IMAGE SCALE AND ROTATION FROM THE PHASE-ONLY BISPECTRUM

Janne Heikkilä

Machine Vision Group, Infotech Oulu, Dept. of Electrical and Information Engineering
P.O. Box 4500, FIN-90014, University of Oulu, Finland
e-mail: jth@ee.oulu.fi

ABSTRACT

This paper deals with the problem of aligning two images under translation, rotation and scaling. The method described utilizes the shift invariance property of the bispectrum to eliminate the effect of the translation component. Only the phase information is preserved from the bispectrum in order to achieve better resilience against nonuniform illumination changes. The scale and the rotation parameters are estimated from the remaining log-polar sampled spectrum using cross-correlation. The examples shown in the paper indicate that the method is quite robust against background clutter and occlusions.

1. INTRODUCTION

There are basically two kinds of approaches available for estimating the scale, rotation and translation parameters between two images: feature based and featureless solutions. Feature based solutions utilize some salient features such as corners, edges, contours or some special interest points to determine the geometric mapping between the images. Pose clustering [1] is a typical example within this category that uses a Hough transform like voting scheme to extract the most likely set of correspondences. The main problem with the feature based approach is that reliable features are difficult to extract, and certain objects do not have good discrete features at all.

In the featureless solutions the parameters are estimated directly from the image data. A well-known method is to use geometric moments for extracting this information, but the major shortage related to the moments is that the objects should be segmented first from the background, and this is not often possible. Another well-known method is the Fourier-Mellin transform [2] that is based on the properties of the image amplitude spectrum. The advantage of the Fourier-Mellin transform is that no segmentation is required.

In this paper, we consider an alternative solution for the Fourier-Mellin transform that is based on the bispectrum of the image. Bispectrum has been used earlier for image reconstruction e.g. in [3] and for pattern recognition in [4], but here we choose another course of direction and develop a method for estimating the image scale factor and rotation angle from it.

2. FOURIER-MELLIN TRANSFORM

In the Fourier-Mellin transform, the amplitude spectrum of the image (magnitude of the Fourier transform) is used to obtain a shift-invariant representation which reduces the number of the dimensions in the parameter space from four to two. The remaining scale and rotation parameters can be derived by utilizing the following properties of the Fourier transform:

$$f(ax, by) \quad \Leftrightarrow \quad \frac{1}{|ab|}F(\frac{u}{a}, \frac{v}{b}) \tag{1}$$

$$f(r, \theta + \alpha) \quad \Leftrightarrow \quad F(\omega, \phi + \alpha) \tag{2}$$

where $f(x, y)$ is a gray scale image function defined over a 2-D Euclidean space, $F(u, v)$ is the 2-D Fourier transform of $f(x, y)$, $f(r, \theta)$ and $F(\omega, \phi)$ are the corresponding representations in the polar coordinates, and a, b and α are some constants. The property (1) makes it possible to estimate the scale differences directly from the displacement of the two logarithmically sampled amplitude spectrums. The property (2) converts the problem of estimating the rotation angle into determining the displacement along the ϕ-axis in the polar coordinates. As a result, both the scale and the orientation can be solved simultaneously simply by determining the shift between two log-polar sampled amplitude spectrums.

There are at least three disadvantages related to the amplitude spectrum. Firstly, it does not provide a full description of the image contents, because it lacks the information carried by the phase spectrum. The amplitude spectrum is symmetric which means that the half of its values are redundant. As a consequence, half of the information is basically lost. Secondly, the amplitude spectrum is sensitive to illumination changes that is not a good basis for illumination invariant matching. Thirdly, it is necessary to perform prewindowing before computing the amplitude spectrum in order to avoid the sinc effect in the resulting spectrum.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

On the other hand, it is well known that phase spectrum contains illumination invariant spatial information about the image contents and it has been successfully used in template matching with so called *phase-correlation* or *phase-only matched filtering*. Interestingly, also in the practical implementation of the Fourier-Mellin transform presented in [5] the phase-only matched filtering is used to derive the scale and the rotation parameters from the log-polar sampled amplitude spectrums.

3. PHASE-ONLY BISPECTRUM

Due to the problems related to the amplitude spectrum it becomes reasonable to consider if there are some other shift-invariant representations that can be used instead. In practice, this representation should also have properties similar to (1) and (2) in order to provide a simple means for estimating the scale and rotation parameters.

Fortunately, the amplitude spectrum is not the only spectral representation that is shift-invariant. Also, certain higher order spectrums have the same property. These spectrums are defined by

$$\Psi_n(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = F^*(\mathbf{s}) \prod_{i=1}^{n} F(\mathbf{u}_i) \qquad (3)$$

where $\mathbf{u}_i$ with $i = 1, \ldots, n$ are vectors in the 2-D frequency space, and $\mathbf{s} = \mathbf{u}_1 + \mathbf{u}_2 + \ldots + \mathbf{u}_n$.

Let us now take a closer look at two special cases Ψ_1 and Ψ_2. For Ψ_1 we get

$$\Psi_1(\mathbf{u}) = F(\mathbf{u})F^*(\mathbf{u}), \qquad (4)$$

which is clearly the power spectrum, i.e., the squared amplitude spectrum of $f(x, y)$. If $n = 2$ the higher order spectrum becomes

$$\Psi_2(\mathbf{u}_1, \mathbf{u}_2) = F(\mathbf{u}_1)F(\mathbf{u}_2)F^*(\mathbf{u}_1 + \mathbf{u}_2), \qquad (5)$$

which is also called the *bispectrum*. For $n = 3$ we get trispectrum and so on. Henceforth, we will concentrate on the bispectrum, but it is straightforward to extend the following discussion to the other cases as well.

In contrast to the amplitude spectrum, the bispectrum has the inverse property, which means that the original image can be reconstructed from it just by adding the missing shift information. A solution for this reconstruction problem is represented for example in [3]. Except for the location, the bispectrum does not loose any essential information of the image contents, and it retains the phase information that is lost by the amplitude spectrum. This makes the bispectrum a much more appealing representation from the image analysis point of view.

As we can see from (5) the bispectrum has two vector arguments containing totally four scalar frequency variables. Assuming that $F(\mathbf{u})$ is an N-by-N discrete Fourier transform of $f(x, y)$, the bispectrum becomes a four-dimensional N-by-N-by-N-by-N matrix. It is therefore not practical to evaluate the whole bispectrum. A better solution is to take 2-D slices of the 4-D spectrum. There are basically various ways of defining these slices, but we shall only consider the case where

$$S_k(\mathbf{u}) = \Psi_2(\mathbf{u}, k\mathbf{u}) \qquad \forall k \in \mathcal{R}. \qquad (6)$$

Although we have now taken only a small portion of the whole spectrum, it can be shown that the reconstruction is still possible and no essential information has been lost [6].

We must next consider the scaling and the rotation properties of these bispectrum slices. From (1), (2), (5) and (6) it follows that

$$f(ax, by) \quad \overset{B}{\Rightarrow} \quad \frac{1}{|ab|^3} S_k(\frac{u}{a}, \frac{v}{b}) \qquad (7)$$

$$f(r, \theta + \alpha) \quad \overset{B}{\Rightarrow} \quad S_k(\omega, \phi + \alpha), \qquad (8)$$

where $\overset{B}{\Rightarrow}$ means the transformation from an image function to its bispectrum slice. These properties are basically the same as for the Fourier transform, which indicates that we can directly use the bispectrum slices in the same way as the amplitude spectrum in the Fourier-Mellin transform to estimate the scale and rotation. However, the bispectrum as well as the amplitude spectrum are sensitive to illumination changes. Especially, non-uniform illumination changes may cause significant problems in template matching or image alignment. Therefore, we decide to use only the phase information included in the bispectrum which is less sensitive to such phenomena. The phase information has also shown to be quite robust against background clutter and occlusions in phase-only matched filtering. This gives us a basis for defining the *phase-only bispectrum* slice as follows

$$P_k(\mathbf{u}) = \frac{S_k(\mathbf{u})}{|S_k(\mathbf{u})|}. \qquad (9)$$

This remaining phase information is then used for determining the scale and the rotation parameters.

4. MATCHING PROCEDURE

Let us assume that $P_k(\mathbf{u})$ and $P'_k(\mathbf{u})$ are phase-only bispectrum slices for two gray scale input images $f(x, y)$ and $f'(x, y)$, respectively. Both the images and the spectrums are assumed to be in a discrete form, and the spectrums are computed using the discrete Fourier transform (DFT). For simplicity, we further assume that both DFTs are N-by-N matrices. Thus, the frequency variable $\mathbf{u}$ is a two dimensional vector of integer numbers between [0, $N - 1$]. As it can be seen from (5), (6) and (9) we need samples from the DFT at points $\mathbf{u}$ but also at $k\mathbf{u}$ and $(k + 1)\mathbf{u}$ to compute the slices. The samples in the two latter cases can be extracted from the DFT by utilizing the conjugate symmetry and the periodicity properties of the DFT. Two slices with $k = 1$ and $k = 2$ are computed for both images to further improve the robustness of the method. Notice that it is not necessary to window the data before computing the DFT, because the amplitude information is not used at all.

Next, we resample the slices using log-polar sampling and bilinear interpolation. The sampling points are selected in the same way as in [5] and they are expressed as

$$\begin{aligned} u_i &= t^{r_i/t} \cos\theta_i \\ v_i &= t^{r_i/t} \sin\theta_i, \end{aligned} \tag{10}$$

where $t = N/2 - 1$, and (r_i, θ_i) are the points in the new uniform sampling lattice. In the experiments $r_i \in \{$ 0, 0.5, 1 , . . . , $t\}$, and $\theta_i \in \{0, 1^\circ, \ldots, 360^\circ\}$ were used. The spectrums obtained after bilinear interpolation are denoted by $B_k(r_i, \theta_i)$ and $B'_k(r_i, \theta_i)$ with $k = 1, 2$.

The following task is to find the displacement between the spectrums by utilizing 2-D cross-correlation so that along the r-axis the correlation is linear and along the θ-axis it is cyclic. This is performed independently between B_1 and B'_1, and B_2 and B'_2. The correlation result along the r-axis can be improved by windowing the spectrums radially in the same way as in the Fourier-Mellin transform. The Hanning window could be used for this purpose as suggested in [5]. The resulting cross-correlation functions are denoted by C_1 and C_2. In order to suppress the noise caused by, e.g., the background clutter and the occlusions, these two correlation functions are added together so that $C = C_1 + C_2$.

Now, the coordinates of the maximum value of C denoted by (r_{max}, θ_{max}) give the estimates for the scale and rotation parameters. The rotation is obtained directly from θ_{max} and the scale is derived using the nonlinear mapping $s = t^{-r_{max}/t}$. The correlation peak in C is often slightly blurred due to the bilinear interpolation performed to the phase-only bispectrum. In practice, one can use a simple 3-by-3 convolution mask [-1 -1 -1; -1 8 -1;-1 -1 -1] to enhance the peak before searching the maximum value. Notice also that the number of the slices k does not need to be two. Only one slice or more than two slices could be used instead.

After solving the scale and rotation parameters it is a straightforward task to determine the unknown translation parameters. One of the images $f(x, y)$ or $f'(x, y)$ is rotated and rescaled, and phase-correlation is then applied to estimate the missing parameters.

5. NUMERICAL EXAMPLES

In the first example, we used synthetic data shown in Figure 1 to test the performance of the method. The image in the left is *a target image* that has been obtained by scaling and rotating the original ”cameraman” image shown in the right side of Figure 1. The corresponding scaling factor is 1.33 and the rotation angle 20 degrees. The third image is a small subwindow that has been extracted from the original image. The original and its subimage are called *reference images*. The objective was then to estimate the scaling factor and the rotation angle, and to align the images. In the first case, we used the whole original image as a reference, and in the second case, only its subimage.

Fig. 1. Target and reference images.

Fig. 2. Results with the Fourier-Mellin transform.

Figure 2 shows the alignment result obtained by using the Fourier-Mellin transform. The two upper images are correlation images derived based on the log-polar sampled amplitude spectrums. The implementation follows the procedure described in [5]. The two lower images show the corresponding alignment results where the contours of the transformed reference images have been superimposed over the target image. As we can see, the alignment works perfectly in the first case and the correlation image has only one peak. However, in the second case the Fourier-Mellin transform fails completely, and the correlation image does not have any clear maximum. In the first case the estimates for the scale and the rotation parameters were 1.3312 and 20°, and in the second case 1.2101 and -3°.

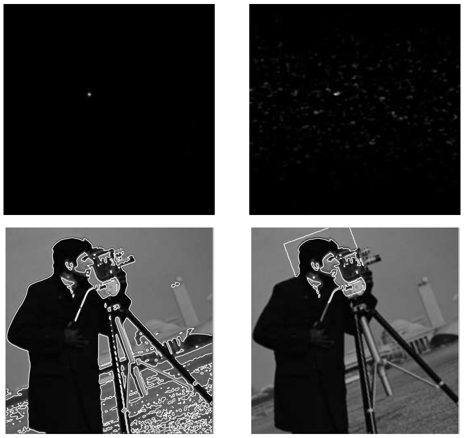

Fig. 3. Results with the new method.

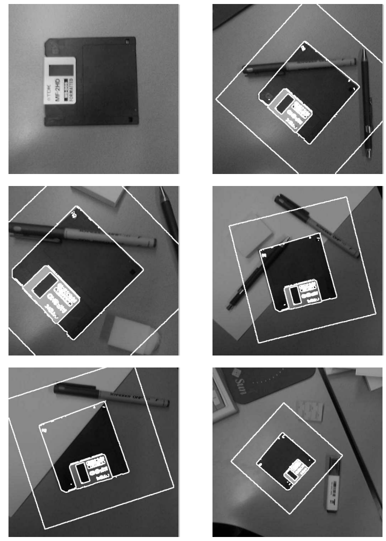

Fig. 4. Alignment results with the new method using real images.

Next, the same experiment was conducted with the new method. The results are shown in Figure 3. Now, we can see that the alignment works correctly in both cases. The correlation images have clear peaks and these peaks are located exactly in the same position giving the estimates 1.3312 and $20°$ for the scale factor and the rotation angle, respectively. The result obtained in the second case indicates that the new method is quite robust against the background clutter, which is an important property in template matching and image alignment.

In the second example, we used real images captured with a Nokia 6600 camera phone. The images are shown in Figure 4. The first one is a reference image representing a floppy disk and it has been aligned with the other five images. All images are noisy, partially unsharp, and they have highlights. There are also some other details in the background and the floppy disk is partially occluded. Moreover, the images have been captured freehand that is likely to cause some non-linear perspective distortion in addition to scaling, rotation and translation. Despite of these disturbing factors the method works well for all the target images.

6. CONCLUSIONS

In this paper, we have presented a method for estimating the scale and rotation parameters between two images. Although the basic idea of the method is quite similar to the Fourier-Mellin transform, it has certain advantages. The bispectrum used as a basis of the method preserves the phase information that is lost in the Fourier-Mellin transform. The phase information is also less sensitive to nonuniform illumination changes and it seems to be more robust against background clutter and occlusions. Moreover, the input images do not need to be windowed before applying the method.

7. ACKNOWLEDGEMENTS

This work was carried out in the project ”Methods for transform domain pattern analysis” supported by the Academy of Finland.

8. REFERENCES

[1] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, 2002.

[2] D. Casasent and D. Psaltis, “Position, rotation, and scale invariant optical correlation,” *Appl. Opt.*, vol. 15, no. 7, pp. 1795–1799, 1976.

[3] B. Sadler and G. B. Giannakis, “Shift and rotation invariant object reconstruction using the bispectrum,” *Journal of the Optical Society of America A*, vol. 9, pp. 57–69, 1992.

[4] V. Chandran, B. Carswell, B. Boashash, and S. L. Elgar, “Pattern recognition using invariants defined from higher order spectra - 2-d image inputs,” *IEEE Trans. Image Processing*, vol. 6, no. 5, pp. 703–712, 1997.

[5] Q. Chen, M. Defrise, and F. Deconinck, “Symmetric phase-only matched filtering of fourier-mellin transform for image registration and recognition,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, pp. 1156–1168, 1994.

[6] A. P. Petropulu and H. Pozidis, “Phase reconstuction from bispectrum slices,” *IEEE Trans. Image Processing*, vol. 46, no. 2, pp. 527–530, 1998.