

Correlation technique for the compensation of diffraction widening of optical reference signals

José Sáez-Landete,^{1,*} José Alonso,² Luis Miguel Sanchez-Brea,³ Tomás Morlanes,⁴ and Eusebio Bernabeu³

¹*Departamento de Teoría de la Señal y Comunicaciones, Escuela Politécnica Superior, Universidad de Alcalá, Campus Universitario, 28805 Alcalá de Henares, Madrid, Spain*

²*Departamento de Óptica, Escuela Universitaria de Óptica, Universidad Complutense de Madrid, Arcos de Jalón 118, 28037 Madrid, Spain*

³*Departamento de Óptica, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain*

⁴*AOTEK, S. Coop. Torrebaso Pasealekua 4, 20540 Eskoriatza, Guipúzcoa, Spain*

*Corresponding author: jose.saez@uah.es

Received April 9, 2009; revised June 26, 2009; accepted June 30, 2009;
posted July 9, 2009 (Doc. ID 109842); published August 4, 2009

Two-grating measurement systems are routinely employed for high-resolution measurements of angular and linear displacement. Usually, these systems incorporate zero reference codes (ZRCs) to obtain a zero reference signal (ZRS), which is used as a stage-homing signal. This signal provides absolute information of the position to the otherwise relative information provided by the two-grating incremental subsystems. A zero reference signal is commonly obtained illuminating the superposition of two identical pseudorandom codes and registering the transmitted light by means of a photodiode. To increase the resolution of the system, a reduction of the grating period and the ZRC widths is required. Due to this reduction, the diffractive effects produce a widening of the ZRS and, in turn, a loss of the measuring accuracy. In this work, we propose a method to narrow the distorted signal obtained with a Lau-based encoder, reinstating the accuracy of the ZRS. The method consists of the inclusion of a correlation mask on the detector. A theoretical model to design the mask has been developed, and experimental results have been obtained that validate the proposed technique. © 2009 Optical Society of America

OCIS codes: 040.1880, 040.5160, 050.1940, 050.1950, 120.3940, 230.0040.

1. INTRODUCTION

Two-grating diffraction systems are widely used in optical encoders for measuring linear and angular displacements. The typical output of an optical encoder is a pair of periodic electrical signals with a 90° phase difference. These electrical signals are used to measure incremental or relative displacement by means of the arctangent algorithm [1]. In most optical encoder configurations, a fixed-scale grating and a reading head are used. The reading head is composed of a set of gratings and photodetectors that move with respect to the fixed scale. Typically, two encoder signals are generated by means of interference, moiré fringes, fringe projection with spatially incoherent light, or any other optical phenomena involving two or more gratings [2–4].

In addition to these two electrical signals used for detection of the relative displacement, a third output, which consists of an isolated pulse, is normally required. This signal is called zero reference signal (ZRS), and it is used to set an absolute reference in the scale. This reference is necessary to define a home position or to obtain an absolute position measurement. To achieve higher accuracy in the measurement, the width of the reference signal and the period of the incremental signal must be equal. The ZRS is normally obtained by means of two identical zero reference codes (ZRCs), one situated on the scale grating and the other on the reading head. As one moves with respect to the other, the two codes overlap. An incident light

beam propagates through both codes and the transmitted light is registered by means of a photodiode. A ZRC consists of a group of unequally spaced slits. In general, a ZRC can be described by the sequence of binary data $\mathbf{c} = [c_0, \dots, c_{n-1}]$, where n is the length of the ZRC, $c_i = 1$ if a transparent slit is located at the i -position, and $c_i = 0$ elsewhere. The transmittance of a code can be expressed as the sum

$$t(x) = \sum_{j=0}^{n-1} c_j \text{rect}\left(\frac{x - jb}{b}\right), \quad (1)$$

where b is the width of the slit of the code. The design of necessary codes to obtain suitable reference signals has been studied in [5–10]. In all these works, a collimated light beam has been considered, and the distance between the two ZRC is neglected. Therefore, a geometrical approach is assumed, neglecting the diffractive effects of the beam as it propagates between the codes. By means of this approximation, the signal obtained is the autocorrelation of the transmittance of the code, and the width of the signal at the base of the central peak is twice the width of the slit. Similar techniques to design arrays with special autocorrelation functions are used in coded aperture imaging [11].

In [10], we proposed a method to design ZRCs taking into account the propagation of the beam between them. The technique is based on a tradeoff between the attainment of ZRCs not very sensitive to diffractive effects

(ZRCs with specially grouped slits) and the attainment of signals with a well-defined central peak (ZRCs with ungrouped slits). This method is based on an iterative optimization process, where the diffractive effects are calculated at each iteration. Therefore, this method can be used only in configurations where the computation of the diffracted code is not a time-consuming process. For this reason, in that paper an incident collimated light beam is considered, and as a consequence, this system can be implemented in a moiré-based encoder.

The use of spatially incoherent light beams (diffused light) has been proposed to increase the resolution of grating measurement systems [2]. Moreover, recent work has demonstrated the advantage of using spatially and temporally incoherent light beams (diffused and partially polychromatic light) to increase the tolerance regarding the gap between the grating scale and the reading grating [4]. This technology allows a reduction of the diffraction grating period and an increase of the distance between diffraction gratings. These improvements in the incremental signals pose a problem in the reference signal. As the grating period becomes smaller, the slit width b of the ZRC must decrease accordingly. This reduction of the slit width, along with a larger gap between codes, makes diffraction effects much more noticeable, producing a widening of the reference signal and, in turn, a loss of system accuracy. Because of the diffractive effects the width of the central peak increases and the ratio between the height of the central peak and the height of the secondary maxima decreases. This ratio represents the “effective” quantity of signal.

In this work, we propose a different approach to this problem. We consider a reference signal with a widened central peak and propose a technique to squeeze and regenerate this central peak. This method can be used in any encoder configuration and it can be complementary to the method of code design presented in [10]. The method consists of the use of two suitable correlation masks located just before two monolithic detectors. Before the signal detection, the light distribution is correlated with the masks and the registered signal is reshaped.

The rest of the paper is organized as follows. In Section 2 we show a suitable technique for the generation of the reference signal as well as a theoretical and experimental demonstration of the degradation and the widening of the reference signal. In Section 3 is presented an algorithm for the design of correlation masks that will allow us to obtain pulses narrow enough. Finally, an experimental verification of the proposed technique is shown in Section 4.

2. GENERATION OF THE REFERENCE SIGNAL IN AN ENCODER BASED ON THE LAU EFFECT

In this section, we show an example of generation of a reference signal that makes evident the diffraction effects in the signal. We present below a theoretical framework that partially explains the broadening and the degradation of the central peak of the signal. The signal is generated by means of the setup shown in Fig. 1. The light source consists of a LED that incorporates a diffuser. It emits a spa-

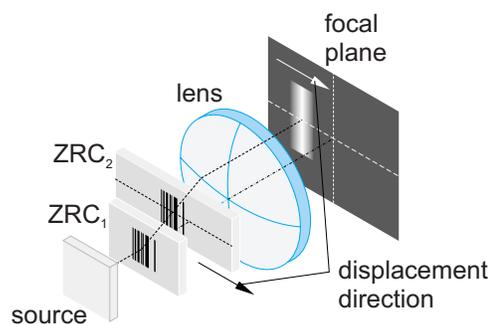


Fig. 1. (Color online) Lau configuration for the reference signals. The detection plane is placed at the focal plane of the lens where the mask of detection and the photodiodes are located.

tially incoherent beam on the detection plane [12]. The light passes through two identical pseudorandom codes, and it is registered in the focal plane of a lens where the detection system is located. To measure the intensity profile and to simulate the behavior of different detection masks on monolithic detectors, we have used a CCD sensor. It is worth mentioning that encoders must work in real time, so they contain neither a CCD nor a processing stage.

To obtain suitable correlation masks for the regeneration of the reference signal, we need to know the intensity profile on the detection plane. By assuming an aberration-free lens and perfectly Lambertian diffuser and by taking into account the diffraction from the first code to the second, it can be demonstrated theoretically, under the Fresnel approximation (see [12]), that the intensity distribution on the detection plane is given by

$$I(x) = \int T(x_0) \hat{T}\left(x_0 + \frac{zx}{f}\right) dx_0 = \int T(x_0) \hat{T}(x_0 + u) dx_0 = I\left(\frac{f}{z}u\right), \quad (2)$$

where $u = zx/f$, z is the direction of propagation, x and x_0 are coordinates in the direction of the displacement, T is the transmittance of the code, and f is the focal length of the lens. \hat{T} is the Fresnel transform of T ,

$$\hat{T}(x + u) = \int T(x') \exp\left\{\frac{i\pi}{\lambda z}[x' - (x + u)]^2\right\} dx', \quad (3)$$

where λ is the wavelength of the light.

Thus, the reference signal is a scaled correlation between a ZRC and the Fresnel transform of the other ZRC. In general, the Fresnel transform generates a widening in the intensity distribution with regard to the geometrical autocorrelation signal (considering a collimated incident beam and absence of diffraction). To show the widening of the ZRS predicted by Eq. (2) we have performed an experimental verification using a green LED ($\lambda = 540$ nm) with a volume diffuser, two identical standard pseudorandom codes with slit widths of $20 \mu\text{m}$, and a lens with a back-focal length of 7.7 mm. The gap between codes is 2 mm.

The broadening of the signal due to the diffraction effects is shown in Fig. 2, along with the autocorrelation signal obtained with the geometrical approach. In this fig-

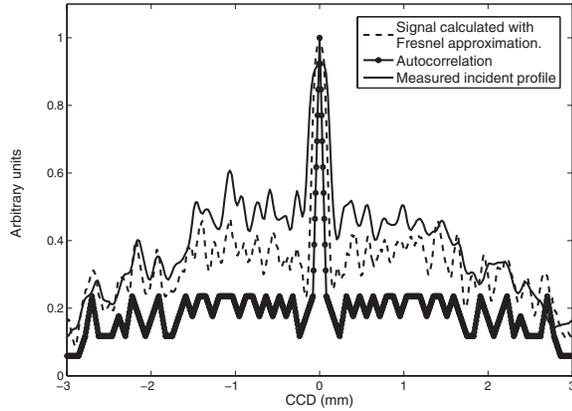


Fig. 2. Intensity profiles. The dashed trace is calculated propagating the field with the Fresnel approximation. The curve with circles is the autocorrelation, neglecting the distance between the codes (and therefore ignoring the diffraction). The continuous curve is the incident profile measured at the CCD.

ure we have also included the experimental profile measured at the focal plane by means of a CCD array. The width of the central peak predicted by Eq. (2) is roughly half the real width and twice the autocorrelation approximation. The larger actual width is due to the CCD resolution, lens aberrations (mainly distortion), spectral width of the light source, and the diffuser being non-Lambertian. Because of the difficulty of characterizing these effects, we have used the experimental profile to compute the correlation mask.

3. DESIGN OF THE MASK OF DETECTION

When one code shifts with respect to the other by an amount η , the intensity at the CCD plane is

$$I'(x) = \int T(x_0 + \eta) \hat{T}\left(x_0 + \frac{zx}{f}\right) dx_0 \\ = \int T(x_0) \hat{T}(x_0 + u - \eta) dx_0 = I \left[\frac{f}{z}(u - \eta) \right]. \quad (4)$$

The intensity pattern shifts an amount proportional to the displacement of the code. The electrical signal registered in the photodiode is the correlation between the optical intensity profile and the spatial detectivity of the detector. This phenomenon will be used to narrow the signal.

When one ZRC passes in front of the other, the electrical signal registered by the photodiode is

$$S(\eta) = K \int I(x - \eta) D(x) dx, \quad (5)$$

where K is the efficiency of the photodiode and D is the spatial detectivity of the detector. The variable detectivity can be achieved in a monolithic detector masking it by means of a set system of slits. The transmittance value at the location of each slit is proportional to the area of that slit. In general, the detectivity should take both positive and negative values in order to obtain a suitable signal S . The negative values can be implemented using a system of two detectors in push-pull configuration. The problem

consists of calculating a mask whose correlation with the intensity pattern is a suitable electrical signal. The quality of a reference signal can be described using different parameters, such as the slope of the main peak [5],

$$SL = \frac{S(0) - S(x_1)}{x_1}, \quad (6)$$

where x_1 is the half-width of the central peak at the base; or the signal effective range

$$R = \frac{S(0) - S_m}{S(0)}, \quad (7)$$

where S_m is the second maximum of the signal. A suitable reference signal will be described by using an appropriate width of the central peak of the signal and a high value of the parameters SL and R .

Numerical Description of the Problem. To find a mask whose correlation with the irradiance profile provides a suitable electrical signal, we must solve the differential Eq. (5). By means of sampling the involved variables, the differential equation is transformed into an inverse problem. These kinds of problems are described in [13,14], provided that S and I are given. In this application we are interested in obtaining a signal that has suitable values of the SL and R parameters. To transform the differential equation in an inverse problem, the variables I , D , and S are sampled with a sufficient number of samples to guarantee a good characterization of the incident intensity profile I and the central peak of the signal S . The sampled detectivity vector is $\mathbf{D}=[D_1, \dots, D_m]$, the sampled electrical signal vector is $\mathbf{S}=[S_1, \dots, S_{n+m-1}]$, and the sampled intensity profile is $\mathbf{I}=[I_1, \dots, I_n]$. Using this approach, the continuous correlation in Eq. (5) becomes a discrete correlation, $S_k = K \sum_{l=k+1}^{\min(n+k,m)} I_{l-k} D_l$. The correlation between \mathbf{I} and \mathbf{D} can be obtained by means of the matrix multiplication $\mathbf{S} = \mathbf{C} \cdot \mathbf{D}$, where \mathbf{C} is the matrix correlation

$$\mathbf{C} = \begin{bmatrix} I_n & 0 & 0 & 0 & 0_{(1,m)} \\ \vdots & I_n & 0 & 0 & 0 \\ I_1 & \vdots & \ddots & 0 & 0 \\ 0 & I_1 & \vdots & \ddots & 0 \\ \vdots & 0 & \vdots & \vdots & I_n \\ 0_{(n+m-1,1)} & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \quad (8)$$

This matrix multiplication represents a linear and over-determined system of $n+m-1$ equations and m unknowns. The solution to these systems is given by the following theorems [14]:

1. A solution to $\mathbf{S} = \mathbf{C} \cdot \mathbf{D}$ exists if and only if there is a conditional inverse of \mathbf{C} , \mathbf{C}' , for which $\mathbf{C} \cdot \mathbf{C}' \cdot \mathbf{S} = \mathbf{S}$. The conditional inverse of a matrix \mathbf{C} is a matrix for which $\mathbf{C} \cdot \mathbf{C}' \cdot \mathbf{C} = \mathbf{C}$.

2. If a solution to $\mathbf{S} = \mathbf{C} \cdot \mathbf{D}$ exists, the solution is unique if and only if the rank of the $n \times m$ matrix \mathbf{C} is equal to m .

A method to obtain matrix \mathbf{C}' is to calculate the generalized inverse [14]. Generalized inverses are conditional inverses, and they can be calculated whenever the rank of the matrix is maximum using the equation

$$\mathbf{C}^{-} = (\mathbf{C}^{\mathbf{T}} \cdot \mathbf{C})^{-1} \cdot \mathbf{C}^{\mathbf{T}}. \quad (9)$$

The coefficients matrix is triangular, so its determinant is nonzero. If we fix an electrical signal \mathbf{S} and we find a mask \mathbf{D} that is a solution, this solution is unique. The solution of the system is

$$\mathbf{D} = \mathbf{C}^{-} \cdot \mathbf{S}. \quad (10)$$

If the system is consistent, the signal must be an eigenvector of $\mathbf{C} \cdot \mathbf{C}^{-}$ with its corresponding eigenvalue equal to one, or equivalently, \mathbf{S} must belong to the subspace defined by the columns of \mathbf{C} . We must find signals for which the system is consistent and, in turn, fulfill the requirements of a reference signal. To approach the problem, for a given a signal \mathbf{S} , we consider the obtained error when the incident profile is correlated with an arbitrary mask:

$$\mathbf{S} = \mathbf{C} \cdot \mathbf{D} + \mathbf{e}(\mathbf{D}), \quad (11)$$

where $\mathbf{e}(\mathbf{D})$ is the error function, which depends on the mask used. It can be demonstrated that the norm of this vector is minimum when $\mathbf{D} = \mathbf{C}^{-} \cdot \mathbf{S}$. Substituting Eq. (10) into Eq. (11), the error function module results,

$$\|\mathbf{e}(\mathbf{S})\|^2 = \mathbf{e}(\mathbf{S})^{\mathbf{T}} \mathbf{e}(\mathbf{S}) = \mathbf{S}^{\mathbf{T}} [\mathbf{I} - \mathbf{C} \cdot \mathbf{C}^{-}] \mathbf{S}, \quad (12)$$

where $\mathbf{S}^{\mathbf{T}}$ is the transpose of \mathbf{S} and \mathbf{I} is the identity matrix.

When the linear system is consistent for a signal \mathbf{S} , then Eq. (12) is zero. Under this consideration, the mask necessary to obtain the signal \mathbf{S} can be calculated by means of Eq. (10). The error modulus function is positive defined, and its zeros can be found by means of a minimization. Though there is not a complete description of the signal \mathbf{S} , it must fulfill some constraints. In this sense, we can find the optimum mask that minimizes the Eq. (12) subject to these constraints. These constraints consist of upper and lower bounds that assure a control of the parameters defined in Eqs. (6) and (7). The incident irradiance patterns, as well as the upper and lower bounds of the signal \mathbf{S} , are displayed in Fig. 3. It can be seen that outside the central peak area, the signal is lower and upper bounded. The value of the bounds determines the parameter R given in Eq. (6). Within the central peak re-

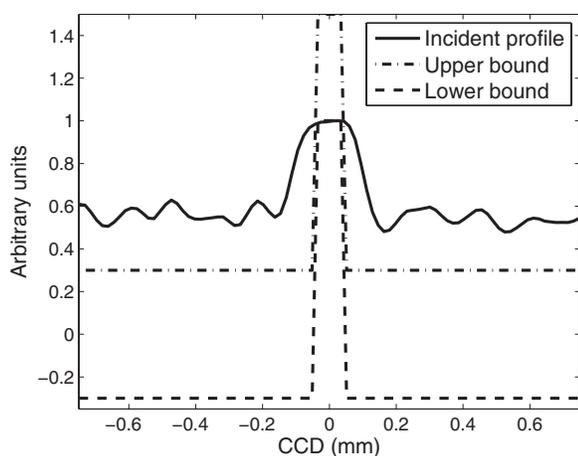


Fig. 3. Incident irradiance profile measured at the CCD (solid curve) and the lower and upper bounds (dashed and dashed-dotted curves) required.

gion, a lower bound is required to be higher than the upper bound. In this way, the width of the central peak is defined by the height difference of the bounds.

4. EXPERIMENTAL DEMONSTRATION

To validate the proposed method we have considered an example of a double diffraction scheme based on the Lau effect (see Fig. 1). In this example, we have calculated a correlation mask and we have demonstrated the experimental narrowing of the reference signal. The ZRC is a pseudorandom set of slits with a width b of $20 \mu\text{m}$, and the gap between the ZRCs is 3 mm. The illumination source is a green diffuse LED located 1 mm in front of the first ZRC. A lens with a back-focal length of 7.95 mm is located behind the second grating, and the CCD chip is in its focal plane. In this example, we have an incident measured profile with $250 \mu\text{m}$ width at the base of the central peak. The optical signal is normalized and is transformed into a narrowed signal of $100 \mu\text{m}$ width by means of the correlation with a mask on the detector. Therefore, the resolution of the absolute position measurement in an encoder is increased by a factor of 2.5. The incident measured profile as well as the upper bound and the lower bound are displayed in Fig. 3. The narrowed signal is constrained at ± 0.3 units in $|x| > 50 \mu\text{m}$, and it is greater than 1.5 in $|x| < 50 \mu\text{m}$. The sampling of the variables is given by the pixel size of the CCD sensor, namely $17 \mu\text{m}$. In this case the algorithm converges and finds a suitable signal with its associated mask.

In Fig. 4, the incident measured optical signal (solid curve), the bounds of the narrowed signal (dotted and dotted-dashed curves), and the narrowed signal (solid circles) calculated by means of the algorithm are shown. The mask that is the solution for this problem is shown in Fig. 5, where the superior region takes positive values and the inferior region takes negative values. It is worth mentioning that there is a tradeoff between the narrowing factor and the possibility of convergence of the algorithm. The sampling of the problem is related to the number of the variables and, therefore, to the complexity of

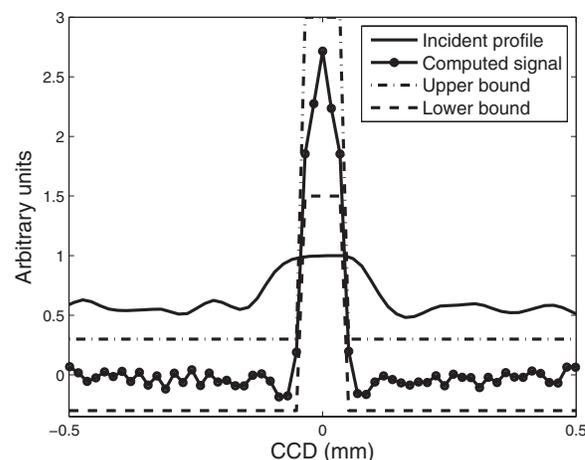


Fig. 4. Incident irradiance profile measured at the CCD (solid curve), the lower and upper bounds of the narrowed signal (dashed and dashed-dotted curves) and the narrowed signal obtained by means of the algorithm (filled circles).

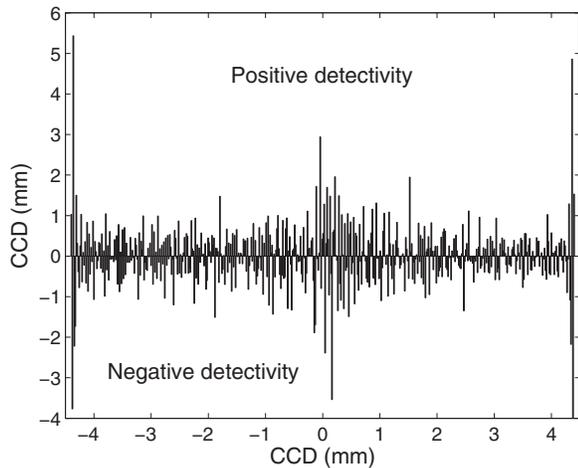


Fig. 5. Mask obtained by means of the algorithm. The superior region is associated with the positive detectivity and the inferior region corresponds to the negative detectivity.

the mask and to the degrees of freedom of the problem. The greater the number of variables, the greater the probability of convergence but the more difficult the mask implementation.

To experimentally demonstrate the narrowing of the signal, the computed mask is simulated at the CCD by multiplying the acquired image by zeros or ones and subtracting the superior and inferior region (see Fig. 5). The measurements have been achieved by locking one of the ZRCs to a very accurate linear stage. This stage moves along the perpendicular direction to the slits of the ZRC while images of the irradiance profile are registered on a CCD array. These images are multiplied by the digital mask and the row and column values are added, making up the numerical correlation between the mask and the incident intensity profile.

In Fig. 6 the incident measured optical profile, the narrowed signal computed by the algorithm, and the narrowed signal measured experimentally are shown. The agreement between theory and experiment is quite good at the central region. The width of the central peak of both signals is the same. Outside the central region, the difference between the two signals is greater; neverthe-

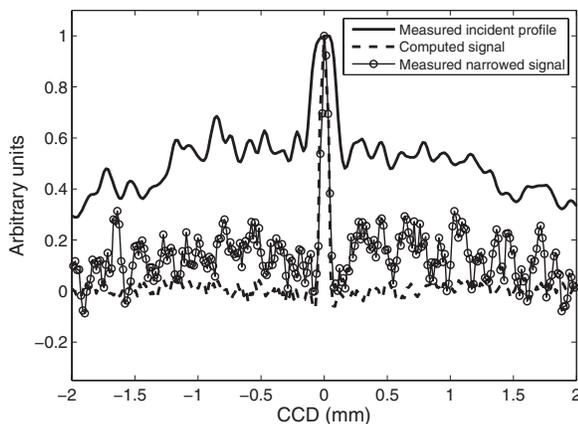


Fig. 6. Incident intensity profile measured by the CCD (solid curve), as well as the signal computed by means of the algorithm (dashed curve) and the experimental narrowed signal (open circles).

less the experimental signal is between the demanded bounds. This discrepancy outside the center region is influenced by two factors. First, the distortion of the lens deforms the profile of the incident optical signal when the ZRC moves. Second, the source exitance is not constant with respect to the emission angle (the diffuser is not truly Lambertian), and the exitance decreases when the angle increases. Both lens distortion and angular spread of the light distort the irradiance profile as it moves away from the central position on the observation plane. This effect has not been considered in Eq. (5), and it affects the lateral region of the signal. Fortunately, this effect is not appreciable in the central region, where the ZRCs are almost aligned and arranged symmetrically with respect to the source.

5. CONCLUSIONS

In this work we have proposed a technique to regenerate diffraction degraded reference signals by means of correlation with a mask located on the detectors. We have developed an algorithm for designing the masks, and it has been demonstrated in an experimental narrowing of a reference signal by a factor of 2.5; hence, the resolution of the encoder can be increased by the same factor. Moreover, the algorithm allows us to design reference signals with pre-established shapes by means of constraints in the minimization process. Therefore, this technique can be used to obtain reference signals with predefined parameters, such as the signal-to-noise ratio, the effective amplitude of the signal, the slope of the central peak, etc. The convergence of the algorithm is related to the number of slits of the mask. In this sense, if we accept masks with fine details we will have more degrees of freedom, and therefore, possibilities of convergence. This technique has been experimentally verified for a Lau-based encoder, although it can be easily applied to other optical encoders. Double-grating encoders are based on a self-imaging process (or similar), and the detected signals are the correlation between a fringe pattern and the detector detectivity. In these cases, the technique proposed can be implemented to increase the resolution of the absolute reference.

ACKNOWLEDGMENTS

This work has been supported by Center for Innovation in Transport (CENIT) project “Tecnologías avanzadas para los equipos y procesos de fabricación de 2015: e-eficiente, e-cológica, e-máquina (eEe)” of the Ministerio de Industria, Turismo y Comercio of Spain and by the Madrid Autonomous Community-University of Alcalá de Henares (CAM-UAH) project CCG08-UAH/TIC-3941.

REFERENCES

1. L. M. Sanchez-Brea and T. Morlanes, “Metrological errors in optical encoders,” *Meas. Sci. Technol.* **19**, 1–8 (2008).
2. D. Crespo, J. Alonso, and E. Bernabeu, “Generalized grating imaging using an extended monochromatic light source,” *J. Opt. Soc. Am. A* **17**, 1231–1240 (2000).
3. D. Crespo, J. Alonso, T. Morlanes, and E. Bernabeu,

- “Optical encoder based on the Lau effect,” *Opt. Eng. (Bellingham)* **39**, 817–24 (2000).
4. L. M. Sanchez-Brea, J. Sáez-Landete, J. Alonso, and E. Bernabeu, “Invariant grating pseudoimaging using polychromatic light and a finite extension source,” *Appl. Opt.* **47**, 1470–1477 (2008).
 5. Y. Xiangyang and Y. Chunyong, “A new method for design of zero reference marks for grating measurement systems,” *J. Phys. E* **19**, 34–37 (1986).
 6. Y. Li, “Autocorrelation function of a bar code system,” *J. Mod. Opt.* **34**, 1571–1575 (1987).
 7. Y. Li, “Optical valve using bar codes,” *Optik (Stuttgart)* **79**, 67–74 (1988).
 8. J. Sáez-Landete, J. Alonso, and E. Bernabeu, “Design of zero reference codes by means of a global optimization method,” *Opt. Express* **13**, 195–201 (2005).
 9. J. Sáez-Landete, S. Salcedo-Sanz, M. Rosa-Zurera, J. Alonso, and E. Bernabeu, “Optimal design of optical reference signals using a genetic algorithm,” *Opt. Lett.* **30**, 2724–2726 (2005).
 10. J. Sáez-Landete, S. Salcedo-Sanz, M. Rosa-Zurera, J. Alonso, and E. Bernabeu, “Generation of optical reference signals robust to diffractive effects,” *IEEE Photon. Technol. Lett.* **19**, 1133–1135 (2007).
 11. E. E. Fenimore and T. M. Cannon, “Coded aperture imaging with uniformly redundant arrays,” *Appl. Opt.* **17**, 337–347 (1978).
 12. J. Jahns and A. W. Lohmann, “The Lau effect (a diffraction experiment with incoherent illumination),” *Opt. Commun.* **28**, 263–267 (1979).
 13. L. Mertz, *Transformations in Optics* (Wiley, 1965).
 14. W. K. Pratt, *Digital Image Processing* (Wiley, 1991).