Resolution limits to object tracking with subpixel accuracy

David Mas,[1,*] Belen Ferrer,[2] John T. Sheridan,[3] and Julian Espinosa[1]

[1]*Instituto Universitario de Fisica Aplicada a las Ciencias y las Tecnologias, Universidad de Alicante, P.O. Box 99, Alicante 03080, Spain*

[2]*Departamento de Ingenieria de la Construccion, Universidad de Alicante, P.O. Box 99, Alicante 03080, Spain*

[3]*School of Electrical, Electronic and Communications Engineering, Communications and Optoelectronic Research Centre, SFI-Strategic Research Cluster in Solar Energy Conversion, College of Engineering and Architecture, University College Dublin, Belfield, Dublin 4, Ireland*
Corresponding author: david.mas@ua.es

Subpixel methods increase the accuracy and efficiency of image detectors, processing units, and algorithms and provide very cost-effective systems for object tracking. Published methods achieve resolution increases up to three orders of magnitude. In this Letter, we demonstrate that this limit can be theoretically improved by several orders of magnitude, permitting micropixel and submicropixel accuracies. The necessary condition for movement detection is that one single pixel changes its status. We show that an appropriate target design increases the probability of a pixel change for arbitrarily small shifts, thus increasing the detection accuracy of a tracking system. The proposal does not impose severe restriction on the target nor on the sensor, thus allowing easy experimental implementation.   © 2012 Optical Society of America

OCIS codes: 100.4999, 100.6640, 040.1490.

Object tracking in video sequences basically consists of looking for differences between two consecutive frames. Independent of the scene complexity, the detection of one single pixel difference can be enough to determine the movement of a target. Therefore, information about changes results in very sparse matrices [1]. Since the majority of the pixels are empty, they do not contribute to the object tracking, resulting in a waste of both data and the available system spatial-bandwidth product. Thus it is reasonable to consider rearranging the information gathered in order to optimize performance.

In [2] the authors discuss the design of target shapes for precise movement detection. Unfortunately, optimal shapes require precise knowledge of target and detecting array geometries, thus impeding realistic experimental implementation. Finally, the authors propose to use targets with known shapes, which turn out to be suboptimal but of easier implementation. The use of geometrical objects permits smart interpolation in the spatial domain, and thus subpixel accuracy can be achieved. However, we could not find in the literature resolution enhancements of greater than three orders of magnitude [3–5].

In this Letter we show that the accuracy of object tracking systems can be theoretically enhanced by several orders of magnitude. Using numerical simulations we demonstrate that micropixel resolutions can be achieved, and that further improvements are possible provided that the number of pixels in the sensor used is large enough.

Consider a detector array [CCD or complementary metal–oxide–semiconductor (CMOS)] in an imaging system with $N \times N$ pixels. Each pixel center is defined by the coordinates (i,j), with $i,j = 1 \ldots N$. For simplicity, we consider that the detection is binary, i.e., the pixel is only able to give an "on/off" response, depending on whether it detects light or not. The exact position of one single image spot on the camera is uncertain, since based on the detected signal all that is known is that it falls inside one pixel active area, i.e., the uncertainty region is $[i - 1/2, i + 1/2; j - 1/2, j + 1/2]$. Movement of the spot within this pixel area is undetectable.

Let us now suppose that the individual pixel area is subdivided up into $M \times M$ rows and columns, M being a natural number, and that the image spot is located on the last (right-hand) column of the pixel (i,j). A horizontal shift of the spot $1/M$ to the right provokes a jump of the spot to the $(i,j + 1)$ pixel in the adjacent column, and therefore the subpixel movement is registered. Detection of an additional equal displacement in the same direction is not possible, since the spot remains incident on the same already excited pixel. In order to register any additional displacement, the state of a different pixel, (i',j'), and its neighbor, $(i',j' + 1)$, must change. This reasoning may be repeated until the whole scene is moved one pixel to the right and the process starts again. Consequently, the detection of movements in the horizontal direction, with an accuracy of $1/M$ of one pixel, requires that different subsections of the object scene are imaged onto subarrays of M pixels. Similarly, vertical subpixel movements can be tracked by considering changes along columns of adjacent pixels. Following this reasoning, what is needed is to construct a target that maximizes the probability of pixel change detection for arbitrarily small shifts.

Let us consider again a two-dimensional array of $N \times N$ pixels, and let us assume a target formed by a cloud of dots that can be imaged to a different spot in the image plane. We define two independent sets of M random positive real numbers with values uniformly distributed in the range $[0, N - 1]$ and limit the arithmetic precision of these values to ε. The integer part of the numbers in each of the two sets represents the x and y coordinates, respectively, of each dot in the target. The origin, (0,0), is chosen to be the upper-left element of the array. Given that the precision has been set to ε, each coordinate unit can be subdivided into $1/\varepsilon$ elements. Thus, the decimal part of the numbers in the sets represents the location of the image spot within each individual pixel. For example,

setting $\varepsilon = 0.25$ implies a subpixel array of 4×4 elements. The pair (3.25, 5.75) represents a dot located at the position (2,4) inside pixel (3,5). Defining the object in this way ensures that, statistically, the spots in the image cloud are located at different locations in different pixels. A graphical representation of a target, made up of dots generated in this way, is depicted in Fig. 1.

If the target now shifts a distance ε to the right, all the dots will be displaced within the subelement structure. Eventually an image spot will move from one pixel to another, provoking a change in at least one pixel state, and the movement is detected (Media 1). Subsequent small displacements may each produce such state changes at randomly distributed pixels. This process is repeated until the entire initial image frame is reproduced shifted one pixel to the right.

Unambiguous movement detection requires independent single pixel state changes for each shift of ε. In an attempt to satisfy this requirement, we choose to have every excited pixel surrounded by unexcited pixels. Applying this rule of thumb, one in every two rows and columns must remain blank and the number of dots M in the target is limited to the 25% of the total array elements. Therefore, based on this value of the maximum number of available pixels, the theoretical tracking accuracy achievable is approximately $4/(N \times N)$ of one pixel. The value of the precision should then be fixed according to this limit, i.e., $\varepsilon \sim 4/(N \times N)$.

At this point, we would like to underline that this technique works on different principles from the one based on known-shape object detection [2–5], thus explaining why better resolutions are achieved.

In the case of known-shape object tracking the input target is continuous; however, in the image domain the field is sampled by the detector array, which imposes an effective band limit on the extent of the Fourier spectrum captured. After capture the object is reconstructed from its samples, and the lower the sample density, the poorer the estimation of the target shape and the more imprecise tracking becomes. Improvements can be achieved, since the target shape is known *a priori*. Thus, the missing information can be filled in following capture, overcoming the original sampling-rate-based Nyquist limit. Such use of *a priori* information can be interpreted as analogous to analytical spatial interpolation. This operation extends the band limits of the frequency spectrum captured, effectively increasing the
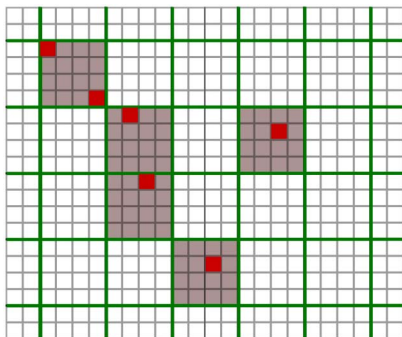
system space bandwidth product, permitting subpixel tracking accuracy, but this technique is limited by the interpolant shape.

In the method presented here, there is no predefined target shape, and subpixel accuracy is only obtained when tracking moving targets. The target is defined so that the dots in the target produce spots at the camera that are smaller than the pixel area. This defines an oversampled array that expands the Fourier limits. According to discrete Fourier theory, spatial interpolation enlarges the frequency bandwidth by adding zeros beyond the Nyquist limits. Nevertheless, no new information is added, so the extra bandwidth remains empty. However, in this case we use the temporal dimension (time varying information), and higher accuracy is obtained provided that there is a temporal change in the target. Detection is produced in the temporal domain, but the particular form of the target permits us to use the time variations to increase our information about the spatial shift. In this sense, the method performs a spatial multiplexing of the information, and the temporal variation is stored in the new frequency band so that a subpixel accuracy improvement is obtained.

The performance of our method has been tested using numerical simulations. For example, take a target consisting of M dots on a support matrix of 512×512 pixels (px) with $M < (512 \times 512)$. Two independent sets of coordinates consisting of M random positive real numbers uniformly distributed in the range [0,511] are generated. The precision is set to $\varepsilon = 1/(4M)$, which is greater than the expected accuracy, $1/M$. The resulting dots cloud is mapped onto the support matrix by rounding up the values toward the nearest equal or greater integer value. This simulates the effect of discretization imposed by a CCD or CMOS array. The original cloud set is then shifted in the x direction by in steps of size ε. After each shift, the target is discretized and pixel changes with respect to the previous step are noted. The process is repeated until a full frame displacement by one pixel is achieved (see Fig. 1).

For the first simulation, $M = 50000$ px and $\varepsilon = 1/4M = 5 \times 10^{-6}$. Due to the statistical nature of the process, not all displacements are detected and a range of threshold shifts between ε and $48 \times \varepsilon$, i.e., $[5 \times 10^{-6}, 2.4 \times 10^{-4}]$ px, are needed to produce a pixel change. In Fig. 2 the resulting histogram and the accumulated histogram of the obtained threshold values are presented.

From the figure, it is clear that that only 20% of movement detections (involving one pixel change) take place for the minimum displacements of 5×10^{-6} px (one ε unit). This happens because of the statistical method used to construct the target; however, this is not a fixed value or relationship, and the same change often requires larger displacements to be generated. The accumulated histogram represents the probability of detecting a pixel change with target displacements equal to or less than the value indicated on the horizontal axis. One can see that 80% of displacements equal to or less than 3.5×10^{-5} px ($\sim 2/M$) are correctly detected. This value is only a little worse than the expected accuracy of $1/M$. Since the target is generated randomly, there is no systematic control of dot positioning. Therefore,

Fig. 1. (Color online) Detection of a horizontal displacement with accuracy 0.25 px with the random cloud target (Media 1).
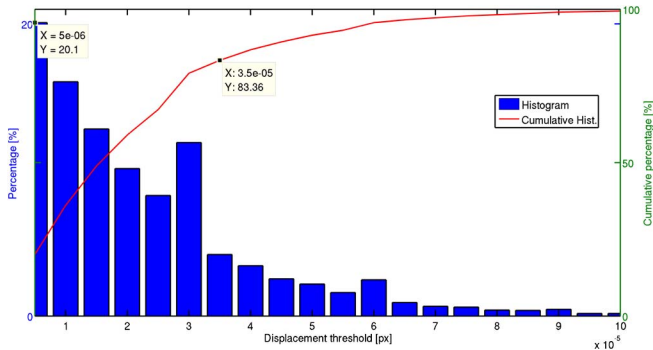
Fig. 2.   (Color online) Histogram and cumulative histogram of threshold shifts needed to detect an incremental displacement in the target with $M = 50,000$ dots.
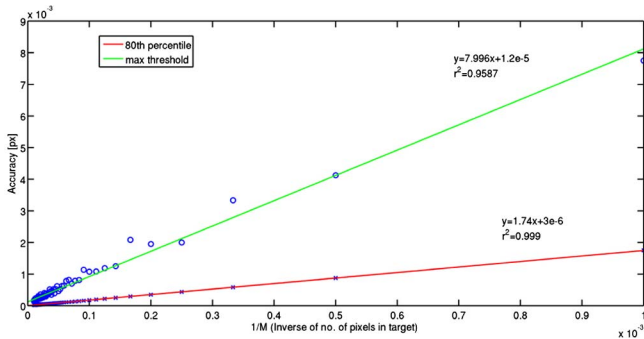

Fig. 3.   (Color online) Detection accuracy of the sparse target method according to the 80th percentile and the maximum threshold criteria for a sparse target size M from 10^3 to 10^5 px, defined on support matrices of 512×512 px.

multiple spots, localized within a single pixel area, may be generated, and this will slightly diminish the accuracy, explaining the reduction in accuracy observed.

Based on these results, two criteria can be used in order to establish the detection accuracy of the method. First, the accuracy can be determined by taking the most conservative criterion, which consists of taking the worst case, i.e., the maximum displacement needed to always produce a pixel change, i.e., 2.4×10^{-4} px or above. The second criterion ensures that the displacement produces an observation with 80% probability, which means detection of the vast majority of displacements, i.e., 3.5×10^{-5} px. The choice of criterion depends on the expected trajectory of the object to be tracked. If it is deterministic, e.g., vibration assessment [4], the statistical criterion is sufficient, as errors can be compensated by fitting the results to the expected path. If, on the other hand, the trajectory is completely unknown or is known to be erratic [6], the most conservative error should be used.

Note that the accuracy of the method depends on the number of dots in the target, M, but not on the size of the support array, which only limits the maximum accuracy (see Fig. 3). In this figure the results are presented for simulations performed for different targets with $10^3 < M < 10^5$ defined on a support matrix of 512×512 px. Following our procedure above, displacements in steps of $\varepsilon = 1/(4M)$ are made, until a full image frame displacement of one pixel takes place and the accuracy achieved when applying both criteria is presented.

In the figure, least square fit lines to the simulated data are presented. The correlation value, i.e., $r^2 = 0.99$, between line fit and the simulated results is especially good in the case of the 80th percentile criterion. For the other criterion the simulated results appear noisier; however, even in this case the results can still be well fit by a straight line, i.e., $r^2 = 0.9587$. Such results indicate that following proper design of a target can take account of the size of the sensor array or, inversely, involve selecting an adequate sensor to achieve the required accuracy.

Real implementation of the method will place additional constraints on the accuracy achievable. Noise will introduce distortions on the target, thus decreasing the accuracy. According to our own simulations, the method can hold 5% of spurious dots in the target with an accuracy loss of one order of magnitude. Above this noise the resolution is rapidly degraded, although this may depend on the processing algorithm.

Nevertheless, the method enhances subpixel resolution in a controlled way and, under ideal conditions, micropixel or even higher accuracies can be obtained, thus producing very cost-effective tracking systems.

References

1. Y. Kashter, O. Levi, and A. Stern, Appl. Opt. **51**, 2491 (2012).
2. A. M. Bruckstein, L. O'Gorman, and A. Orlitsky, IEEE Trans. Inf. Theory **44**, 3156 (1998).
3. S. H. Lee, M. Baday, M. Tjioe, P. D. Simonson, R. Zhang, E. Cai, and P. R. Selvin, Opt. Express **20**, 12178 (2012).
4. D. Mas, J. Espinosa, A. B. Roig, B. Ferrer, J. Perez, and C. Illueca, Appl. Opt. **51**, 2664 (2012).
5. R. Kurita and E. R. Weeks, Phys. Rev. E **82**, 011403 (2010).
6. J. P. Ryle, M. Al-Kalbani, N. Collins, U. Gopinathan, G. Boyle, D. Coakley, and J. T. Sheridan, J. Biomed. Opt. **14**, 014021 (2009).