Binary Optical Filters for Scale Invariant Pattern Recognition

Max B. Reid, John D. Downie, and Butler P. Hine

April 1992
Binary Optical Filters for Scale Invariant Pattern Recognition

Max B. Reid, John D. Downie, and Butler P. Hine, Ames Research Center, Moffett Field, California

April 1992
Summary

Binary synthetic discriminant function (BSDF) optical filters which are invariant to scale changes in the target object of more than 50% are demonstrated in simulation and experiment. Efficient databases of scale invariant BSDF filters can be designed which discriminate between two very similar objects at any view scaled over a factor of 2 or more. The BSDF technique has considerable advantages over other methods for achieving scale invariant object recognition, as it also allows determination of the object’s scale. In addition to scale, the technique can be used to design recognition systems invariant to other geometric distortions.

Introduction

Optical pattern recognition involves the use of optical correlation to distinguish one or more spatial patterns, or images, from another set of patterns. Optical correlation is being investigated at Ames Research Center as a basis for autonomous vision systems (ref. 1). Much of this research concerns the development of new types of optical filters for use in optical correlation, including binary synthetic discriminant function (BSDF) filters.

BSDF filters have previously been proposed (refs. 2 and 3) and demonstrated (refs. 4 and 5) which were designed to be invariant to either in-plane or out-of-plane rotations of a target object. General complex-valued synthetic-discriminant function filters cannot be encoded on commercially available binary spatial light modulators (SLMs). The BSDF technique includes the modulation characteristics of the binary SLM in the filter synthesis equations in order to overcome this limitation.

The filter design procedure begins with a set of centered training images, \( t_n(x,y), n = 0, 1, \ldots, k \), spanning the desired distortion invariant feature range. This image set is used to construct the synthetic function, \( s'(x,y) \), for a given filter modulation. The desired peak correlation response of \( s'(x,y) \) is a constant, \( c_n \), for each training image \( t_n(x,y) \):

\[
\int \int t_n(x,y)s'^*(x,y) \, dx \, dy = c_n
\]  

where * is the complex conjugate operator and the integral is taken over the area of the input field. The function \( s'(x,y) \) includes the filter modulation, \( M \), through the equation

\[
s'(x,y) = F^{-1}M[F[s(x,y)]^*]
\]  

where \( F \) is the fourier transform operator. The purpose of the filter generation procedure is to determine a function \( s(x,y) \) which solves equation (1) given a particular modulation function, \( M \). The function \( s(x,y) \) is chosen to be a linear combination of the training images as

\[
s(x,y) = \sum_{n=0}^{k} a_n t_n(x,y)
\]  

A general synthesis equation results from substituting equations (3) and (2) into equation (1):

\[
\left< t_n(x,y) \left| F^{-1}M \left[ \sum_{m=0}^{k} a_m^* t_m(x,y) \right] \right> = c_n
\]  

For binary phase-only filters, BPOFs, the modulation function is of the form

\[
M[S(u, v)] = \begin{cases} 
1, & \text{Re}[S(u, v)] \geq 0 \\
-1, & \text{Re}[S(u, v)] < 0
\end{cases}
\]

where \( S(u,v) \) is a two dimensional complex function, and equation (4) becomes a system of nonlinear equations which may be solved using an iterative procedure based on the Newton-Raphson algorithm. The filter coefficients, \( \bar{a} \), are constrained to be real, are initialized to the desired response vector, \( \bar{c} \), and iterated based on the formula (ref. 2)

\[
a_n^{i+1} = a_n^i + \beta \left[ c_n - c_0 \left( \frac{m_n}{m_0} \right) \right]
\]

where \( i \) is the iteration number, \( \beta \) is a damping constant, and \( m_n^i \) is the modulus of the peak correlation response of image \( t_n(x,y) \) with the filter constructed with \( \bar{a}^i \).

Scale Invariant BSDF Design

Here we apply the BSDF technique to training sets consisting of scaled views of a target object. The target used was the 128 x 128 pixel binary silhouette of the box-end wrench shown at 100% scale in figure 1(a). A set of 51 training/test images was made from the target object, where the scale of the images varied from 50% to 100% of the original at 1% intervals. The smallest member of the test set is shown in figure 1(b).
Filters were designed to produce nearly equal peaks for a specified range of sizes of box-end wrenches. After synthesis, the filters were correlated with all 51 scaled views of the box-end wrench to test their selectivity to scale. They were then correlated with similar sets of 51 scaled views of both an open-end wrench and a visegrip. The 100% and 50% scaled images of the out-of-class objects are shown in figure 1(c)-(f). All three tool images were scaled so that the 100% scale views contained equal mean light levels. The open-end wrench was chosen to test the capability of scale-invariant BSDFs to discriminate between very similar objects. The visegrip provides a test of discrimination against less similar objects.

In simulation, filters were made for six different scale distortion ranges with ratios of maximum-image size to minimum-image size from 1.06 to 2.0 (see table 1). That is, the end points of each training set were chosen so that the scale of tmax divided by the scale of tmin equaled the desired distortion range, such that the mean of their scales equaled the 71% image, the geometric mean of the entire test set.

Two different filters were made for each distortion range. The first set was made with training images spaced so that each image was ~5% smaller than the next largest image and ~5% larger than the next smallest. For instance, for a scale range of 1.25, 5 training images were used which were 63%, 67%, 71%, 75%, and 79% of the size of the original target shown in figure 1(a). For the second set of filters (shown in table 1), every image within the distortion range was used in the training set. The 1.25 scale filter in this case used 17 training images, i.e., all of the test images between 63% and 79%.

The number of training images used for both sets of filters is given in table 1. The learning parameter used was $\beta = 0.6$, and convergence was defined as the point where the correlation peak intensity for each of the training images varied by no more than $\pm 4\%$ from the average of the n training image correlation intensities. The first set of filters is made using training images spaced at $=5\%$ relative scale changes. The second set of filters was made using every test image in the desired distortion invariant scale interval. To achieve stability and ensure the procedure would terminate, a value of $\beta = 0.2$ was used in equation (6), and convergence was defined as the point where each training image correlation intensity varied by no more than $\pm 10\%$ from the average. The table also gives

<table>
<thead>
<tr>
<th>Scale invariant distortion range</th>
<th>Training images at 5% intervals</th>
<th>All test images used to train</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of training images</td>
<td>Number of training images</td>
</tr>
<tr>
<td></td>
<td>Number of iterations</td>
<td>Number of iterations</td>
</tr>
<tr>
<td>1.06</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>1.12</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>1.25</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>1.50</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>1.75</td>
<td>11</td>
<td>41</td>
</tr>
<tr>
<td>2.00</td>
<td>13</td>
<td>51</td>
</tr>
</tbody>
</table>

Table 1. BSDFs made for six different distortion ranges using two different methods for choosing the training set.

Figure 1. Target object for scale invariant filter design: (a) box-end wrench at 100% scale, (b) box-end wrench at 50% scale. Out-of-class objects for testing BSDF filter discrimination capability: (c) open-end wrench at 100% scale, (d) open-end wrench at 50% scale, (e) visegrip at 100% scale, (f) visegrip at 50% scale. All full size objects are scaled to contain the same number of pixels with value "1."
the number of iterations in a computer simulation that were required before the procedure converged, defined by equations (4) and (6). For the images spaced at 5% relative scale changes, convergence was defined as the point where the correlation peak intensities of the training images each varied by no more than ±4% from the average of the number of intensities (n). A value of β = 0.6 was used in equation (6), and in table 1, convergence was very rapid as is evident.

More iterations were required for convergence when the training images were spaced more closely together, as given in column 4 of table 1. In this case, the cross-correlation matrices of the training images are less orthogonal, and a smaller learning constant of β = 0.2 was required in equation (6) in order to ensure stability in the iterative algorithm. It was also necessary to use a relaxed definition of convergence in order for the procedure to terminate. Convergence was defined, in this case, as the point where the correlation peak intensities varied by no more than ±10% from the average value.

Simulation Results

Both sets of BSDFs were tested against all 51 scaled views of the box-end wrench or target object. For comparison, the correlation response of a binarized matched spatial filter (BMSF) made to recognize the 71% scale box-end wrench is in figure 2 as a function of input image scale size. The figure shows the peak intensity of each cross-correlation normalized to the intensity of the auto-correlation peak. The peak intensity of the correlation of different scale box-end wrenches with the BMSF are plotted normalized to the peak of the 71% scale-image correlation. Scale is given as a percentage of the size of the image in figure 1(a). Such a binary filter made from only one scale of the image is extremely sensitive to scale distortions. The peak intensity drops by approximately a factor of 2 for a scale change of only ±1.5% (from the 71% view to the 70% or 72% views).

This extreme sensitivity forces the use of closely spaced training images to produce scale invariant BSDFs. The correlation response of a BSDF designed to be invariant over a scale range of 1.06 using just two training images is shown in figure 3. The peak intensity of the correlation of different scale box-end wrenches with the BSDF are plotted normalized to the correlation peak intensity of the 71% scale image with its BMSF, as in figure 2. The figure shows that while the correlation peaks for the training images are nearly equal, peaks for test images in-between training images drop substantially. The high peaks for training images are wasted because the lower peaks between limit the discrimination capability of the filters.

For binary images, which almost fill a 128 × 128 pixel window, scale resolution is limited to about 1% of the size.
of the original image. A way to avoid the problem of lower peaks for test images between training images is to make filters that use every scaled view that can be drawn in the given window. Filter synthesis takes longer in this case, but not an unreasonable length of time for a $128 \times 128$ window as given in table 1.

Typical responses of scale invariant filters made using all test images in the given distortion range as training images are shown in figures 4 through 6. Comparing the 1.06 scale factor filters in figures 4 and 3, the correlation peaks for all 5 in-range images are now seen in figure 4 to be very nearly equal. The highest peaks have dropped, but the lower peaks, which limit discrimination capability, are considerably higher. Figure 4 also shows a very sharp drop-off for sizes of box-end wrench outside the desired invariant range. A similar scale response is seen in figures 5 and 6 for the scale ranges of 1.12 and 1.50.

The results for all six BSDFs made with all in-range test images as training images are summarized in figure 7. The curve connecting the open squares plots the lowest peak resulting from the correlation of each filter with all test images which lie in the specified distortion range. For example, the peak value of 0.57 for the 72% image in figure 4 is plotted as the in-class result for the 1.06-scale factor filter in figure 7. The data plotted for a scale range of 1.0 (no scale invariance), are the results for the BMSF.

Figure 5. Normalized simulation response of a BSDF filter designed to be invariant over a 1.12-scale factor range. The filter was created using all nine test images between 67% and 75% as training images.

Figure 6. Normalized simulation response of a BSDF filter designed to be invariant over a 1.5 scale factor range. The filter was created using all 30 test images between 58% and 87% as training images.

Figure 7 also shows the worst case response of the two out-of-class objects when correlated with each filter. The open circles denote the highest peak resulting from correlating all 51 scaled views of the open-end wrench with the BSDFs. The diamonds show the highest peak for the vise-grip. All in-range box-end wrenches are discriminated.

Figure 4. Normalized simulation response of a BSDF filter designed to be invariant over a 1.06 scale factor range. The filter was created using all five of the test images between the 69% and 73% images as training images. The endpoints of the distortion range are delineated with dashed vertical lines.
Figure 7. BSDF discrimination capability for filters designed using all test images in the specified scale distortion invariant interval as training images. The lowest correlation peak intensity for any box-end wrench within the specified distortion range is plotted as boxes and normalized to the correlation peak of the 71% image with its BMSF. The BMSF is included in the plot as the filter with distortion range 1.0. The highest peak produced by correlating any view of the open-end wrench scaled from 50% to 100% with the given filters is also plotted as circles, as is the highest peak for any scale of the visegrip, plotted as diamonds.

From all open-end wrenches up to a scale distortion range of a factor of 1.25, and from all visegrips up to a scale range of 2.0. It is seen that the BSDF with the greatest scale invariance, which can discriminate against all views of the out-of-class objects, is the scale factor 1.25 filter. Thus, scale invariance of up to 25% is achievable while still allowing discrimination between two similar objects.

For the less restrictive case, discriminating against the visegrip only, a scale range of 100% can be achieved. These distortion invariant ranges are quite sufficient to allow very efficient filter databases designed to first recognize an object invariant to scale, then subsequently determine the exact scale of the input view, which is discussed in the conclusion.

Experimental Procedure

Because of the superior performance of the BSDFs which were designed to use every scale of test image in the training set, this design technique was chosen for experimental verification on a laboratory correlator. A diagram of the correlator is shown in figure 8. Two 128 x 128 pixel magneto-optic spatial light modulators (MOSLMs) produced by Semetex were used in the input and filter planes. Binary phase-only operation was obtained by placing the second Glan-Taylor calcite polarizer after the filter MOSLM perpendicular to the polarization defined by the input MOSLM polarizer. The polarizers were both placed as far downstream as possible in the optical train to minimize the effect of phase distortions caused by polarizer curvature (ref. 6). The polarizers were found to have substantial curvature and were not nearly as flat as implied by the Melles Griot catalog.

Lenses 1 and 2 were chosen to produce a focal length of \( f = 1168 \) mm to match the spatial frequency of the MOSLMs. The input and filter MOSLMs were placed at the front and back focal planes of the transform lens pair, respectively. A more compact system utilizing a phase correction lens at the focal plane (ref. 4) was not used because of the additional alignment errors inherent in such designs. The third lens was chosen to be \( f_3 = 1000 \) mm for convenience.

![Figure 8. Three lens correlator using magneto-optic spatial light modulators (MOSLMs) in the input and filter planes.](image)
The transmittance of the MOSLMs used is low (=5%), and in addition, a high percentage of transmitted light is diffracted into higher orders by the MOSLM pixel structure. By imaging through both light modulators, the intensity of light transmitted to the correlation plane is reduced from the input plane by approximately six orders of magnitude. This necessitated the use of a silicon intensifier target camera in the correlation plane, along with a 35 milliwatt helium-neon laser to provide sufficient input intensity.

The iterative filter construction was performed directly on the optical correlator under automatic computer control. Constructing BSDFs on the correlator has the advantages of using a continuous Fourier transform and of compensating for aberrations in the optical system. Both the input and filter MOSLMs and the frame grabber used to record the correlation output were controlled directly by the computer. The intensities of the actual optical correlation peaks were used in the iterative algorithm described by equations (4) and (6) in order to determine the BSDF coefficients.

The output of the frame-grabber/camera system varied by ±10% from frame to frame, as is normal for a thermally uncontrolled charge coupled diode (CCD) camera. As a result of this variation, an average of five measurements was taken of the correlation peak of each image with the trial filter during each iteration to avoid making random changes in BSDF coefficients that could preclude convergence. The same learning parameter of β = 0.2 was used as in simulation. Iteration was stopped when the peak correlation intensity for each image in the training set varied by no more than ±10% from the average of all peaks in the set. In general, the number of iterations required was very close to the values given in column 4 of table 1 for the simulated filters.

**Experimental Results**

After BSDFs were synthesized on the correlator for the six distortion invariance ranges, they were correlated with all scaled views of the box-end wrench, open-end wrench, and visegrip. The responses of the filters designed for invariance to scale ranges of factors of 1.06, 1.12, and 1.5 are shown in figures 9 through 11. The iterative filter synthesis procedure was performed using correlation peaks measured on the actual laboratory correlator. The peak intensity of correlations of different scale box-end wrenches with the BSDF as measured on the correlator are plotted normalized to the correlation peak intensity of the 71% scale image with its BMSF. The plots are seen to agree very well with the simulated results of figures 4 through 6. The difference is a greater variability in the experimental results, caused by the variance in experimental correlation peak measurements, as discussed above.
Greater variance in correlation peaks of test views results in a lower minimum correlation peak for the in-class objects. This causes some reduction in the maximum scale invariant range which can be covered with a single filter while retaining the ability to discriminate against out-of-class objects. The results for the experimental filters are summarized in figure 12. The curve for the in-class box-end wrenches is somewhat lower than in figure 7 for simulation. The greatest scale range over which a BSDF can discriminate against all scaled views of the open-end wrench is reduced to a factor of 1.12. For the less similar object, the experimental filters still distinguish between all in-range box-end wrenches and all scaled views of the visegrip out to a scale invariant range of a factor of 2. Scaled box-end wrenches are discriminated from visegrips with a significant margin for error up to a scale range of 1.5. There is little margin for error in the results for the 1.75 and 2.0 scale range filters, however, which leads us to prefer the more conservative conclusion of invariance to changes in scale up to 50%.

Conclusions

We have demonstrated that a single binary phase-only filter can be designed as a binary synthetic discriminant function filter capable of producing nearly constant correlation peaks for all scaled views of a target object within a specified distortion range. In simulation, filters were demonstrated invariant to changes in scale up to a factor of 1.25 which could discriminate between images of two very similar types of wrench. The scale invariant range could be extended to a factor of two if the out-of-class object was a less similar image of a visegrip. Experimental filter results confirmed the simulations, though measurement variance reduced the scale invariant distortion range which could confidently be claimed to a factor of 1.12 for the very similar open-end wrench and 1.5 for the visegrip.

Binary SDFs have definite advantages over competing methods for achieving scale invariance in optical correlator systems. They may be directly implemented on rapidly updatable SLMs, as used in correlators under development by numerous groups. No coordinate transformations are required for the input images, as with the Mellin transform (ref. 7). Other approaches for achieving full scale invariance are wedges in wedge-ring detectors (ref. 8), and scale invariant moment detection (ref. 9), but potentially significant information is lost with these techniques. A database of BSDFs can be designed which not only performs rapid discrimination invariant to scale changes, but also subsequently specifies the scale of the input view to high precision. Because the scale invariant range of
each filter may be specified, a hierarchical tree of filters can be synthesized, as previously demonstrated for rotation distortions (ref. 5).

Figure 13 displays the kind of hierarchical filter database which may be designed for scale invariant recognition/discrimination and subsequent scale determination. The figure shows part of a filter database which allows discrimination between a box-end wrench and a vise-grip. Here, either tool may be scaled over a factor of 2. The wrench can be identified by evaluating the correlations of a test input with just the first two filters at the highest level of the database. Subsequently, only 12 further filters need be examined to determine the image’s scale to a precision of 1%. With a linear database of BMSFs (ref. 10) approximately 50 filters would require evaluation to achieve the same precision.

The total number of filters with which the input must be correlated is not greatly increased if the in- and out-of-class objects are very similar. Figure 14 shows part of a filter database which recognizes a box-end wrench over a scale factor of 2, is capable of distinguishing from scaled open-end wrenches, and can determine the scale of the box-end wrench to 1% precision after sequencing through a total of 15 filters. The highest level filters are invariant over a much smaller scale range than in figure 13, but the length of the database search is increased by only one filter. In both cases, the speedup over a linear database of BMSFs is about a factor of 3, and would be even greater for databases covering larger distortion ranges.

Figure 13. Portion of a hierarchical database of scale invariant filters which could be used to discriminate between a target and an out-of-class object, such as a box-end wrench and a vise-grip, either of which could be at any scale over a range of a factor of 2. Only two filters must be evaluated to perform discrimination at the highest level of the hierarchy. Sequencing through twelve further filters allows specifying the scale of the target object to 1% precision.

Figure 14. Portion of a hierarchical database of scale invariant filters to discriminate between a target and a very similar out-of-class object, such as a box-end wrench and a open-end wrench, either of which could be at any scale over a range of a factor of 2.
Hierarchical databases of BSDFs provide both the rapid invariant recognition achievable with wedge-ring detection and the highly precise scale determination possible with large databases of binarized matched spatial filters. Further, the BSDF technique can be extended to produce filters from training sets of images subjected to multiple distortions, including rotations in- and out-of-plane, providing the basis for extremely powerful autonomous vision systems simultaneously invariant to all in- and out-of-plane geometric distortions.

References


Binary optical filters for scale invariant pattern recognition are demonstrated, allowing efficient databases of filters to discriminate between two very similar objects scaled over a factor of 2 or more. This technique offers advantages over other methods for scale invariant object recognition. Additionally, the technique can be used to design recognition systems invariant to other geometric distortions.