

# Phase Fourier vector model for scale invariant three-dimensional image detection

José J. Vallés, Pascuala García-Martínez, Javier García, and Carlos Ferreira

*Departamento de Óptica. Universitat de València. C/ Dr. Moliner 50. 46100Burjassot Spain*  
[pascuala.garcia@uv.es](mailto:pascuala.garcia@uv.es)

**Abstract:** A scale invariant 3D object detection method based on phase Fourier transform (PhFT) is addressed. Three-dimensionality is expressed in terms of range images. The PhFT of a range image gives information about the orientations of the surfaces in the 3D object. When the object is scaled, the PhFT becomes a distribution multiplied by a constant factor which is related to the scale factor. Then 3D scale invariant detection can be solved as illumination invariant detection process. Several correlation operations based on vector space representation are applied. Results show the tolerance of detection method to scale besides discrimination against false objects.

©2007 Optical Society of America

**OCIS codes:** (070.2590) Fourier transforms; (070.5010) Pattern recognition and feature extraction; (100.6890) Three-dimensional image processing

---

## References and links

1. B. Javidi, ed., *Image Recognition and Classification: Algorithms, Systems, and Applications*, (Marcel Dekker, New York, 2002).
2. J. Rosen, "Three-dimensional electro-optical correlation," *J. Opt. Soc. Am. A* **15**, 430–436 (1998).
3. J. Rosen, "Three-dimensional joint transform correlator," *Appl. Opt.* **37**, 7538–7544 (1998).
4. T. Poon and T. Kim, "Optical image recognition of three-dimensional objects," *Appl. Opt.* **38**, 370–381 (1999).
5. B. Javidi and E. Tajahuerce, "Three-dimensional object recognition by use of digital holography," *Opt. Lett.* **25**, 610–612 (2000).
6. B. Javidi, I. Moon, S. Yeom, and E. Carapezza, "Three-dimensional imaging and recognition of microorganism using single-exposure on-line (SEOL) digital holography," *Opt. Express* **13**, 4492–4506 (2005).
7. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," *Appl. Opt.* **22**, 3977–3882 (1983).
8. J. J. Esteve-Taboada, D. Mas, and J. García, "Three-dimensional object recognition by Fourier transform profilometry," *Appl. Opt.* **38**, 4760–4765 (1999).
9. J. J. Esteve-Taboada, J. García and C. Ferreira, "Rotation invariant optical recognition of three-dimensional objects," *Appl. Opt.* **39**, 5998–5352 (2000).
10. J. García, J. J. Vallés and C. Ferreira, "Detection of three-dimensional objects under arbitrary rotations based on range images," *Opt. Express* **11**, 3352–3358 (2003).
11. J. J. Esteve-Taboada, N. Palmer, J.-Ch. Giannesini, J. García and C. Ferreira, "Recognition of polychromatic three-dimensional objects," *Appl. Opt.* **43**, 433–441 (2004).
12. J. J. Esteve-Taboada, J. García and C. Ferreira, "Optical recognition of three-dimensional objects with scale invariance using classical convergent correlator," *Opt. Eng.* **41**, 1324–1330 (2002).
13. M. Rioux, "Laser range finder based on synchronized scanners," *Appl. Opt.* **23**, 3837–3844 (1984).
14. E. Paquet, M. Rioux and H. H. Arsenault, "Invariant pattern recognition for range images using the phase Fourier transform and a neural network," *Opt. Eng.* **34**, 1178–1183 (1995).
15. E. Paquet, P. García-Martínez, and J. García, "Tridimensional invariant correlation based on phase-coded and sine-coded range images," *J. Opt.* **29**, 35–39 (1998).
16. S. Chang, M. Rioux, and C. P. Grover, "Range face recognition based on the phase Fourier transform," *Opt. Commun.* **222**, 143–153 (2003).
17. Y. Li and J. Rosen, "Scale invariant recognition of three-dimensional objects by use of a quasi-correlator," *Appl. Opt.* **42**, 811–819 (2003).
18. D. Lefebvre, H. H. Arsenault, P. García-Martínez, and C. Ferreira, "Recognition of unsegmented targets invariant under transformations of intensity," *Appl. Opt.* **41**, 6135–6142 (2002).
19. H. H. Arsenault and P. García-Martínez, "Intensity-invariant nonlinear filtering for detection in camouflage" *Appl. Opt.* **44**, 5483–5490 (2005).

20. H. H. Arsenault and D. Lefebvre, "Homomorphic cameo filter for pattern recognition that is invariant with change of illumination," *Opt. Lett.* **25**, 1567-1569 (2000).
  21. D. Lefebvre, H. H. Arsenault, and S. Roy, "Nonlinear filter for pattern recognition invariant to illumination and to out-of-plane rotations," *Appl. Opt.* **42**, 4658-4662 (2003).
  22. S. Roy, D. Lefebvre, and H. H. Arsenault, "Recognition invariant under unknown affine transformations of intensity," *Opt. Commun.* **238**, 69-77 (2004).
  23. J. J. Vallés, J. García, P. García-Martínez, and H. H. Arsenault, "Three-dimensional object detection under arbitrary lighting conditions," *Appl. Opt.* **45**, 5237-5247 (2006).
  24. F. M. Dickey and L. A. Romero, "Normalized correlation for pattern recognition," *Opt. Lett.* **16**, 1186-1188 (1991).
- 

## 1. Introduction

Recent interest in three-dimensional (3D) optical information systems has increased because of its vast potential in applications such as object recognition, image encryption as well as 3D display [1]. Regarding 3D object optical recognition, joint transform correlator (JTC) architecture in combination to electronic processing can be used [2, 3]. However it is a complex setup and it is computationally intensive, although it provides localization of 3D targets in 3D space. Other techniques use digital holography as a recording method to perform correlation between planar holograms of the 3D functions [4-6]. In addition, techniques based on Fourier transform profilometry (FTP) are used for 3D object detection [7,8]. The FTP technique deals with the projection of gratings onto the 3D object besides a registration of a 2D image carrying the 3D information. It has been applied to many pattern recognition tasks, as rotation invariance [9, 10], color pattern recognition [11] and scale invariance [12].

Other recognition procedures are based on the application of optical or digital processing to range imagery. A range image contains the depth information of an object from a given view line, that defines the  $z$  axis [13]. One of the main advantages of the range image is that all 3D information is stored in a 2D image containing only geometrical information. The encoding of the depth information has been used in the literature to extend the possibilities of range images recognition using the phase Fourier transform (*PhFT*) [14,15]. They showed that it is possible to characterize a 3D object by using a limited number of normals which they calculated using the *PhFT* operation. The idea is to detect the planar surface of a 3D range image when the Fourier transform of a phase-coded image (the phase being proportional to the elevation) is calculated using feed-forward neural networks [14]. Another application of this technique is face recognition applications, for which the range image plays an important role since it contains richer and more accurate shape information than 2D intensity images [16]

However, a main limitation of most 3D matching methods is the limited tolerance to scale changes in the object. Several methods have been shown in the literature to show scale tolerance detection applied to 3D images [12, 14, 17]. In Ref. [12] the authors combine profilometry techniques with optical matched filtering based on Mellin radial harmonics to obtain 3D scale invariant pattern recognition. Moreover, other optical methods based on performing a correlation between a set of images of a tested 3D scene with a logarithmic radial harmonic filter is applied for scale invariance [17]. However the method suffers from complexity and a certain limitation of the space bandwidth product. To avoid such inconveniences, the use of *PhFT* of range images is studied for shift, rotation and scale invariance using feed-forward neural networks [14]. However, the process of training the neural network may be complex if one wants to provide a good discrimination result.

In this paper we have addressed scale invariant 3D object detection using the *PhFT* applied to range imagery in combination with nonlinear statistical operations that we have defined recently for intensity invariant 2D pattern recognition [18,19]. We show that scaling a 3D range object implies the multiplication of the *PhFT* distribution by a constant factor. So, a change of scale for a range image is transformed in a change in illumination (multiplication by a constant) of the *PhFT* amplitude. In common 2D correlation, if the illumination model consists in multiplying a target by an unknown constant factor, the correlation peak will change by the same amount. In such cases dark targets can be missed. Arsenault and

Lefebvre [20] used a homomorphic transformation to change a multiplicative-intensity problem into an additive-intensity problem that can be addressed with the synthetic discriminant filter mentioned above. Moreover, Lefebvre et al [18] defined a nonlinear filtering method known as the LACIF (Locally Adaptive Contrast Invariant Filter), which is invariant under any linear intensity transformation. This LACIF operation uses three correlations involving local statistics and nonlinearities. It was applied directly to scenes containing unsegmented targets. One of the advantages of the LACIF method is that no a priori information about the constant values involved in the linear illumination model is assumed. The LACIF method can also be combined with synthetic discrimination filters to achieve both illumination invariance and out-of-plane rotation invariance [21]. In Ref. [22] the authors generalized the LACIF filtering for situations where a linear intensity gradient across an object is presented. It is interesting to consider the LACIF technique in the context of a vector space interpretation. In addition, in Ref. [19], the authors applied LACIF algorithm to unsegmented natural camouflage scenes while maintaining intensity invariance. Recently, we have applied another filtering connected with LACIF to 3D shading to achieve illumination invariance [23]. In section 2 we review the basis of *PhFT* and how affects a change of scale in range imagery. LACIF filtering is described in Section 3. The application of the method to 3D scaled targets is in Section 4.

## 2. Scale changes for range image phase Fourier transforms

A range image,  $z(x, y)$ , contains the depth information of an object from a given view line, that defines the  $z$  axis. One of the main advantages of a range image is that three-dimensional information is stored in a 2D image containing only geometrical information. In fact, the range image is considered as a set of facets which may be described by their normals to the surface. The encoding of the depth information has been used in the literature to extend the possibilities of range images for pattern recognition [10, 14, 15]. Those techniques are based on phase coding. We encode the range as a phase distribution as follows:

$$P(x, y) = \exp[imz(x, y)] \quad (1)$$

where  $m$  is a constant that permits the adjustment of the phase slope of the object. From now, without losing generality, we will assume  $m=1$ .

A way to deal with range images keeping translation invariance is to use their Fourier transform. The Fourier transform of the phase encoded range image (*PhFT*) is

$$PhFT_z(u, v) = F_{2D} \{ \exp[iz(x, y)] \} \quad (2)$$

where  $F_{2D}$  stands for two dimensional Fourier transform.

One of the most important properties of the *PhFT* is that it contains information of all the orientations of the surfaces that defines a given 3D object [14, 15]. For instance, if the object is defined by a planar surface, after a phase encoding, it will become a planar phase distribution. Thus its Fourier transform will be peaked around a well defined location. So *PhFT* maps a facet into a peak. The position and distribution of the peak represent the orientation and the boundary of the facet, respectively. The intensity of the *PhFT* exhibits a crucial property: it is invariant to arbitrary translations of the object. This property is obvious for translation in the  $(x, y)$  plane, as they will produce just a linear phase factor in the *PhFT*. On the other hand, from the definition of the phase encoded range image [see Eq. (1)], a shift along the view line ( $z$  axis) will influence just as a constant phase factor. In both cases the *PhFT* is just altered by a phase that is irrelevant in intensity [15]. Furthermore, changing the orientation of the object (i.e. rotating all the facets that compose the range image) implies a translation of angular *PhFTs*. This connection between rotation in spatial domain and translation in angular *PhFT* domain makes it advantageous for 3D object correlation based recognition methods [9, 10].

As an example we consider Fig. 1(a), that contains a simple range image (pyramid) with four facets and the *PhFT* amplitude of Fig. 1(a) is shown in Fig. 1(b). The four facets (four normals) correspond to four location peaks in the Fourier plane. On the other hand, if the curvature of the object's surface is a continuous function (smooth object surface, i. e. with large number of facets), the *PhFT* will be continuous.

Now we show the effect of a scale change in *PhFT* domain. Let assume that the scaled object

is given by  $z'(x, y) = z\left(\frac{x}{k}, \frac{y}{k}\right)$  with  $k$  being the scale factor. Although the scale is changed,

the shape of the object will remain the same as well as the orientation of the normals, except for a constant value which is related to the scale factor. In Appendix A, we show the connection between *PhFTs* of scaled objects. In fact, from Appendix A, the *PhFT* of the scaled object can be approximated as

$$PhFT_z'(u, v) \cong k^2 PhFT_z(u, v) \quad (3)$$

So a scale change in the range image will only lead to a multiplication of the amplitude of the non-scaled *PhFT* domain by a constant factor, with no change in the pattern distribution of the *PhFTs*. Only a global change in the intensity is observed.

Fig. 1(c) shows a scaled version of Fig. 1(a), and Fig. 1(d) shows the *PhFT* of Fig. 1(c). Both *PhFTs* [Fig. 1(b) and Fig. 1(c)] are the same except for a global multiplicative constant factor.

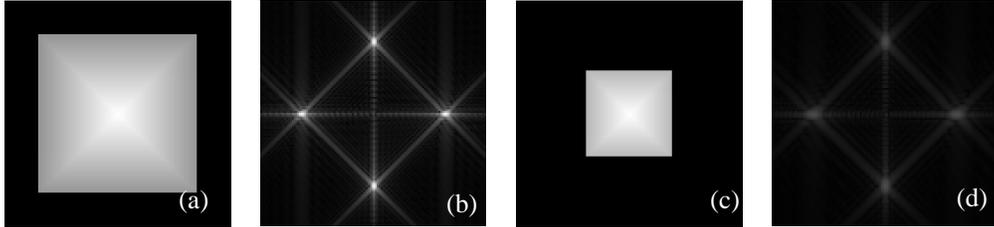


Fig. 1. (a). Range image of a diamond shaped object. (b). *PhFT* amplitude of Fig. 1(a). (c) Scaled version of Fig. 1(a). (d) *PhFT* amplitude of Fig. 1(c).

Moreover, if the surface of the 3D object is a bounded curve, the change of scale causes some new problems. Because of digitalization processes, a larger curved surface contains more small facets than those corresponding to a smaller curved surface (See Fig. 2).

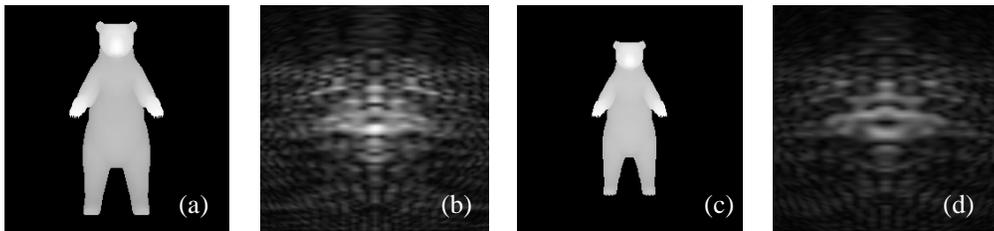


Fig. 2. (a). Range image. (b). *PhFT* amplitude of Fig. 2(a). (c) Scaled range image. (d) *PhFT* amplitude of Fig. 2(c)

Note that the relation between the *PhFTs* shown in Figs. 2(b) and 2(d), respectively, is not a global constant factor. In section 4 we will deal with this factor in more details. But, except for those digitalization errors, changes in scale involve changes in intensity or amplitude. There are several methods for intensity invariant pattern recognition based on correlations [18-22, 24]. In this paper we will apply the LACIF method. Following section briefly review the essentials.

### 3. Intensity invariant correlation method

Images may be considered as vectors in a Hilbert space. Then any vector can be expressed in terms of a given basis. Correlation is a measure of the similarity between images, because one

way to understand the correlation operation is to consider it as an inner product between two functions, the object and the reference. From the point of view of vector spaces, intensity invariant pattern recognition consists of recognizing vectors independently of their length, which can be viewed as an angle measurement between vectors in vector spaces. The vector space defined here consists of *PhFT* amplitude distributions.

Because a change of intensity implies a change in the length of the vector, two equal targets with different illuminations will be parallel, so the normalized correlation and the cosine will be equal to one. For 2-D images, Lefebvre et al [18] defined LACIF operation, which is invariant under any linear intensity transformation. This LACIF uses three correlations involving local statistics and nonlinearities. It was applied directly to scenes containing unsegmented targets. One of the advantages of the LACIF method is that no a priori information about the constant values involved in the linear illumination model is assumed. We will consider that the target is the *PhFT* amplitude in the Fourier domain. So, a linear transformation of intensity over a target can be expressed as

$$PhFT_z(u, v) = \alpha PhFT_z(u, v) + \beta \diamond(u, v) \quad (4)$$

where  $\diamond(u, v)$  is the binary support which is equal to unity over the support of the target  $PhFT_z(u, v)$ , and equals to zero everywhere else, and  $\alpha, \beta$ , are unknown constants. An orthogonal basis for the subspace is selected. We define  $PhFT_z^o(u, v)$  as

$$PhFT_z^o(u, v) = PhFT_z(u, v) - \mu_f \diamond(u, v) \quad (5)$$

where  $\mu_f$  is the mean of  $PhFT_z(u, v)$ . Note that  $PhFT_z^o(u, v)$  is a zero-mean target in the region of support. Then, the target can be defined as a linear combination of two orthogonal images (a silhouette and a zero-mean target) as

$$PhFT_z(u, v) = \alpha' PhFT_z^o(u, v) + \beta' \diamond(u, v) \quad (6)$$

where  $\alpha'$  and  $\beta'$  are constants. The basis defined by  $\{PhFT_z^o(u, v), \diamond(u, v)\}$  is not orthonormal, but we can normalize it to unit length as  $\{\phi_1(u, v) = \diamond(u, v) / \|\diamond(u, v)\|; \phi_2(u, v) = PhFT_z^o(u, v) / \|PhFT_z^o(u, v)\|\}$ , where

$\|PhFT_z^o(u, v)\| = \sqrt{PhFT_z^o(u, v) * PhFT_z^o(u, v)}$ , and  $*$  denotes the correlation, which is the inner product. Taking into account this orthonormal basis,  $\{\phi_1(u, v), \phi_2(u, v)\}$ , the target can now be defined as

$$PhFT_z(u, v) = \alpha'' \phi_1(u, v) + \beta'' \phi_2(u, v) \quad (7)$$

LACIF filtering operation at the output is defined as Refs. [18,19]

$$C_{LACIF}(u, v) = \frac{(PhFT_s(u, v) * \phi_2(u, v))^2}{\sqrt{N} (PhFT_s(u, v)^2 * \phi_1(u, v) - (PhFT_s(u, v) * \phi_1(u, v))^2)} \quad (8)$$

where  $N$  is the number of pixels inside the region of support. Then for a given range target  $s(x, y)$ , if  $PhFT_s(u, v)$  is a linear combination of the orthonormal basis, then the correlation peak will be equal to one, and it will be smaller than one if it is not. We have applied LACIF filtering for scale invariant 3D object detection codified in terms of *PhFTs*. Note that for our 3D detection process only constant  $\alpha$  is considered (see Eq. (3), where  $\alpha = k^2$ ), whereas the constant  $\beta$  equals zero. It means that there is no global constant added to the *PhFT*, but multiplied. Moreover, the definition of Eq. (8) in terms of correlations makes the approach feasible to be implemented optoelectronically. The correlations can be performed by

conventional optical correlators like Vander Lugt correlator or joint transform correlator architectures and the local calculations can be obtained using computer interface.

#### 4. Results of detection

We have carried out experiments to show the performance of the detection method. All of those experiments are based on the calculation of Eq. (8) when the reference target is scaled between 0.6 and 1.2 factors. We have used a square window around the *PhFT* amplitude distribution as the region of support for all the experiments. We have chosen as reference target the average between different scaled reference targets in order to minimize the possible sampling errors due to scaling digitalization process. Results are shown in Fig. 3. As we see from Fig. 3(a), the LACIF is almost constant for all scale factors (60%-120%) and its value oscillates between 0.7 and almost 0.9 correlation peak value. A multimedia file shows the appearance of the range images and the result of the detection.

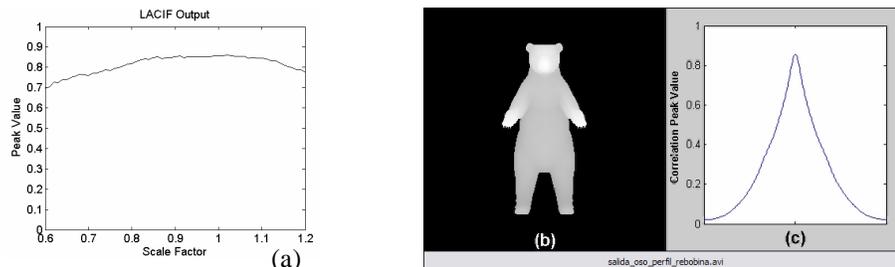


Fig. 3. (a). LACIF value for different scale factors (b). Multimedia file (1.09MB). Scaled Range Targets (c). Profile 1-D with the LACIF correlation peak value

Note from Fig. 3 that the LACIF correlation is not equal to one. In terms of vector space, a correlation value which is not a maximum means that the average target is not an element of the basis. However, the average target is a good candidate to represent of all the scaled targets, and it contains enough information to give a higher correlation peak value.

Another experiment deals with face recognition. Since the surface of human face can be expressed as a set of facets, the *PhFT* of the human face provides a new signature of the face. Ref. [16] deals with invariant to scale, translation and rotation for face human recognition. However the tolerance of the method presented in Ref. [16] to all the changes is quite limited. Our results with LACIF show an improvement of the robustness of the method. At the same time we will consider other false object to verify the discrimination capability of the LACIF method.

Figure 4(a) shows a human face target and Fig. 4(b) shows a false target (a different person) to discriminate. Results are shown in Fig. 4(c) and 4(d), respectively. Note that the different scaled faces are correctly detected above around 70%, whereas the false target is rejected with 35% of discrimination. This means that choosing a detection threshold of around value 0.3 the correct object will be detected and the false rejected. So the method is robust for detection and discrimination for all scale factors considered.

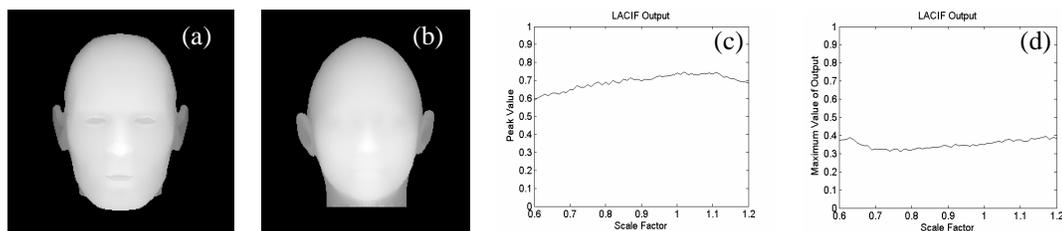


Fig. 4. (a). Target range image. (b). False range image. (c). LACIF peak value for different scaled targets of Fig. 4(a). (d). Idem for the false target.

In summary, scale invariant 3D object recognition is converted to intensity invariant pattern recognition technique. This can be possible if the recognition method implies a phase codification and a Fourier transformation of the 3D range images. A scale change of the *PhFT* of a range images is a multiplication of a constant factor in the *PhFT* of the reference range target. LACIF filtering detects targets which have been multiplied by constant values. Then, applying LACIF to Fourier phase encoded range images will solve the scale changes in 3D range images detection. Various experiments were carried out to validate the scale invariance. For real applications, the method has applied to face human recognition. We successfully tested the method when other false targets are tested.

### Acknowledgments

The authors gratefully acknowledge the editor's comments. This work was supported by the Spanish "Ministerio de Educación y Ciencia" (Ref: FIS2004-06947-C02-01) and Generalitat Valenciana (Ref: GV06/077). José J. Vallés acknowledges a FPU grant from the Ministry of Education and Science.

### Appendix A

This appendix justifies Eq. (3), which shows the effect of a scale change in the *PhFT* distribution. The *PhFT* contains information of all the orientations of the surfaces that defines a 3D object. It can be assumed that a 3D object is composed by multiples planar facets in different position and orientation. All of them define the object. Then, the *PhFT* of the complete 3D object is the contribution of all the *PhFTs* of the different facets.

Let consider one individual facet defined as

$$f(x, y) = (a \cdot x + b \cdot y + c) \cdot w(x, y) \quad (\text{A.1})$$

where  $w(x, y)$  is a boundary function equals to one if the point  $(x, y)$  belongs to the plane and zero otherwise. Its phase Fourier transform is

$$\begin{aligned} PhFT_f(u, v) &= FT \left\{ e^{i \cdot f(x, y)} \right\} = \int_{\mathbb{R}^2} e^{i(ax+by+c) \cdot w(x, y)} e^{-i2\pi(ux+vy)} dx dy \\ &= \int_{(x, y) \in w} e^{i(ax+by+c)} e^{-i2\pi(ux+vy)} dx dy + \int_{(x, y) \notin w} e^{-i2\pi(ux+vy)} dx dy \\ &= e^{ic} \delta \left( u - \frac{a}{2\pi}, v - \frac{b}{2\pi} \right) \otimes W(u, v) \end{aligned} \quad (\text{A.2})$$

where  $\delta(u, v)$  is the delta function,  $\otimes$  is the convolution and  $W(u, v) = FT \{ w(x, y) \}$ . Note from Eq. (A.2) that the integral outside the boundary area is infinite and the exponential is a periodical function, so the integral of the second term of Eq. (A.2) is zero. Then, the *PhFT* of an individual facet is centered in a peak given by the orientation of the normal to the plane and the shape of the peak is given by a certain factor of form which is the Fourier transform of the boundary. This result is similar to the interpretation of diffraction limited coherent optical imaging due to the pupil of the lenses.

A scale change can be described as

$$f'(x, y) = (a \cdot x + b \cdot y + c) \cdot w(x/k, y/k) \quad (\text{A.3})$$

where  $k$  is the scale factor. Note that a scale change has no influence in the orientation of the individual facets. The *PhFT* of the scaled object,  $f'(x, y)$ , is

$$PhFT_f(u, v) = k^2 e^{ic} \delta\left(u - \frac{a}{2\pi}, v - \frac{b}{2\pi}\right) \otimes W(ku, kv) \quad (\text{A.4})$$

The *PhFT* of the scaled object is given by a peak in the same position than the non-scaled object, but it is convolved by a scaled factor of form. Considering that a continuous 3D object is formed by numerous facets and the range of scale factors is not too large (i.e.  $k$  varies from 0.6-1.2), it can be shown from the experiments (see Fig. 1) that a scale change in the factor of form will be negligible in comparison with the influence of the peaks position, so in a first approximation we can assume that

$$PhFT_f(u, v) \cong k^2 e^{ic} \delta\left(u - \frac{a}{2\pi}, v - \frac{b}{2\pi}\right) \otimes W(u, v) = k^2 PhFT_f(u, v) \quad (\text{A.5})$$