the cases explored before. No 3D scattered interpolation was applied for the image of Fig. 4f in order to provide a better visualization of the structures superimposed in the observation axis.

Voxel size and resolution limit.

The voxel size is calculated by combining the XY pixel size, Δ_r, obtained from each phase retrieval reconstruction and its corresponding projection on the Z-axis given by:

$$\Delta_z = \frac{\Delta_r}{tan\theta_1 + tan\theta_2} \tag{2}$$

The spatial resolution is estimated by combining the XY spatial resolution obtained for each stereo scene (using 10-90% Rayleigh criterion) and the respective Z depth resolution estimated using eq. 2. Note that this estimation does not account for resolution degradations arising from the numerical processing.

Data availability.

The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

We acknowledge financial support from the European Union through the Future and Emerging Technologies (FET) Open H2020: Volumetric medical X-ray imaging at extremely low dose (VOXEL) and the integrated initiative of European laser research infrastructure (LASERLAB-EUROPE) (grant agreement no. 654148). Support from the French ministry of research through the 2013 Agence Nationale de Recherche (ANR) grants ”NanoImagine”, 2014 ”ultrafast lensless Imaging with Plasmonic Enhanced Xuv generation (IPEX)”, 2016 “High rEpetition rate Laser for Lensless Imaging in the Xuv (HELLIX)”; from the Centre National de Compétences en Nanosciences (C’NANO) research program through the NanoscopiX grant; the LABoratoire d’EXcelence Physique Atoms Lumière Matière - LABEX PALM (ANR-10-LABX-0039-PALM), through the grants ”Plasmon-X” and “HIgh repetition rate Laser hArmonics in Crystals (HILAC)” and, finally, the Action de Soutien à la Technologie et à la Recherche en Essonne (ASTRE) program through the “NanoLight” grant are also acknowledged. We acknowledge, as well, the financial support from the Deutsche Forschungsgemeinschaft, grant KO 3798/4-11, and from Lower Saxony through “Quanten- und Nanometrologie” (QUANOMET), project NanoPhotonik. We acknowledge Anton Barty and Yoshinori Nishino for data availability and discussions.

Authors contribution

J. D., R. C., J. H., W. B., B. I and H. M. carried out the experiment. The samples were produced by F.F., L.D and W.B.. H.C. provided the pyramid 2D images. Simulations were performed by J.D. and R.C.. Data analysis was performed by J. D., R. C., J. H., B. I., M. K. M. F., H. C., W. B. and H.M.. W.B., M.K. and H.M. proposed the physical concept. All authors discussed the results and contributed to the writing of the manuscript.

Competing interest

The authors declare no competing financial and non-financial interest.

Additional information

Correspondence and requests for materials should be addressed to H.M..

References

1. Seibert, M. M. et al. Single mimivirus particles intercepted and imaged with an X-ray laser *Nature* **470**, 78-81 (2011).

2. Redecke, L. et al. Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an X-ray laser. *Science* **339**, 227-230 (2013).

3. Ravasio, A. et al. Single-Shot Diffractive Imaging with a Table-Top Femtosecond Soft X-Ray Laser-Harmonics Source. *Phys. Rev. Lett.* **103**, 028104 (2009).

4. Gauthier, D. et al. Single-shot femtosecond X-ray holography using extended references. *Phys. Rev. Lett.* **105,** 093901 (2010).

5. Barty, A. et al. Ultrafast single-shot diffraction imaging of nanoscale dynamics. *Nat. Photon.* **2,** 415–419 (2008).

6. Miao, J. et al. Beyond crystallography: Diffractive imaging using coherent x-ray light sources. *Science.* **348,** 530-535 (2015).

7. Nishino, Y., Takahashi, Y., Imamoto, N., Ishikawa, T. & Maeshima, K. Three-Dimensional Visualization of a Human Chromosome Using Coherent X-Ray Diffraction. *Phys. Rev. Lett.* **102**, 018101 (2009).

8. Takahashi, Y. et al. Three-Dimensional Electron Density Mapping of Shape-Controlled Nanoparticle by Focused Hard X-ray Diffraction Microscopy. *Nano Lett.* **10**, 1922–1926 (2010).

9. Chapman, H. N. et al. High-resolution ab initio three-dimensional x-ray diffraction microscopy. *J. Opt. Soc. Am. A* **23**, 1179-1200 (2006).

10. Howells, M.R. et al. An assessment of the resolution limitation due to radiation-damage in x-ray diffraction microscopy. *J. Electron Spectrosc. Relat. Phenom.* **170**, 4–12 (2009).

11. Thibault, P. et al. High-Resolution Scanning X-ray Diffraction Microscopy. *Science* **321**, 379–382 (2008).

12. Dierolf, M. et al. Ptychographic X-ray computed tomography at the nanoscale. *Nature* **467**, 436-439 (2010).

13. Holler M. et al. An instrument for 3D x-ray nano-imaging. *Rev. Sci. Instrum.* **83**, 073703 (2012).

14. Chapman, H. N. et al. Femtosecond diffractive imaging with a soft-X-ray free-electron laser**.** *Nat. Phys.* **2**, 839–843 (2006).

15. Ekeberg, T. et al. Three-Dimensional Reconstruction of the Giant Mimivirus Particle with an X-Ray Free-Electron Laser. *Phys. Rev. Lett*. **114**, 098102 (2015).

16. Thomson, E. Stereoscopic Roentgen pictures. *Electr. Eng.* 21, 256 (1896).

17. Andruleit, H., Geisen, M. & Stäger, S. Stereo-microscopy of coccolithophores-modern applications for imaging and morphological analysis. *J. Nannoplankton Res*. **28**, 1–16 (2006).

18. Hoshino, M. et al. Development of an X-ray real-time stereo imaging technique using synchrotron radiation. *J. Synchrot. Radiat*. **18**, 569–574 (2011).

19. Gleber, S.-C., Thieme, J., Chao, W. & Fischer, P. Stereo soft X-ray microscopy and elemental mapping of haematite and clay suspensions. *J. Microsc*. **235**, 199–208 (2009).

20. Miao, J. et al. High Resolution 3D X-Ray Diffraction Microscopy. *Phys. Rev. Lett*. **89**, 088303 (2002).

21. Schmidt, K. E. et al. Tomographic femtosecond X-ray diffractive imaging. *Phys. Rev. Lett*. **101**, 115507 (2008).

22. Gallagher-Jones, M. et al. Macromolecular structures probed by combining single-shot free-electron laser diffraction with synchrotron coherent X-ray imaging. *Nat. Commun*. **5**, 3798 (2014).

23. Wei, H. Fundamental limits of 'ankylography' due to dimensional deficiency. *Nature* **480**, E1 (2011).

24. Wang, G., Yu, H., Cong, W. & Katsevich, A. Non-uniqueness and instability of 'ankylography'. *Nature* **480**, E2 (2011).

25. Thibault, P. Feasibility of 3D reconstructions from a single 2D diffraction measurement. Preprint at http://arXiv.org/abs/0909.1643v2 (2010)

26. Chen, C.-C. et al. Three-dimensional imaging of a phase object from a single sample orientation using an optical laser. *Phys. Rev. B* **84**, 224104 (2011).

27. Xu, R. et al. Single-shot three-dimensional structure determination of nanocrystals with femtosecond X-ray free-electron laser pulses. *Nature Comm.* **5**, 4061 (2014).

28. Yabashi, M. & Tanaka, H. The next ten years of X-ray science. *Nat. Photon.* **11**, 12-14 (2017).

29. Marr, D. & Poggio, T. Cooperative computation of stereo disparity, *Science* **194,** 283-287 (1976).

30. Elser, V. Phase retrieval by iterated projections. *J. Opt. Soc. Am. A* **20**, 40–55 (2003).

31. Furukawa, Y. & Hernández, C. Multi-View stereo: A tutorial. *Found. Trends Comput. Graph. Vis.* **9**, 1-148 (2013).

32. Longo, E. et al. 3D imaging of theranostic nanoparticles in mice organs by means of x-ray phase contrast tomography. *Proc. SPIE 10573*, 105734I (2018).

33. Longo, E. et al. 3D map of theranostic nanoparticles distribution in mice brain and liver by means of X-ray Phase Contrast Tomography. *J. Instrum.* **13**, C01049 (2018).

34. Menk, R. H. et al. Gold nanoparticle labelling of cells is a sensitive method to investigate cell distribution and migration in animal models of human disease. *Nanomed. Nanotechnol. Biol. Med.* **7**, 647–654 (2011)

35. Ortis, A., Rundo, F., Di Giore, G. & Battiato, S. Adaptive compression of stereoscopic images. *In Proc. International Conference on Image Analysis and Processing,* 391-399 (Springer, 2013).

36. Lu, W. et al. Development of a hard X-ray split-and-delay line and performance simulations for two-color pump-probe experiments at the European XFEL. *Rev. Sci. Instrum.* **89,** 063121 (2018).

37. Osaka, T. et al. Wavelength-tunable split-and-delay optical system for hard X-ray free-electron lasers. *Opt. Express* **24**, 9187-9201(2016).

38. Chen, B. et al. Multiple wavelength diffractive imaging. *Phys. Rev. A* **79**, 023809 (2009).

39. Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient subpixel image registration algorithms. *Opt. Lett*. **33**, 156–158 (2008).

40. Amidror, I. Scattered data interpolation methods for electronic imaging systems: a survey. *J. Electron. Imaging* **11**, 157–177 (2002).

41. Hartley, R. I. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, **19**, 580–593 (1997).

42. Papadimitriou, D. V. & Dennis, T. J. Epipolar line estimation and rectification for stereo image pairs. *IEEE Trans. Image Process.* **5**, 672–676 (1996).

43. Sobel, I. An isotropic 3×3 image gradient operator. *Machine Vision for three-dimensional Sciences*, 376-379 (Academic Press, 1990).

Figures captions

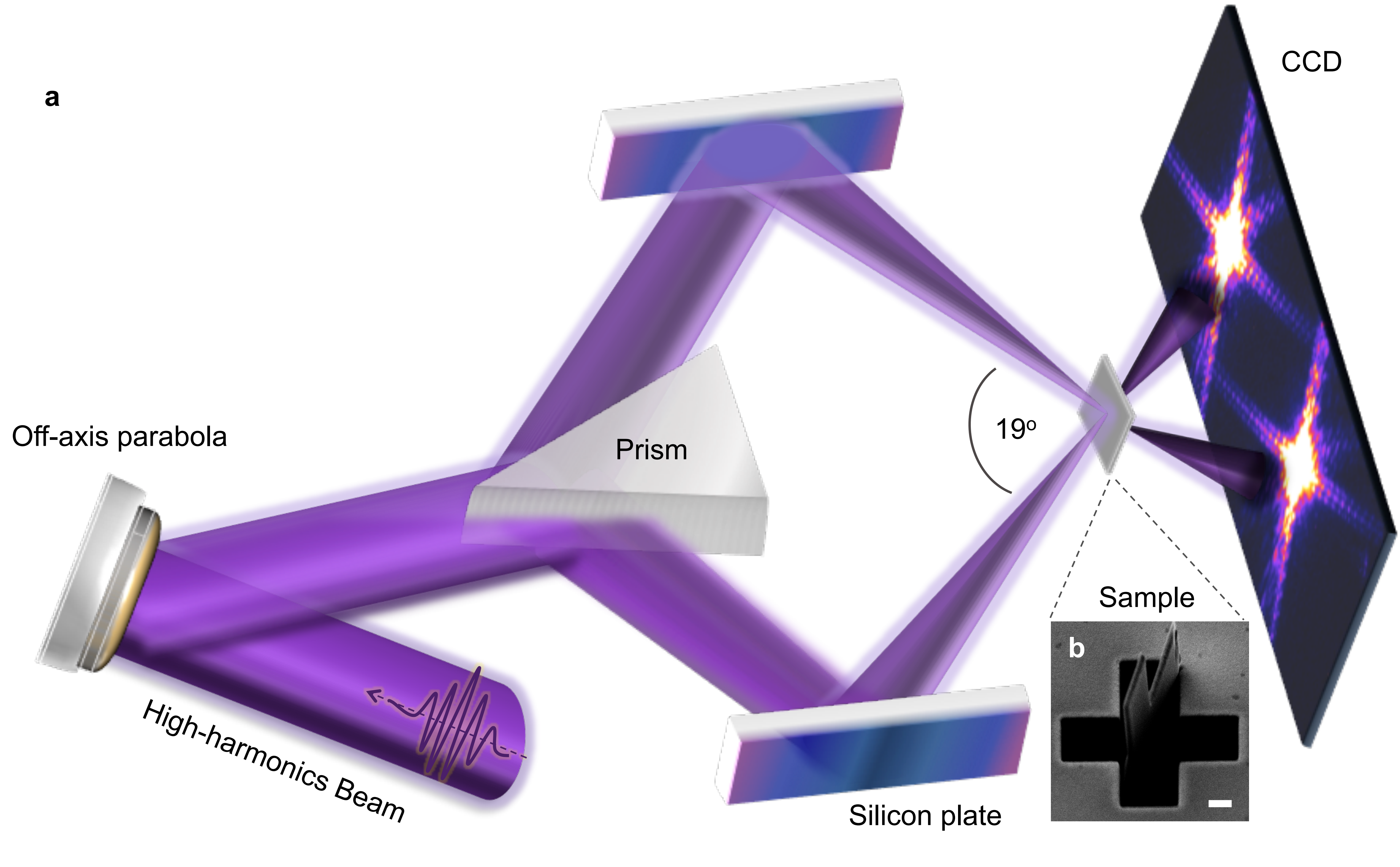
Fig. 1 | Experimental setup and sample for single acquisition 3D stereo imaging. a, A multilayer coated off-axis parabola selects harmonic 33 from the laser beam (λ = 24 nm) and focuses it into the sample. A grazing-incidence prism inserted after the focusing optics splits the beam in two stereo pairs. Controllable silicon mirrors are used to reflect each sub-beam onto the sample. A single CCD camera is positioned in the far field and used to simultaneously record the two diffraction patterns, which constitute the input images to the stereo views. **b,** SEM (scanning electron microscopy) image of the sample observed at a 60° angle. The scale bar has a length of 1 µm.

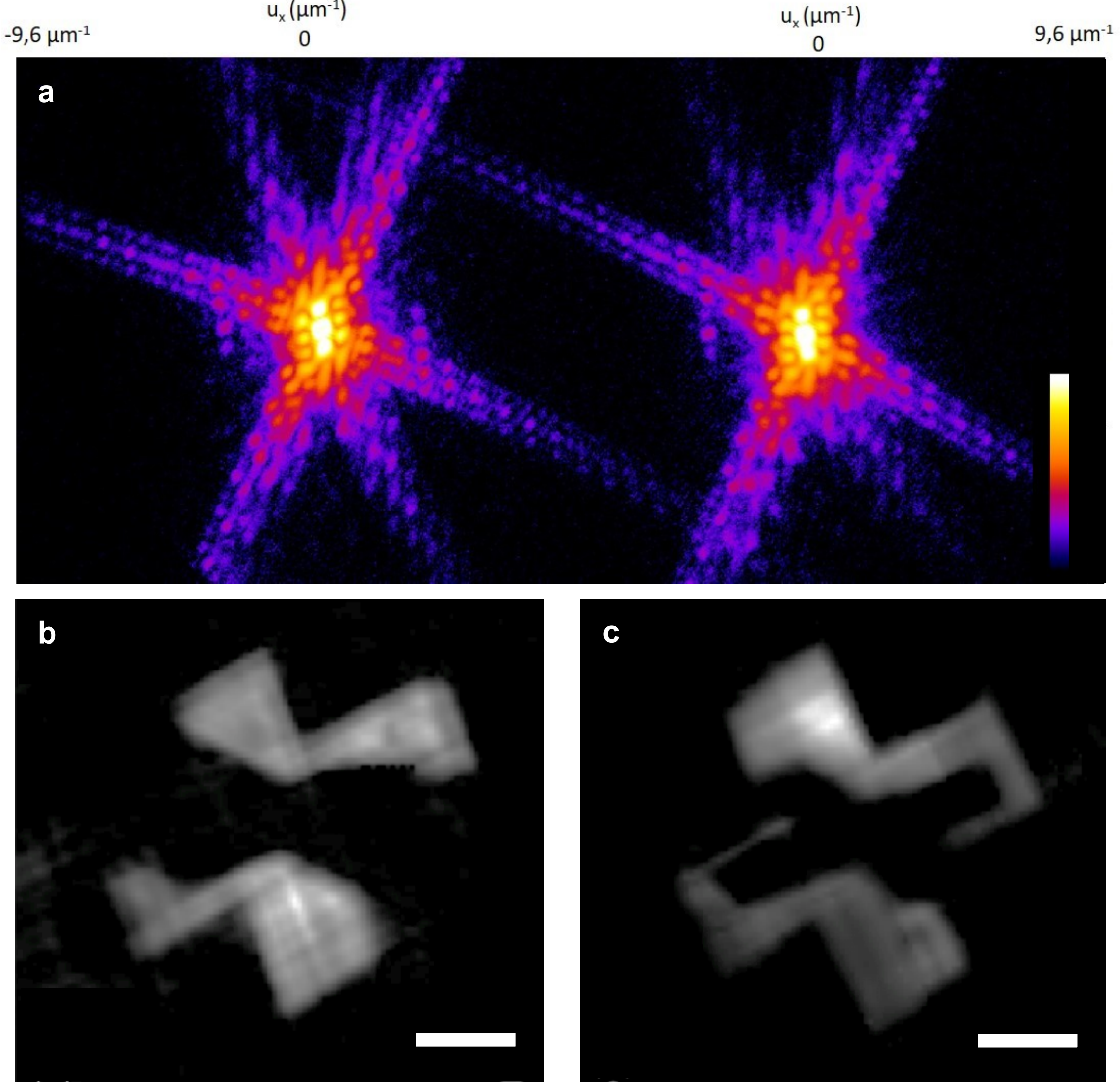
Fig. 2 | Typical dual diffraction pattern and stereo reconstruction of the amplitude sample. a, Diffraction patterns, achieved from the two stereo angles, recorded on a single X-ray CCD. The left (right) diffraction pattern corresponds to the beam coming on the sample from the right (left). The image is shown in logarithmic scale. The accumulation time was 180s. **b-c,** 2D-amplitude reconstructions, corresponding to the left and right views of the sample, respectively. They are obtained from high dynamic range *(HDR)* diffraction patterns and correspond to the coherent average of 45 reconstructions from independent runs of the CDI algorithm. Each view reaches a spatial resolution of 127 nm. The scale-bar length is 2 µm.

Fig. 3 | Amplitude and phase computed stereo reconstructions. a-b, Experimental disparity map and computed depth-map of the "cross" pure amplitude sample, respectively**.** The disparity map is obtained from the correspondence between the left and right views from Fig. 2b-c. **c-d**, Simulated stereo reconstruction of a phase sample using a similar object but adding X-ray transparency, assuming different refraction-index materials for the cross and the membrane. Phase images of each view can be extracted and a disparity map is computed in **c**, from which depth information is retrieved, **d**. The scale bar in **a**, **c** denotes the disparity in pixels, while the colour scale in **b**, **d** represents the depth value in µm.

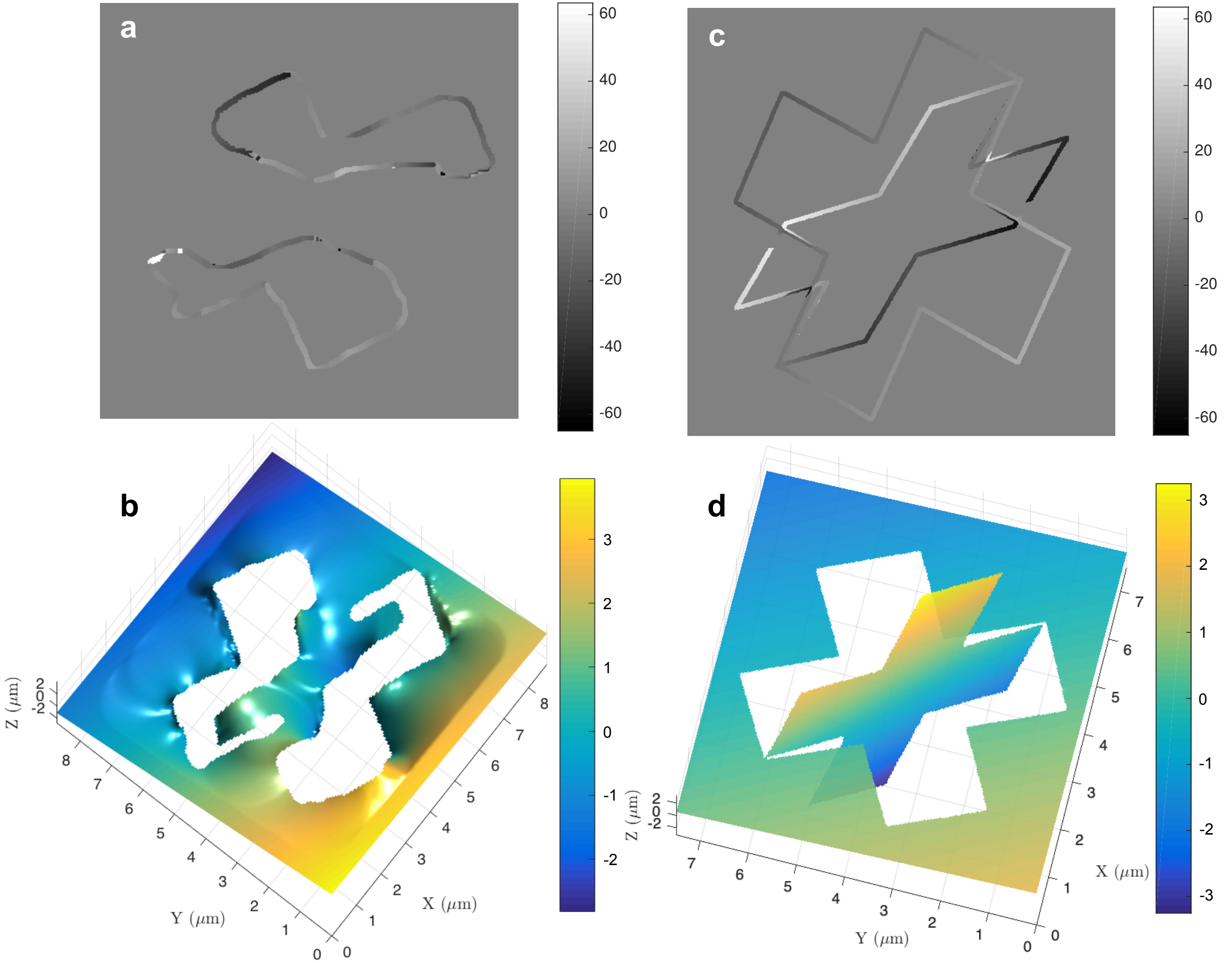
Fig. 4 | 3D stereo imaging of an arrangement of nanoparticles. a-**b,** SEM image and respective iso-surface rendering of the test object composed by 50-nm-diameter gold spheres lining the inside of a pyramid-shaped notch lying a 100-nm-thick silicon nitride membrane. The pyramid base has a width of 2.5 µm and 1.8 µm of height. The scale-bar length is 1 µm. The inset in **a** shows a zoom emphasizing a detail of 105-nm depth, δz. **c-d,**

2D CDI reconstructions, corresponding to 7^{o}-apart views of the sample. Each view reaches a spatial resolution of 42 nm. **e**, Disparity map of the stereo views in **c-d**. The colour bar represents the disparity in pixels. **f,** 3D reconstruction of the nano-pyramid, retrieved directly from the disparity maps, without 3D interpolation for surface rendering. The colour bar represents the depth, in μm. The circle displays a zoom of the detail corresponding to the one shown in **a** as an inset. The image is inverted at it seen from the other side.

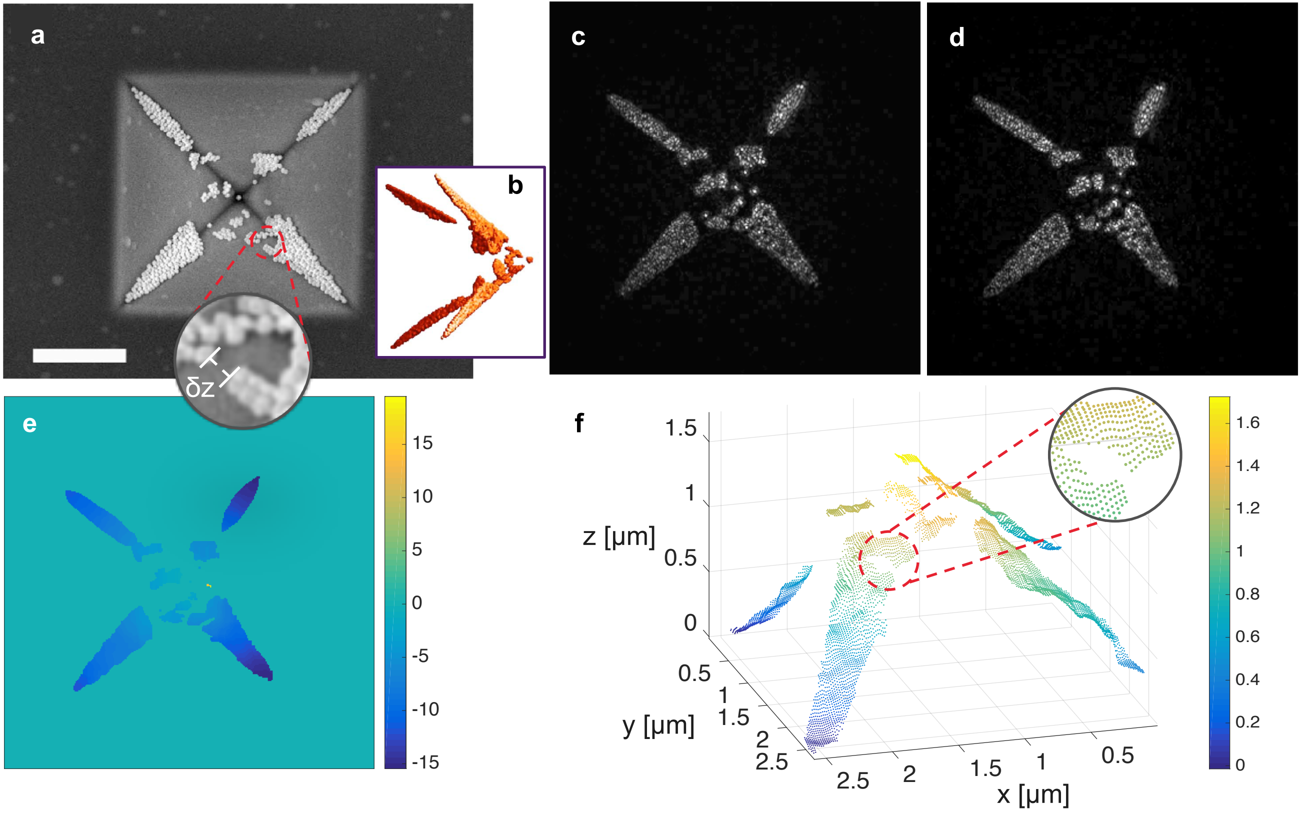

a
CCD
Off-axis parabola
Prism
19°
Sample
b
High-harmonics Beam
Silicon plate



-9,6 µm-1
ux (µm-1)
0
ux (µm-1)
0
9,6 µm-1
a
b
c



a
60
40
20
0
-20
-40
-60
b
3
2
1
0
-1
-2
Z (μm)
2
0
-2
X (μm)
Y (μm)
c
d
X (μm)
Y (μm)
Z (μm)


a
b
c
d
δz
e
15
10
5
0
-5
-10
-15
f
z [µm]
1.5
1
0.5
0
y [µm]
0.5
1
1.5
2
2.5
x [µm]
2.5
2
1.5
1
0.5
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0