Detection and orientation evaluation for three-dimensional objects

José J. Esteve-Taboada *, Javier García

Departament d'Òptica, Facultat de Física, Universitat de València, C/ Doctor Moliner, 50, 46100 Burjassot (València), Spain

Abstract

This paper presents a novel method for detection and orientation evaluation for three-dimensional (3-D) objects based on range images. It is based on the calculation of the phase Fourier transform of the 3-D object range image. The method permits the estimation of the angular position of the object. It works with rotations around an axis perpendicular to the line of sight, although it exhibits a certain nodding tolerance. Simulated results demonstrate the ability of the new introduced method.
© 2003 Elsevier Science B.V. All rights reserved.

1. Introduction

In the last years many efforts have been devoted to the three-dimensional (3-D) object recognition field [1–8]. Many of the derived techniques are based on obtaining a range image of the 3-D object as an intermediate step of the recognition task. A range image of a 3-D object is defined as $x(y, z)$, and the intensity image is defined as $I(y, z)$. Each point (y, z) of the 3-D object has a corresponding intensity I and a range x with respect to a reference plane. Since $I(y, z)$ contains both photometric and geometrical structure information, the range image contains only geometrical information on the 3-D object. Thus, if the lighting changes, $I(y, z)$ will

change but the range image will remain the same whatever the surrounding lighting be. The main advantage of the range image is that all the 3-D information is stored in a two-dimensional (2-D) image containing only geometrical information.

The range image $x(y, z)$ may be obtained by transforming the data given by any of the available techniques for acquiring images of 3-D objects, such as range cameras [9], triangulation (see, for instance [10]), or fringe projection systems [11].

Regarding 3-D object recognition based on range images, in [7] correlation of range images is used to establish the shape departure between a reference object and an acquired one. As the authors demonstrate, smooth and small amplitude defects of the 3-D objects can be detected by enhancing the correlation result using coding and filtering operations.

Other method for obtaining 3-D object recognition based on range images is described in [8], in

---

* Corresponding author. Tel.: +34-96-354-4611; fax: +34-96-354-4715.

E-mail address: jose.j.esteve@uv.es (J.J. Esteve-Taboada).

which a 3-D discrete Fourier transform algorithm is used to correlate several contour images obtained from points at the same depth level in a range image of a 3-D object. As the authors showed, higher computational cost is needed to obtain the 3-D correlation function.

Other techniques use the phase-coding of the range image to perform the 3-D recognition process. In this way, a method for recognizing 3-D objects based on the normals to the surface of the object is introduced in [1]. They showed that it is possible to characterize a 3-D object by using a limited number of normals which they calculated using the phase Fourier transform (PhFT) operation. The 3-D recognition was performed with the help of a feed-forward neural network. The problem, however, was that the process of training the neural network may be complex if one wants to provide a good discrimination result. Another method based on phase-coding of the range image is presented in [2]. They used phase-coding of the range image to obtain translation invariance along the vision axis in the correlation process. They also suggested the sine-coding of the range image in order to obtain better discrimination than that obtained with the phase-coding technique.

In this manuscript we introduce a new technique to perform 3-D object recognition and also an evaluation of the 3-D orientation based on range images. The method is based on the calculation of the phase Fourier transform of the 3-D object range image. As will be shown, it permits the estimation of the angular position of the object. In Section 2 the main mathematics are derived. In Section 3 some simulated results are presented to demonstrate the utility of the proposed method. Finally, in Section 4 the main conclusions will be outlined.

2. Basic theory of the method

It is easy to see that any range image of a 3-D object can be considered as a set of facets which may be described by their normals to the surface. If one calculates the PhFT of the range image, each of these facets will contribute to increase the signal in the Fourier plane in a defined position according to the orientation of the facet. Therefore, a range image from a complex 3-D object will have a characteristic PhFT, the 3-D information of the object being univocally encoded in the PhFT.

For simple 3-D objects (small number of facets) the contribution to the PhFT will be concentrated around small number of locations, each one corresponding to each facet that characterize the object. Thus, a description based on normals may be simpler owing to the low number of data that is needed to describe the object (only the significant maxima in the PhFT). On the other hand, in the cases where the curvature of the object's surface is continuous (smooth surface objects, i.e., with large number of facets), the PhFT is a continuous signal, and then the description of the 3-D shape by a range image $x(y, z)$ or by the PhFT has the same complexity. In any case, the characteristic PhFT will allow for obtaining 3-D object recognition based on correlation in the Fourier domain.

Besides, the method permits the estimation of the angular position of the considered 3-D object. The basic idea is that a change in the orientation of the object is converted in a displacement of the corresponding signals in the Fourier plane. Thus, the two PhFTs corresponding to two rotated versions of the same 3-D object will have a high similarity which could be used for obtaining a recognition of the objects. As we will show later, an extension of this idea to the full range of possible orientations around one axis ($-180°$ to $180°$) permits the use of a full map of PhFTs, which will be called 3-D object orientation map (3-DOOM), that can be used to determine the angular position of the 3-D object performing the correlation between the PhFT of the considered range image and the 3-DOOM. As will be shown in the results section, the described method works with rotations around an axis perpendicular to the line of sight, although it exhibits a certain nodding tolerance. For rotations around an axis non-perpendicular to the line of sight the change obtained in the PhFT is a combination of a displacement and a rotation. In these cases the full analysis would be more complex. Rotation-invariant methods [12,13] should be used.

In order to explain the basic theory of the proposed method, consider a range image obtained from a 3-D object

$$x = x(y, z). \tag{1}$$

This range image can be encoded in a phase function as follows:

$$\exp[iwx] = \exp[iwx(y, z)], \tag{2}$$

where w is a constant factor. The PhFT of this range image is defined as [14]

$$\mathrm{PhFT}(x) = F_{2D}\{\exp[iwx(y, z)]\}, \tag{3}$$

where F_{2D} means the 2-D Fourier transform. As it was said above, a range image can be considered as a set of facets which may be described by their normals to the surface. Now, if we consider a rotation in the 3-D object of a defined angle about an axis perpendicular to the line of sight, all the facets that compose the range image will change their orientation according to the rotation of the 3-D object. To clarify this idea, an example considering a very simple 3-D diamond shape object is shown in Fig. 1. Fig. 1(a) shows a range image of this 3-D object in a particular orientation, and Fig. 1(b) shows the corresponding range image of the same object rotated by an axis perpendicular to the line of sight. If one calculates the PhFT of the range image, each of the facets contributes to the signal in the Fourier plane in a defined position according to the orientation of the facet. A change in the orientation of the object, and thus a change in the orientations of all the facets that compose the

range image, is converted, as is shown next, in a displacement of the corresponding signals obtained in the Fourier plane.

To analyze the changes in the PhFT let us consider a planar facet $F_{\varphi\theta}$ of the 3-D object, defined by the angles φ and θ of the vector normal to the surface (see Fig. 2). The PhFT corresponding to this facet may be written, except for constants, as

$$\mathrm{PhFT}(F_{\varphi\theta}) = F_{2D}[\exp\{iw[y\tan(\alpha_y) + z\tan(\alpha_z)]\}]. \tag{4}$$

Because it is the Fourier transform of a linear phase factor, it will be peaked around the spatial frequencies (u, v) in the Fourier plane given by

$$(u, v) \equiv \left(\frac{\tan(\alpha_y)}{2\pi}, \frac{\tan(\alpha_z)}{2\pi} \right). \tag{5}$$

From this peak location in the Fourier plane it is possible to obtain the angular orientation of the normal to the facet with the conventional spherical angles (φ, θ) (see Fig. 2), using the following coordinate transformations:

$$\alpha_y = \varphi, \tag{6}$$

$$\tan(\alpha_z) = \frac{\sin\theta}{\cos\theta\cos\alpha_y} = \frac{\tan\theta}{\cos\varphi}. \tag{7}$$
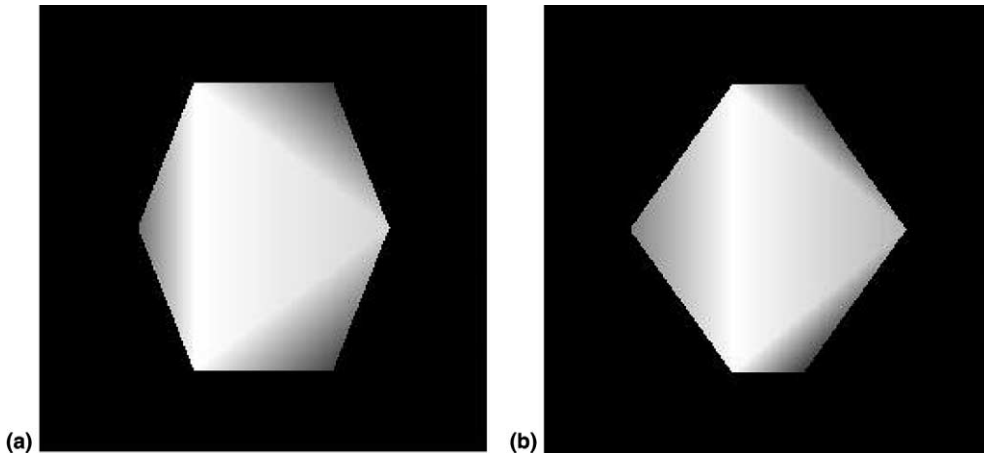
And thus, Eq. (5) may be written

Fig. 1. An example considering a very simple 3-D diamond shape object: (a) range image of this 3-D object in a particular orientation, and (b) range image of the same object rotated by an axis perpendicular to the line of sight.
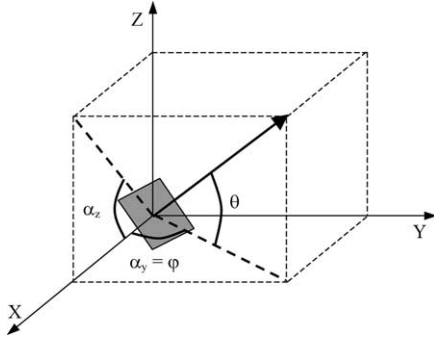
Fig. 2. A particular facet of the range image that is defined by the angles φ and θ of the vector normal to the surface.

$$(u, v) = \left(\frac{\tan(\varphi)}{2\pi}, \frac{\tan(\theta)}{2\pi\cos(\varphi)} \right). \qquad (8)$$

Performing these transformations to the Fourier plane coordinates (u, v) one can obtain the corresponding PhFT in the conventional spherical (φ, θ) coordinates. Therefore, in this new representation of the PhFT, each facet of the range image, which can be characterized by the angles φ and θ of the

vector normal to the surface, will contribute to the PhFT to a location given by the (φ, θ) coordinates.

3. Simulated results

An example of application is depicted in Fig. 3. Fig. 3(a) shows a range image of a 3-D object and Fig. 3(b) shows the modulus of the corresponding PhFT in the (φ, θ) coordinates obtained using Eqs. (3) and (8).

Considering now a rotation in the 3-D object of an angle $\Delta\varphi$ around the Z axis (see Fig. 3(c)), the peak corresponding to one facet will be just displaced, in the (φ, θ) representation of the PhFT, to the new angular orientation $(\varphi + \Delta\varphi, \theta)$. Therefore, a rotation in the 3-D object made of facets is, in principle, directly converted in a displacement of the PhFT in the (φ, θ) domain. An example is presented in Fig. 3(d), in which the modulus of the PhFT in the (φ, θ) coordinates that corresponds to the range image in Fig. 3(c) is shown. This PhFT is just a shifted version of the
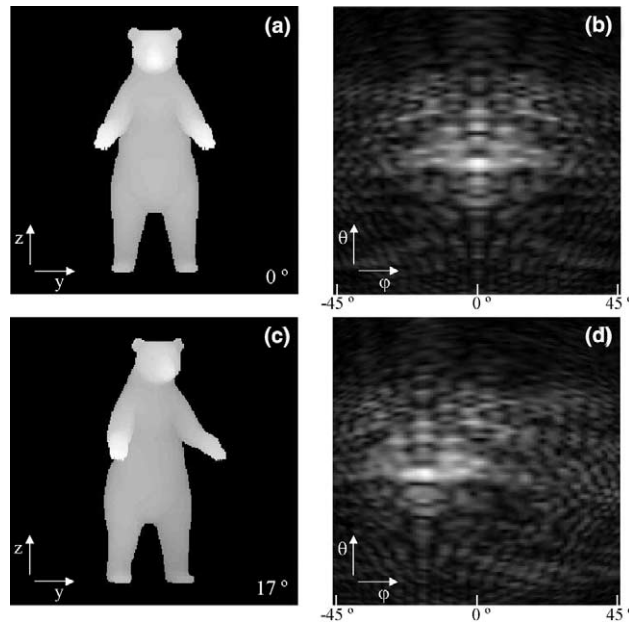

Fig. 3. (a) Range image of a 3-D object rotated 0° around the Z axis. (b) Modulus of the PhFT obtained for the range image shown in (a). (c) Range image of the 3-D object rotated 17° around the Z axis. (d) Modulus of the PhFT obtained for the range image shown in (c).

one obtained for the orientation of the 3-D object shown in Fig. 3(a).

When varying the orientation of the 3-D object, small changes may be appreciated in the obtained PhFT. These changes are due, basically, to the following three reasons: (1) in the rotation process new zones of the 3-D object appear in the range image and other zones disappear from this image, (2) there exist relative changes in the phases of the facets depending on the distance from the particular facet to the rotation center, and (3) the size of the cross-section of each facet from the viewpoint changes along the rotation process, which makes the contribution of each facet have different energies in the different orientations of the 3-D object. However, as has been shown in the example of application depicted in Fig. 3, the effects of these factors are minor compared with the effect of the lateral displacement of the PhFT that occurs in the (φ, θ) domain. A quantitative assessment of the distortion obtained in the PhFTs when varying the orientation of the 3-D object can be obtained by performing the correlation between the PhFT corresponding to the non-rotated 3-D object and the PhFTs of the different rotated 3-D objects within a certain interval around the Z axis. The obtained result is plotted in Fig. 4. As can be seen the correlation peak value has a smooth decay with the rotation angle (rotation up to $\pm 20°$ only influence in a 10% of the maximum correlation peak value).
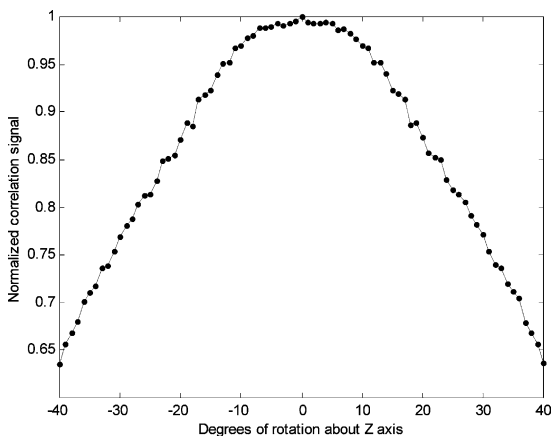

Fig. 4. Normalized correlation signal for the PhFTs corresponding to different rotations around the Z axis of the 3-D object shown in Fig. 3(a).

The high similarity between the PhFTs corresponding to the two rotated versions of the 3-D object previously shown indicates the possibility of making the 3-D recognition of these objects by comparison (through correlation) of the coordinate transformed PhFTs. Thus, in order to extend the detection capabilities for the full range of rotations, a full map of PhFTs for all possible orientations is needed. Considering rotations around the Z axis, this map (the above called 3-DOOM) is obtained by pasting on a large image, covering the full φ range (−180° to 180°), the coordinate transformed PhFTs. The rotation is performed every 1°, and thus the PhFTs are highly overlapping. This permits an averaging in the pasting process that diminishes the effect of the above mentioned small changes in the PhFTs. Bearing in mind that the distortion of the PhFT is greater for the points far from the center of the image, it could be possible to obtain a 3-DOOM taking into account different weights for the different points of the images. However, the differences between the distinct methods are marginal.

Therefore, the 3-DOOM is just the unfolded complete map of the modulus of the PhFTs corresponding to the 3-D object in all the possible orientations around one axis. As an example, the resultant 3-DOOM for the previously used 3-D object (see Fig. 3), when considering rotations about the Z axis, is the one shown in Fig. 5(a).

This 3-DOOM can be used to achieve 3-D object recognition and also to determine the angular position of the particular 3-D object just by performing the correlation between the 3-DOOM and the corresponding modulus of the PhFT of the considered 3-D object. As for each point (y, z) the height of the 3-D object is codified in a phase function before the Fourier transforming step, the recognition process will be based in the comparison of the relieves of the 3-D objects. For instance, Fig. 5(b) shows the 3-D object in a particular orientation (as indicated, rotated 93.5°) and Fig. 5(c) shows the modulus of its PhFT in the (φ, θ) coordinates. Performing the correlation between the previously obtained 3-DOOM (Fig. 5(a)) and this modulus of the PhFT (Fig. 5(c)) one can obtain the correlation plane shown in Fig. 5(d). A
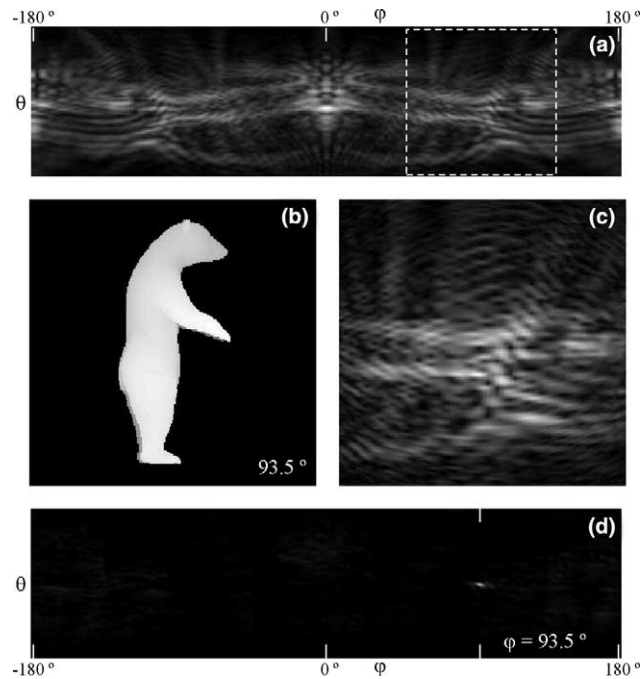
Fig. 5. (a) 3-DOOM of the 3-D object shown in Fig. 3. (b) Range image of the 3-D object rotated 93.5° around the Z axis. (c) Modulus of the PhFT obtained for the range image shown in (b). (d) Correlation plane obtained for the modulus of the PhFT shown in (c). As indicated, the orientation of the 3-D object can be obtained from the position of the correlation peak.

correlation peak is obtained because the 3-D object coincides with the target considered to obtain the 3-DOOM. Besides, as indicated, it is easy to get the orientation of the 3-D object from the position of the correlation peak in this image using the coordinate transformation shown in Eq. (8).

In the recognition process, a certain nodding tolerance is also obtained. This means that the recognition of the 3-D object is also achieved when it presents small rotations around the other axis perpendicular to the line of sight, in our case, the Y axis. To demonstrate this idea we have considered small rotations around the Y axis in the 3-D object shown in Fig. 5(b). The obtained correlation results, normalized to the maximum value, are plotted in Fig. 6. As can be seen, in the case of the considered reference 3-D object, a nodding tolerance of about ±10° can be allowed in order to keep the correlation peak over 50% of the maximum value.

The limited tolerance to the nodding of the 3-D object deserves some explanation. The proposed

method requires the conversion of the polar coordinates to Cartesian coordinates in order to display them in a 2-D image. We have used an equidistant cylindrical projection, in which the
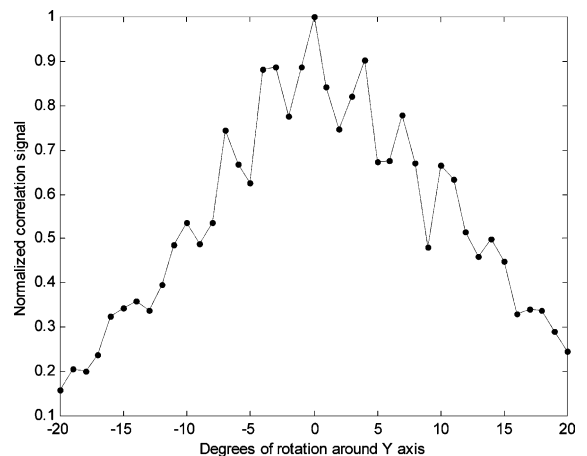

Fig. 6. Normalized correlation signal for different rotations around the Y axis for the 3-D object shown in Fig. 5(b).

(φ, θ) angles are linearly scaled. This introduces a distortion, mainly in the polar areas ($\theta = 0$, $\theta = \pi$). Every rotation around any other axis different from the Z one (which defines the polar coordinate system) will produce a distortion in the PhFT avoiding a perfect matching with the previously computed 3-DOOM. Possible solutions to this problem would be the use of additional 3-DOOMs with other coordinate choices and/or the use of different cartographic projections.

As the PhFTs contains all the information of the particular 3-D object, it is clear that the discrimination ability of the proposed method will be comparable to that of the commonly used methods for 3-D object recognition (see, for example, [1–8]). In order to show the recognition and discrimination abilities of the proposed method, we consider the correlation between the 3-DOOM of the reference 3-D object (see Fig. 7(a)) and the modulus of the coordinate trans-

formed PhFT obtained from a different 3-D object in a particular orientation around the Z axis (see Figs. 7(b) and (c)). The resultant correlation image is the one shown in Fig. 7(d). As can be seen, no correlation signal is obtained, which means that the considered 3-D object does not have correspondence with the 3-DOOM of the reference object.

Finally, it is worthy to note that the computation of the 3-DOOM has to be performed only one time for each 3-D object to be recognized, and that using this 3-DOOM it is possible to obtain the 3-D recognition process for all the possible orientations of the object around a certain axis. Besides, as is said above, the 3-DOOM is a unique function that only depends on the considered 3-D object and that characterizes it following a one-to-one relation. This can be seen in Fig. 8, in which some examples of 3-DOOMs are depicted for different 3-D objects. Figs. 8(a) and (c) are the range images
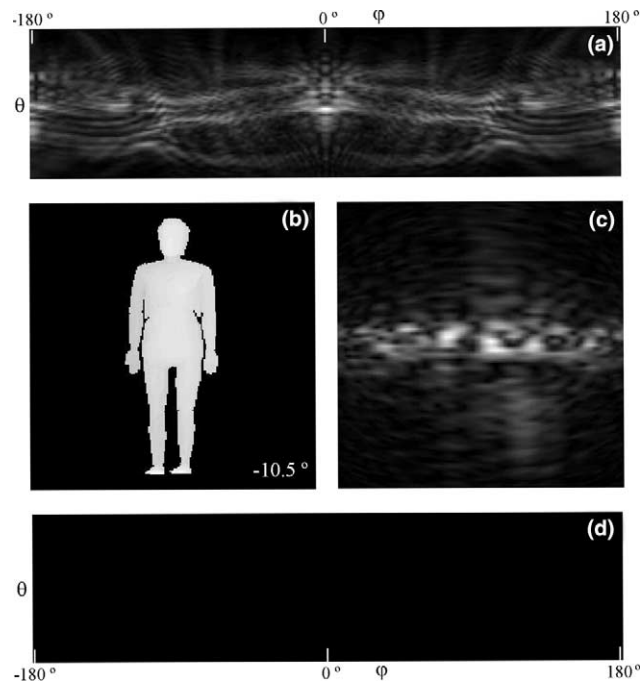

Fig. 7. (a) 3-DOOM of the 3-D object shown in Fig. 3. (b) Range image of a different 3-D object rotated −10.5° around the Z axis. (c) Modulus of the PhFT obtained for the range image shown in (b). (d) Correlation plane obtained for the modulus of the PhFT shown in (c). As can be seen, no correlation signal is obtained, which means that the considered 3-D object does not have correspondence with the 3-DOOM of the reference object.
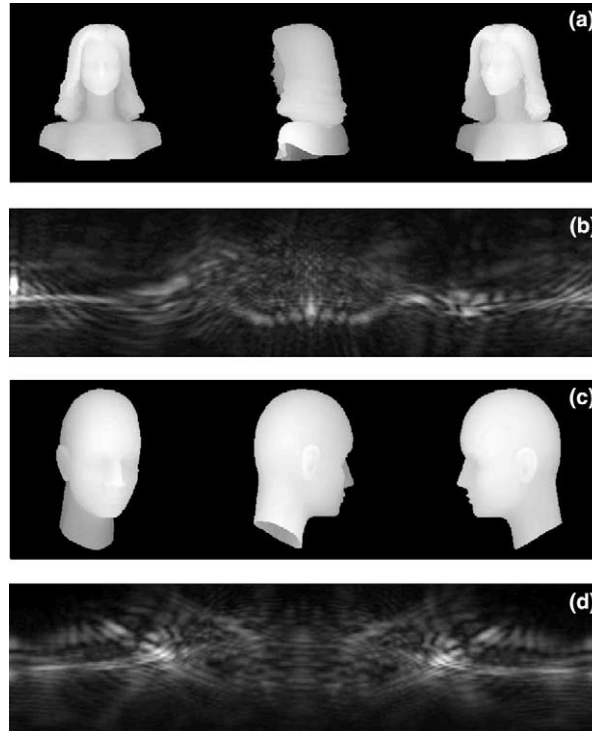
Fig. 8. (a, c) Range images corresponding to different views of these 3-D objects. (b, d) Resultant 3-DOOMs for these 3-D objects.

corresponding to different views of the 3-D objects, and Figs. 8(b) and (d) are the resultant 3-DOOMs for these 3-D objects.

4. Conclusions

In conclusion, we have presented a new method for 3-D object recognition. It is based on obtaining the PhFT of a range image of the faceted 3-D object, which is a characteristic function that allows for obtaining the 3-D recognition process using correlation in the Fourier domain. Besides, the method permits the estimation of the angular position of the 3-D object, because rotation in the object is directly converted in a displacement of the PhFT after a coordinate transformation. This permits to obtain an unfolded complete map of the coordinate transformed PhFTs of the 3-D object, the 3-DOOM (a characteristic function of the considered 3-D object), which can be used to determine the angular position of the object by means of correlation. The method works with rotations around an axis perpendicular to the line of sight, although a certain nodding tolerance is also obtained. Results verify the derived theory and show the utility of the introduced method.

References

[1] E. Paquet, M. Rioux, H.H. Arsenault, Opt. Eng. 34 (1995) 1178.

[2] E. Paquet, P. García-Martínez, J. García, J. Opt. 29 (1998) 35.

[3] J. Rosen, J. Opt. Soc. Am. A 15 (1998) 430.

[4] T. Poon, T. Kim, Appl. Opt. 38 (1999) 370.

[5] B. Javidi, E. Tajahuerce, Opt. Lett. 25 (2000) 610.

[6] J.J. Esteve-Taboada, D. Mas, J. García, Appl. Opt. 38 (1999) 4760.

[7] P. Parrein, J. Taboury, P. Chavel, Opt. Commun. 195 (2001) 393.

[8] J. Guerrero-Bermúdez, J. Meneses, O. Gualdrón, Opt. Eng. 39 (2000) 2828.

[9] M. Rioux, Appl. Opt. 23 (1984) 3837.

[10] S. Vallmitjana, I. Juvells, A. Carnicer, M. Montes, E. Martín, I. Labastida, F. Rodríguez, R. Tudela, in: Ph. Réfrégier, B. Javidi, C. Ferreira, S. Vallmitjana (Eds.), Optoelectronic Information Processing: Optics for Information Systems, SPIE Critical Review CR 81 (2001) 148.

[11] M. Takeda, K. Mutoh, Appl. Opt. 22 (1983) 3977.

[12] Y.N. Hsu, H.H. Arsenault, G. Abril, Appl. Opt. 21 (1982) 4012.

[13] Y.N. Hsu, H.H. Arsenault, Appl. Opt. 21 (1982) 4016.

[14] E. Paquet, H.H. Arsenault, M. Rioux, Opt. Eng. 32 (1993) 2173–2180.