Detection of small-size synthetic references in scenes acquired with a pixelated device

Xavier Fernández, MEMBER SPIE
Industrias de Optica, S.A. (INDO)
08902 L'Hospitalet (Barcelona), Spain
E-mail: xavier@indo.es

Josep Amat
Universitat Politècnica de Catalunya (UPC)
Dept. d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial
Pau Gargallo 5
08028 Barcelona, Spain

Abstract. A synthetic reference (SR) or fiducial mark is a target that has been designed and placed within a scene in order to help locate objects. When a scene containing a small SR is acquired through a pixelated device (such as a CCD camera) and this SR is located using a digitally computed normalized correlation search, there is a lowering of the correlation peak if the target SR is displaced a nonwhole number of pixels with respect to the position of the previously acquired model SR. We propose a framework where the lowering is related to detection reliability, in order to compare the performance of different sizes and shapes of SR. This framework has been applied to eight SRs having simple shapes, using synthetic scenes. Results show large differences in their detection reliability according to their shape. A SR in the shape of a 45-deg-tilted square shows the best behavior among all the studied shapes, and one can take advantage of this fact when designing an application dealing with small-size SRs.

Subject terms: synthetic references; fiducial marks; normalized correlation; similarity measures; detection reliability; small targets.

Paper 34017 received Jan 31, 1997; revised manuscript received June 10, 1997; accepted for publication June 10, 1997.

1 Introduction

The task of accurately locating an object in a scene often becomes difficult when there is a lack of definite and repeatable elements (edges, corners) suitable to be used as position references. A useful solution is to apply to the object some marks specifically designed in order to easily locate their position, and through it, to derive the position and/or orientation of the object. We call those marks synthetic references (SRs). Among other names, the terms cooperative targets and fiducial marks are also employed.

SRs are commonly designed as binary targets of finite extent. The task of locating binary targets in a gray-level scene, taking into account the geometric properties of the targets, has been handled recently in several fields, for example in order to simplify the computation of correlation, allow the use of statistical tests, or optimize a binary-domain similarity measure.

Noise is a major restrictive factor for detection reliability, and this is specially true for small targets in natural scenes, thermal images, or optical correlation. Detection reliability, however, does not depend only on noise. The detection of a target in a noiseless scene, using a numerically computed normalized correlation search can fail if the target has been degraded so that the correlation peak is lower than the partial correlations with the background objects.

When a pixelated device (such as a CCD camera) is used to acquire the image, the resulting numerical representation of the target depends on the position of the target relative to the pixel lattice (Fig. 1). This variation can be seen as a degradation, which becomes worse as the target size decreases, and this degradation can make the detection unreliable in the terms stated above.

The freedom of the designer to choose the SR shape has often been used to achieve some specific properties, particularly rotational and scale invariance, and localization accuracy. In the present work, we intend to use this freedom for obtaining good detection performance for small sizes, in the presence of the degradation caused by the acquisition through a pixelated device. We try to determine which shape shows the highest detection reliability, and to obtain the minimum detectable size.

The sections are structured as follows: Section 2 reviews and formalizes the use of similarity measures in a template-matching scheme, focusing on the detection decision, and normalized correlation is presented as a similarity measure. In Sec. 3 the spatially discretized evaluation of the similarity measure, inherent in the use of a pixelated device and negligible at large target sizes, is studied, and its influence on small-target detection analyzed. Section 4 introduces the framework to define and compare minimum detectable sizes. Within this framework, an experimental study of different shapes is presented in Sec. 5, where the minimum detectable sizes for each SR are obtained.

In Sec. 6, the framework is modified in order to take into account the need to avoid false detection of background objects appearing in the scene. Section 7 proposes a representative set of those objects and uses them to extend the previous experimental work, obtaining new minimum sizes according to the added conditions. Finally, a conclusions section summarizes and comments on the results obtained, and proposes further work to complement the present developments.
Locating SRs through Template Matching and Similarity Measures. Normalized Correlation

When the position of a target having known characteristics (shape, size, orientation) is to be located in a scene, the template-matching technique is widely used. Let \( f(x, y) \) be the model (ideal target), \( U_f \) its associated support region, \( g(x, y) \) the scene, and \( J_{fg}(u, v) \) a similarity measure between \( f \) and \( g \) displaced to \((u, v)\) and evaluated inside \( U_f \). We assume that the similarity measure holds:

\[
J_{fg}(u, v) < J_{ff}(0, 0).
\]

We call \( J_{ff}(u, v) \) the autosimilarity function of \( f \). Equality occurs only when \( g(x, y) = f(x-u, y-v) \), that is, the image contains a perfect target at the position \((u, v)\). Consequently, in an ideal case, the locations where the similarity measure achieves \( J_{ff}(0, 0) \) show the position of the target in the scene.

In the real world the target’s appearance may be altered by noise, small changes in scale and orientation, optical distortions, and other artifacts, which imply that the maximum in Eq. (1) may not be reached. Hence, the presence of a target at a location \((u, v)\) is decided if at this point the similarity measure presents a local maximum that equals or exceeds a threshold value \( \tau \), somewhat lower than \( J_{ff}(0, 0) \).

Choosing a value for \( \tau \) is always the result of a compromise: If the threshold is too high, there is a risk of not detecting some degraded targets, whereas if it is too low, some background objects presenting partial similarities with the model may be incorrectly detected as a target (false detection or anomalous error). Within the scope of this work detection reliability has to be interpreted as the existence of a threshold that guarantees detection (100% probability of detection) and discriminates against background objects (0% probability of false detection). Those extreme probability values, however, can only be achieved if target degradations and partial similarities of background objects are both bounded.

The normalized correlation (NC) between a model \( f \) and a target \( g \),

\[
J_{fg} = \frac{\left[ \Sigma (f - \bar{f})(g - \bar{g}) \right]^2}{\left[ \Sigma (f - \bar{f})^2 \right] \left[ \Sigma (g - \bar{g})^2 \right]},
\]

is a well-known similarity measure that fulfills Eq. (1), and has been widely used for template matching and registration purposes. Equation (2) corresponds to the square of the correlation coefficient. The NC shows invariance to linear gray-level transformations, and has become a similarity measure of reference when other similarity measures are to be evaluated. Squaring is a frequently used way to avoid the time-consuming square-rooting of the denominator. The value of \( J_{fg} \) ranges from 0 to 1.

The square of the correlation coefficient is a dimensionless number, which measures similarity on a linear scale as the proportion of the total target variance that is explained by a linear dependence between target and model. Thus, the NC gives absolute meaning to similarity values, independently of the particular model and image being compared. For example, if a target shows a NC of 0.7, it means that the target has a 70% measure of linear dependence on...
the model. From this interpretation, a detection threshold $\tau$ also acquires an absolute meaning on the scale of linear dependence.

3 Limitations on the Detection of Small Noise-Free SRs Caused by Spatially Discretized Evaluation of the Similarity Measure

When the acquired image is spatially quantized, displacements of $g$ to obtain its local similarity with $f$ can only be done at discrete intervals, which implies that the similarity measure will be evaluated only at discrete points, coinciding with the nodes of the sensor pixel lattice (Fig. 1). Consequently, if the sensor lattice has a regular square spacing step $\Delta$, then $J_{fg}$ is evaluated at the points $(u,v)$, $u=i\Delta$, $v=j\Delta$.

\[ (u,v), \quad u=i\Delta, \quad v=j\Delta, \quad (3) \]

Suppose that a scene contains a noise-free, not degraded target placed at an arbitrary position $(x_0,y_0)$:

\[ g(x,y) = f(x-x_0,y-y_0). \quad (4) \]

To locate the target, the position among the evaluation points of Eq. (3) that maximizes $J_{fg}$ is found:

\[ \max_{i,j} J_{fg}(i\Delta,j\Delta) = \max_{i,j} J_{ff}(i\Delta-x_0,j\Delta-y_0) \leq J_{ff}(0,0). \quad (5) \]

Equality in Eq. (5) will only hold when the position $(x_0,y_0)$ of the target coincides with an evaluation point as defined in Eq. (3). In any other position, the maximum of $J_{fg}$ will never reach the highest theoretical value $J_{ff}(0,0)$ ($=1$ for NC), even when the target in the scene meets the model exactly.

Thus, in order to guarantee the detection (with 100% probability) of such an ideal target placed at an arbitrary position, a threshold $\tau$ has to be chosen, lower than or equal to the worst expectable maximum value for $J_{fg}$:

\[ \tau \leq \min_{x_0,y_0,i,j} J_{ff}(i\Delta-x_0,j\Delta-y_0). \quad (6) \]

Notice that, generally, autosimilarity measures (such as autocorrelation) are even functions:

\[ J_{ff}(u,v) = J_{ff}(-u,-v). \quad (7) \]

Further assume that $J_{ff}$ decreases monotonically with respect to the displacement $r$, for a given direction $\Theta$ (which usually holds for small displacements):

\[ r' > r \Rightarrow J_{ff}(r' \cos \Theta, r' \sin \Theta) \leq J_{ff}(r \cos \Theta, r \sin \Theta) \forall \Theta. \quad (8) \]

Then detection can be guaranteed if

\[ \tau \leq \min_{u,v \in L_\Delta} J_{ff}(u,v) \equiv \tau_0(f), \quad (9) \]

\[ L_\Delta = \left\{(u,v) : \max(|u|,|v|) = \Delta/2 \right\}. \]

Denote by $\tau_0(f)$ the minimum locating similarity of $f$. Then $\tau_0$ is an upper bound for a threshold $\tau$ that guarantees the detection of a perfect target in an arbitrary position, and can be computed as the minimum value for the autosimilarity function taken on the perimeter of a square of side $\Delta$ centered at $(0,0)$. This perimeter is the locus of the points that are equidistant from the two nearest lattice nodes (Fig. 2).

Dvornychenko showed for one-dimensional autocorrelation functions that the decrease in their value caused by a displacement is dependent on the ratio of the displacement to the extent of the target. This behavior can be also found in other two-dimensional autosimilarity functions. In our case, a displacement as low as $\Delta/2$ may cause a significant decrease of $\tau_0$ when the target is a small-size SR.

Two SRs having different sizes, or having the same size but differing in shape, may yield different values for $\tau_0$. This means that they do not have the same threshold freedom, and the SR with higher $\tau_0$ will discriminate against background objects showing higher partial similarity with the model (Fig. 3).

4 Minimum Detectable Size, Equivalent Size, and Advantageous Shapes

Let $f$ be a SR. Then $f^n$ will be the $n$-fold scaled version of $f$ if:

\[ f^n(x,y) = f\left(\frac{x}{n},\frac{y}{n}\right). \quad (10) \]
We assume that the associated support region $U_f$ is scaled accordingly.

Given a threshold $\tau$, $N = N(f, \tau)$ is the minimum detectable size of $f$ if

$$N = \min\{n; \tau \leq \tau_0(f^n)\}. \quad (11)$$

This means that a size of $f$ equal to or larger than $N$ guarantees the detection of the ideal SR in an arbitrary position with 100% probability, while sizes smaller than $N$ do not.

As a convention, we will say that two SRs $f_a, f_b$ have equivalent sizes when the respective boxing squares, oriented along the axis directions of the pixel grid, have equal sizes (Fig. 4). In that case, for a given size $N$, we can say that $f_a$ is advantageously shaped with respect to $f_b$ if $f_b$ needs a lower threshold than $f_a$ to guarantee detection:

$$\tau_0(f_b^N) < \tau_0(f_a^N). \quad (12)$$

5 Experimental Study of Minimum Sizes and Advantageous Shapes for a Set of SRs Having Equivalent Sizes

The framework above introduced has been applied to study the behavior of eight SRs having simple shapes and equivalent sizes, as depicted in Fig. 4. A square support region for similarity evaluation has been chosen, leaving a 10% blank frame at each side of the SR boxing square (Fig. 4). Narrow structures, when present in the shape, have a 16% width. The side length of the support region defines the size of the SR in terms of the sensor lattice step $\Delta$ (pixel units).

A synthetic image has been generated for each SR, and the acquisition process has been emulated by convolving each image with a uniform $\Delta \times \Delta$-square sensor response function, and uniformly sampling on a $\Delta \times \Delta$ grid. An ideal, diffraction-free optical system is assumed.

The normalized correlation has been computed digitally using Eq. (2). For each SR, and for each size from 2 to 20×20 pixels in one-pixel increments, displacements to points on the $L_\Delta$ locus have been evaluated in order to find $\tau_0$ as defined in Eq. (9). In Fig. 5 a curve for each SR is depicted, showing the dependence of $\tau_0$ on size.

It can be seen in Fig. 5 that a large decrease in $\tau_0$ results when small-size SRs are used. In general terms, the more compact shapes have their minimum locating similarity higher than the less compact ones, although SQUARE behaves poorly in comparison with the other compact SRs. The difference is dramatically evident in the case of the extreme shapes: TILTED-SQUARE shows a value $\tau_0 = 0.67$ at 7×7 size (a 49-pixel SR), while for FRAME only $\tau_0 = 0.21$ is achieved at the same size.

In Fig. 6, the minimum detectable sizes given $\tau = 0.5$ are shown, computed as in Eq. (11) (in fact, the least integer size for which $\tau$ is exceeded). It can be seen that TILTED-SQUARE will be safely detectable down to sizes as small as 5×5 pixels, followed by CIRCLE, which adds one pixel. On the opposite side, FRAME needs a size of 12×12 pixels to guarantee detection in an arbitrary position.

6 Conditioning the Detection to a Discriminating Scheme

Nothing has been said so far about a lower bound for a threshold to avoid false detection of background objects present in the scene; they are very much dependent on the particular application. When no information is known in advance from these structures, using $\tau_0$ is all we can do to express the detection reliability of a SR.
Fig. 5 Minimum locating similarity $r_0$ as a function of SR size.

Fig. 6 Minimum detectable sizes for $\tau = 0.5$. 
Suppose, however, that we have prior knowledge of those background objects to be discriminated, and let \( H = \{ h(x, y) \} \) be the set of their shapes that are able to be located at any point \((u, v)\) in the image \( g(u, v)\). Thus, in order to avoid false detection, threshold \( \tau \) must fulfill, in addition to Eq. (9),

\[
\tau > \max_{h \in H} \max_{u, v} J_{f h}(u, v) = \rho_0(f, H).
\]

Hence, a SR \( f(x, y) \) will be detectable with reliability, conditioned by \( H \), when a threshold \( \tau \) is chosen such that

\[
\rho_0(f, H) < \tau \leq \tau_0(f).
\]

Now, a margin \( \mu_0 \) can be defined for a SR under the added discriminating conditions:

\[
\mu_0(f, H) = \tau_0(f) - \rho_0(f, H).
\]

Then \( \mu_0 \) measures the interval length, within the linear dependence scale of the NC, where a threshold \( \tau \) can be chosen according to Eq. (14). If the target in the scene is degraded, \( \tau_0 \) will decrease, and the margin \( \mu_0 \) will be reduced. Thus, \( \mu_0 \) expresses how much a target in an arbitrary position can be degraded until Eq. (14) is violated and detection becomes unreliable (Fig. 3).

This allows us to redefine the minimum detectable size and advantageous shapes, for a conditioning scheme, considering a given margin \( \mu \) instead of \( \tau \) and subtracting the lower bound \( \rho_0 \) from each \( \tau_0 \) term in Eqs. (11) and (12). Thus, \( M \) is now the minimum detectable size of \( f \), given a margin \( \mu \), if

\[
M = \min\{m; \ \mu < \mu_0(f^m, H)\},
\]

and \( f_a \) is advantageously shaped with respect to \( f_b \) when

\[
\mu_0(f_b, H) < \mu_0(f_a, H).
\]

To summarize, prior knowledge of the shapes whose false detection has to be avoided, added to the foregoing analysis about the effects of the spatially discretized similarity evaluation, allows the settling of concise upper and lower bounds to decide a safe threshold. The margin between the two bounds is then a good qualifier of the detection reliability for a given SR.

### 7 Involving a Representative Conditioning Set of Shapes to be Discriminated

When we have to deal with small SRs, their associated support region is small too; thus the portion of the scene that is extracted at each point to be compared with the model is also small.

When the image contains large and medium-size objects and structures, which are common in typical scenes, the contents of those small portions may be roughly classified in two types: regions of nearly uniform brightness, and regions containing a part of an edge or a corner. The first type has very low variance, and evaluating a similarity function in such conditions is not meaningful; but regions of the second type can pose a multiple false-detection problem if the threshold has not been properly chosen.

Figure 7 depicts the conditioning set \( H \) of shapes that have been chosen as representative prototypes of fractional edges and corners that are expected to be found locally in a typical scene. Only inclinations every 45 deg and corners of 90 deg are represented, but they can serve efficiently to estimate the expected partial similarity values.

A new set of curves, conditioned by \( H \) and displaying margin \( \mu_0 \) as computed in Eq. (15), can be seen in Fig. 8. It is interesting to compare these curves with those previously obtained for \( \tau_0 \) in Fig. 5. In general terms, the curves are about 25% lower than their previous values because the average value of \( \rho_0 \) for all SRs is about 0.25, but now some of them have their level dramatically decreased because of their high degree of similarity with some element of \( H \), as is the case of SQUARE, banished to the last position. On the other hand, TILTED-SQUARE significantly increases its relative advantage with respect to the other SRs.

Minimum sizes \( M \) obtained from Eq. (16) are displayed in Fig. 9, for a given margin \( \mu = 0.25 \). Having this margin to take into account target degradations in the scene, TILTED-SQUARE can still be safely detected at a size as low as \( 5 \times 5 \) pixels. SQUARE needs, in contrast, a \( 12 \times 12 \) pixel size. The other SRs show small variations of 1 pixel with respect to their previously computed sizes in the non-conditioned detection scheme.

### 8 Conclusions

Spatially discretized evaluation of the similarity measure, when the image is acquired with a pixelated device in a template-matching scheme, is not usually taken into account, because its effects are negligible when targets are large compared to the sensor grid spacing period. But when targets are small, a significant lowering of the normalized correlation peak values appears for targets displaced to an arbitrary position, even if they have not been degraded in the scene. This imposes an upper bound on the detecting threshold for a SR.
On the other side, false alarms may arise from the presence of common small-scale structures in the background, such as edges and corners, depending on the degree of similarity between them and each SR. From this fact, a lower bound for a threshold is established. The difference between upper and lower bounds is the remaining margin for dealing with target degradations, and can be efficiently used as a figure of merit for a given SR.
If a shape has to be chosen for a synthetic reference, experiments show that a 45-deg-tilted square appears to be, among the studied shapes, the most advantageous for reliable detection when the similarity measure is the normalized correlation. Using this SR, we can safely detect sizes as small as 5 × 5 pixels, while having a 25% margin left to deal with degradations such as noise, orientation and scale variations, and other shape departures.

The detection reliability characteristic of the SRs shows a very strong dependence on their shape, and among those in the studied set, there can be found differences in the minimum detectable size in a ratio as high as 2.4:1. Hence, when a small SR has to be designed, the decision about its shape must take this behavior into account; otherwise, reliable detection cannot be guaranteed.

Future work ought to be aimed in several directions: First, a geometric interpretation of the observed behavior would help to predict detection reliability for an arbitrarily shaped SR. Second, for practical applications it would be necessary to take into account the influence of a real optical system, especially the diffractive characteristic of the aperture stop. Similarly, the global behavior of the proposed schema under noise conditions should be analyzed.

References


Xavier Fernández received his degree in telecommunications engineering in 1980 from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain. Since 1984 he has been with the Engineering Department of Industrias de Optica, S.A. (INDO), Barcelona, conducting industrial projects on machine vision and advanced automation for ophthalmic optics production. His main research area is the development of techniques for accurate centering and alignment of ophthalmic optical surfaces throughout the manufacturing process. Currently he is pursuing a PhD degree at the Automatic Control Department, UPC, in the field of optimal design of small-size fiducial marks. He is a member of SPIE.

Josep Amat received his degree in electrical engineering in 1967 from the School of Engineering of Barcelona, and obtained his PhD in 1977 from the Universitat Politècnica de Catalunya (UPC). In 1968 he joined that school to teach digital systems. Now, he is a professor of computer technology in the Automatic Control Department of the UPC. His research work initially was in the field of digital signal processing, but since 1980 has been oriented to the computer vision field, for robotics applications, mainly in the development of high-speed image processors. He is the author of many publications and has participated in different national and international conferences and symposia. He has also participated in several projects related to perception, multisensory fusion, and underwater robots.