Position, Rotation and Scale Invariant Object Detection using LabVIEW

Sukrit Dhar and Theja Tulabandhula
Department of Electrical Engineering, IIT Kharagpur, India
E-mail: sukritdhar2@yahoo.co.in, t.theja@iitkgp.ac.in

Category: Student

Products: NI CVS 1450

LabVIEW 8.5 (Professional Development System with MathScript)

Vision 8.5 (NI Vision).

The Challenge

Our objective is to develop and implement an efficient algorithm to recognize an object in a given image in the presence of variations like rotation, scaling and translation with respect to a reference image.

The Solution

Using the CVS 1450 hardware and the LabVIEW 8.5 software an efficient algorithm which is an extension of the well known Phase Only Correlation (POC) technique is developed to cover translation, rotation, and scaling invariance. Phase Only Correlation determines the translational movement between two images. Fourier scaling and Fourier rotational properties are used to further find scale and rotational difference between these two images of the same object. This algorithm shows excellent robustness against random noise apart from invariance to translation, rotation, scaling and illumination making it attractive for embedded image registration applications.

Theory

Image registration is a fundamental task in image processing. The registration method presented here uses the Fourier domain approach to match images that are translated, rotated, and scaled with respect to one another. This method relies on the translation property of the Fourier transform, which is also known as the Fourier shift theorem. Let f_1 and f_2 be two images that differ only by a displacement (x_0, y_0). That is,

$$f_2(x, y) = f_1(x - x_0, y - y_0)$$

Then, their corresponding Fourier transforms F_1 and F_2 will be related to each other as,

$$F_2(\xi,\eta) = e^{-j2\pi(\xi x_0+\eta y_0)} \times F_1(\xi,\eta)$$

where we see that the translation effect is confined only to the phase ($e^{-j2\pi(\xi x_0+\eta y_0)}$). The cross-power spectrum of two such images f and f‘ with Fourier Transforms F and F' is defined as

$$\frac{F(\xi,\eta)F'^{*}(\xi,\eta)}{\left|F(\xi,\eta)F'^{*}(\xi,\eta)\right|} = e^{i2\pi(\xi x_0+\eta y_0)}$$

where F^{*} is the complex conjugate of $F.$ The shift theorem guarantees that the phase of the cross-power spectrum is equivalent to the phase difference between the images. By taking inverse Fourier transform of the above representation in frequency domain, we will have a function which turns out to be an impulse; i.e., it is approximately zero everywhere except at the displacement that is needed to optimally register the two images. This transform can also be shown to be invariant of illumination. This is called the Phase Only Correlation (POC).

Now, we present the necessary theory to match images which are translated, rotated, and scaled with respect to each other using Fourier translation, rotation, and scaling properties.

Without Scale Change

If $f_2(x, y)$ is a translated and rotated replica of $f_1(x, y)$ with

$$f_2(x,y) = f_1(x\cos\theta_0 + y\sin\theta_0 - x_0, -x\sin\theta_0 + y\cos\theta_0 - y_0)$$

According to the Fourier translation property and the Fourier rotation property, transforms of f_2 and f_1 are related by

$$F_2(\xi,\eta) = e^{-j2\pi(\xi x_0+\eta y_0)} \times F_1(\xi\cos\theta_0 + \eta\sin\theta_0, -\xi\sin\theta_0 + \eta\cos\theta_0)$$

Let M_1 and M_2 be the magnitudes of F_1 and F_2. Therefore *we* have

$$M_2(\xi,\eta) = M_1(\xi\cos\theta_0 + \eta\sin\theta_0, -\xi\sin\theta_0 + \eta\cos\theta_0)$$

If we consider the magnitudes of F_1 and F_2, it is easy to see that the magnitudes of both the spectra are the same, but one is a rotated replica of the other. Rotational movement without translation can be deduced in a similar manner using phase correlation by representing the rotation as a translational displacement in the polar coordinate system:

$$M_1(\rho,\theta) = M_2(\rho,\theta-\theta_0)$$

With Scale Change

If f_1 is scaled replica of f_2 with scale factors *(a, b)* for the horizontal and vertical directions then, according to the Fourier scale property, the Fourier transforms of F_1 and F_2 are related by

$$F_2(\xi,\eta) = \frac{1}{|ab|} F_1(\xi / a, \eta / b)$$

By converting the axes to logarithmic scale, scaling can be reduced to a translational movement ignoring the multiplication factor *1/ab,* i.e.

$$F_2(\log\xi, \log\eta) = F_1(\log\xi - \log a, \log\eta - \log b)$$

i.e. $F_2(x, y) = F_1(x - c, y - d)$

where $x = \log\xi$, $y = \log\eta$, $c = \log a$ and $d = \log b$.

The translation *(c, d)* can be found by the phase correlation technique and the scaling *(a, b)* can be computed from the translation *(c, d)* as $a = e^c$ and $b = e^d$
If *(x,* y) is scaled to (*x / a , y / a) ,* their polar representation will be

$$\rho_1 = (x^2 + y^2)$$
$$\phi_1 = \tan^{-1}(y / x)$$
$$\rho_2 = ((x / a)^2 + (y / a))^2 = \rho_1 / a$$
$$\phi_2 = \tan^{-1}((x / a)/(y / a)) = \tan^{-1}(y / x) = \phi_1$$

i.e., if f_1 is a translated, rotated, and scaled replica of f_2 their Fourier magnitude spectra in polar representation are related by

$$M_1(\rho,\phi) = M_2(\rho / a, \phi - \phi_0)$$
$$M_1(\log\rho,\phi) = M_2(\log\rho - \log a, \phi - \phi_0)$$

i.e. $M_1(\xi,\phi) = M_2(\xi - d, \phi - \phi_0)$

where $\xi = \log\rho$, $d = \log a$

Using the phase correlation technique, scale a and angle ϕ_0 can be found out. Once the scale and angle information are obtained, the image with the highest resolution is scaled and rotated by amounts a and ϕ_0 respectively, and the amount of translational movement is found out using phase correlation technique.

System Configuration

In some applications of image processing an objective is to recognize the object in a given image in the presence of variations. These variations can be conveniently broken down as shift, rotation and scale. Fourier Mellin Transform(FMT), analytical Fourier Mellin Transform (AFMT), Two Dimensional Fast Fourier Transform (2D FFT), Log Polar Transform (LPT) and Phase Only Correlation (POC) belong to this Position Rotation and Scale Invariant (PRSI) class of transforms which can detect the above variations. These transforms have very important role to play in Content based Image Retrieval (CBIR) systems, active vision systems, image registration as well as image recognition. Most common applications include:

- Automatic object recognition
- Fingerprint Matching
- Face recognition (can also be used as shown here)

The main steps in system configuration are:

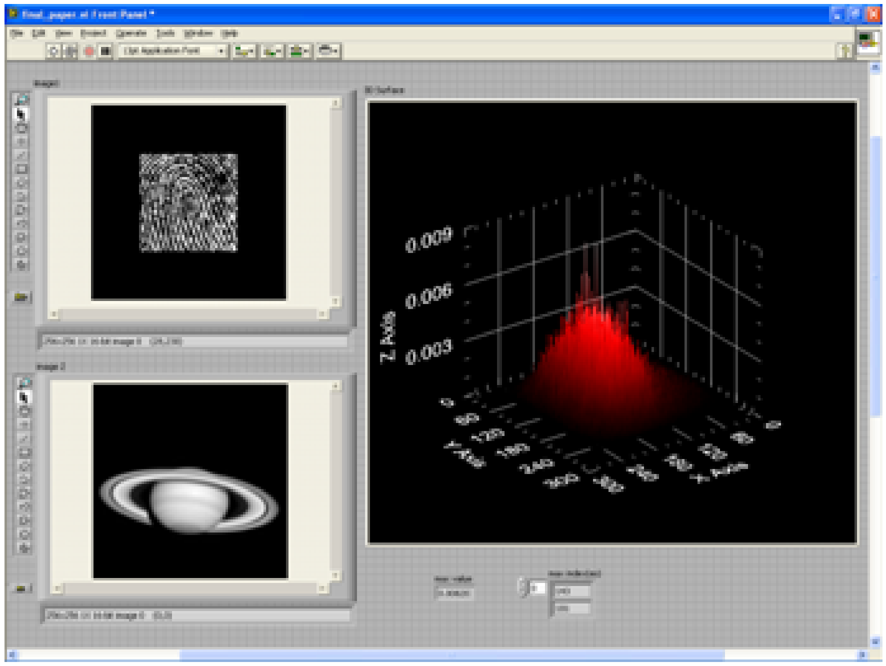
1. A database of the requisite images is created and stored in the computer.
2. The image which is to be compared is acquired with the help of NI CVS 1450.
3. The comparison algorithm developed in LabVIEW using NI Vision is applied.
4. The result is displayed.

System Implementation

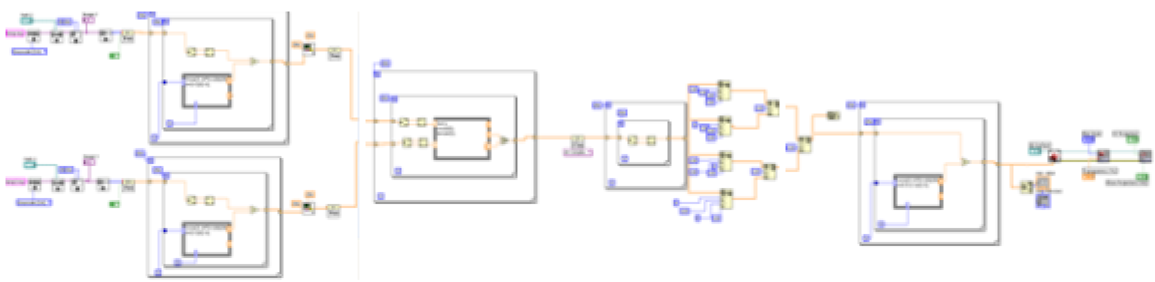
The first step in system implementation is to acquire the image in LabVIEW. Another image from the image database is selected. The magnitudes of the FFT of the two images are taken and the dc components are shifted to the centre. A smooth high pass filtering is done to enhance the output. Now as it is evident from the theory if one of the image is a scaled and rotated version of the other image then their magnitude responses should be identical in the log polar domain(pixel coordinates) but only with translation .

Next we do the log polar transform to convert the pixel coordinates from the Cartesian system to the log polar system. After the log polar transformation we do the phase only correlation (POC) between the two obtained image arrays to find out the degree of association between the two images and also the translation of one of the arrays if the two arrays are identical. Now from the value of this translation we can recover the degree of scaling and rotation of the two images if one is the rotated and scaled version of the other.

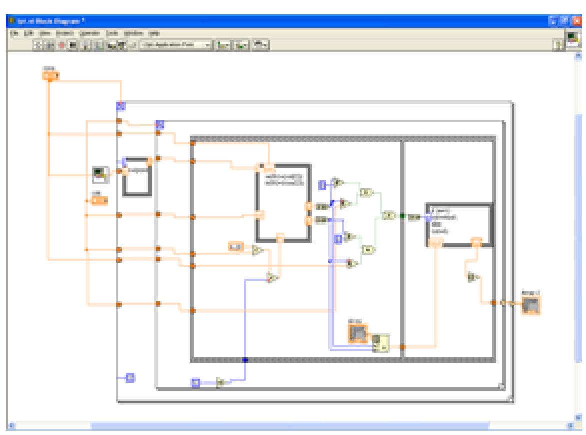
However one thing has to be kept in consideration that all these signal processing steps cannot be discretized. So signal processing blocks and coding techniques are utilized to find the output rather than using custom VIs from the IMAQ Vision utility.



Object recognition VI Front Panel



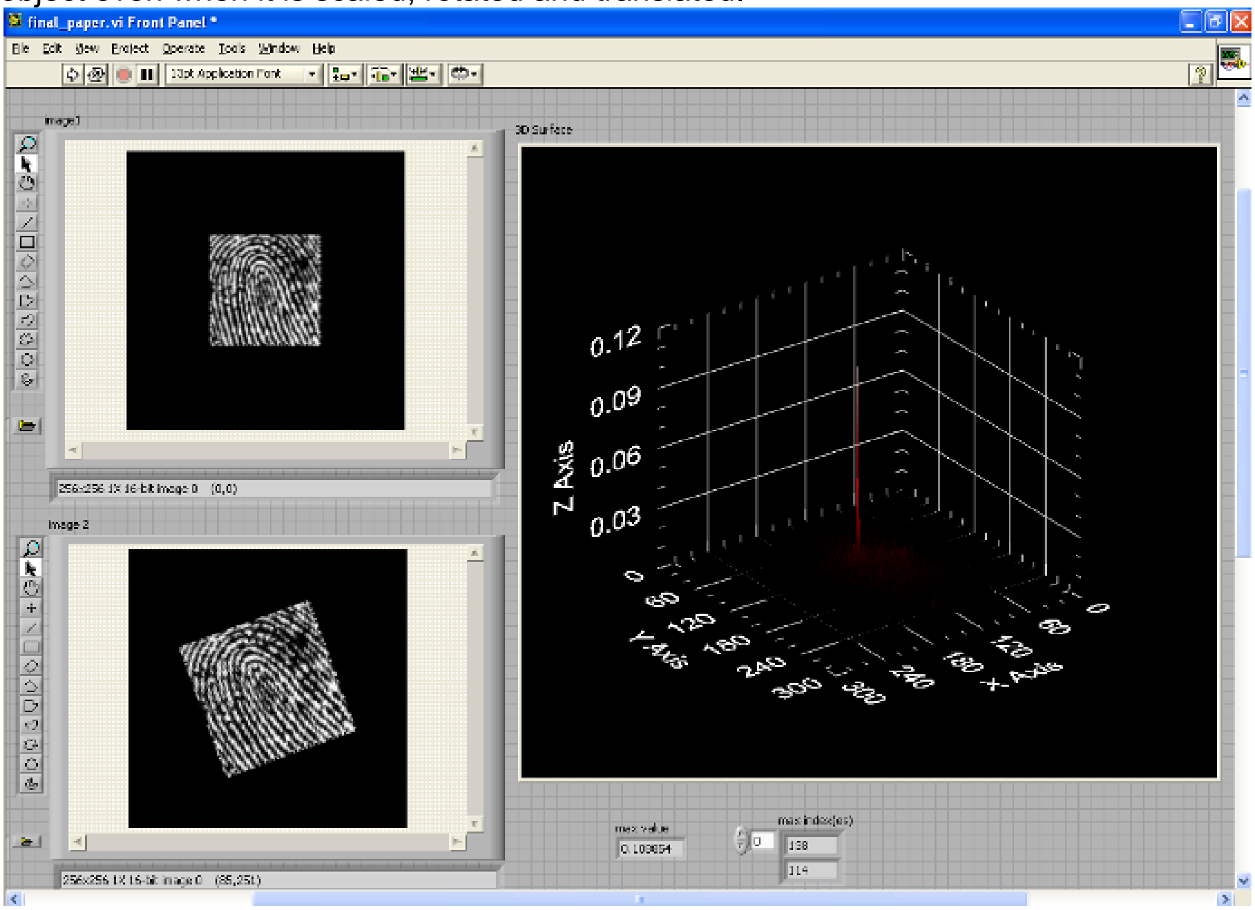
Object Recognition VI Block Diagram



Log Polar Transform Sub VI Block Diagram used in Object Recognition VI

Results

This technique is found to be efficient for object as well as face recognition.
For object recognition in case of similar objects which are scaled, rotated or translated it will give a distinct peak with the maximum value of the final matrix being of the order of $\geq 6*10^{-2}$ (worst case scenario if there is too much scaling, above 200%) where as in case of different objects it will give a random noise, the maximum value being of the order of $\leq 1*10^{-2}$. So a threshold can easily be set at $0.5*10^{-2}$ for the detection of an object even when it is scaled, rotated and translated.



Result for two images which are scaled and rotated versions of one another

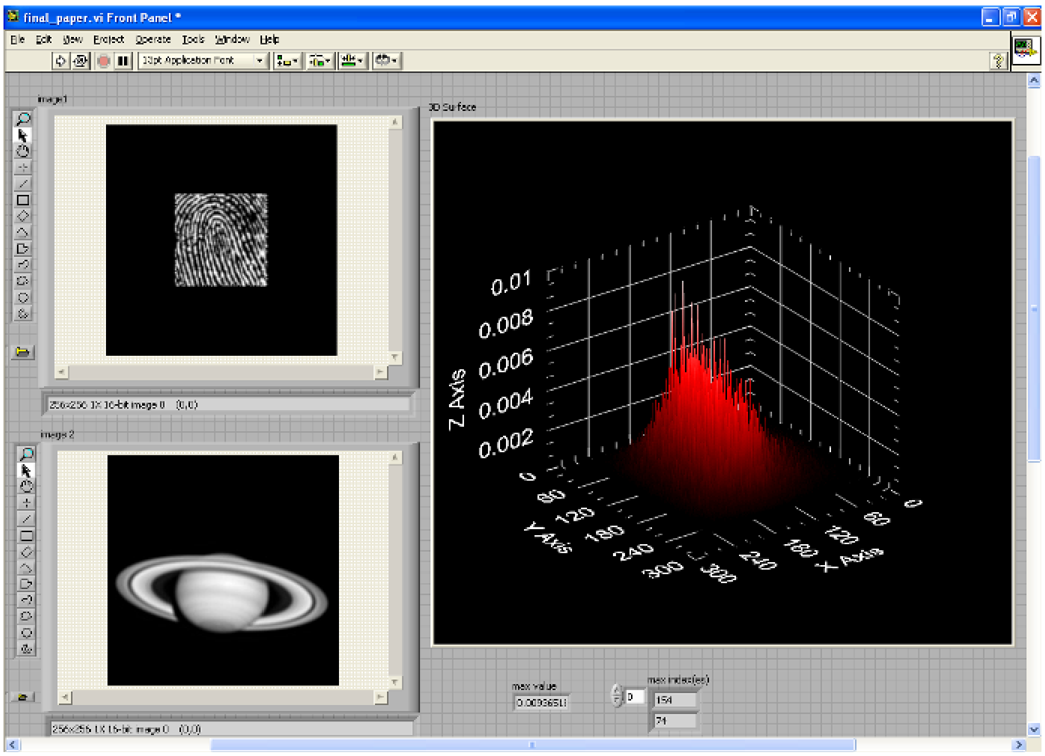
Maximum value of the matrix = 0.108854 that is of the order of 10^{-1}. Characterized by one distinct peak. From the coordinates of the peak the scaling and rotation could be calculated. The images have a resolution of 256*256 pixels. The y coordinate of the peak is at 114 .That is, it is shifted from the central point by (128-114) =14 pixels. As the phase is linearly dependant so the angle of rotation is calculated to be:

$(14*2\pi/256)$ =19.6875. (Actual rotation given is 20^0) : Error =0.3125^0.

The x coordinates of the peak is at 138, shifted from the central point by 10 pixels .So, the scaling comes to be:

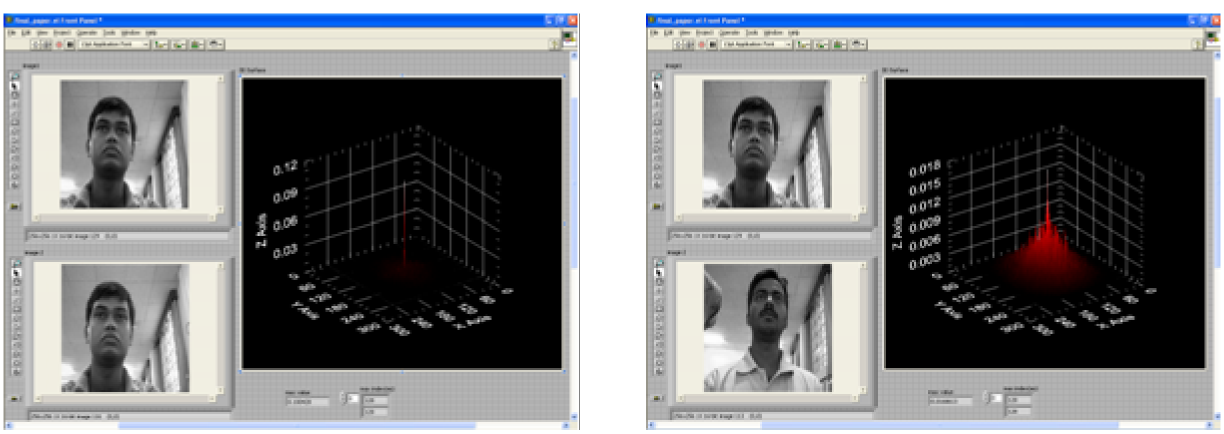
(-log (10/256)-1)*100%=39.79 % (Actual scaling was 25%): Error = 14%

This error is exponential in nature. If we recapitulate the theory we will remember that during the derivation we have taken the product of the two scale factors of the two axes (*1/ab)* to be approximately equal to one. So, for scaling up to 20%, it gives reasonably accurate result, where as it will continue to give correct output for object recognition even above a scaling of 300%. This proves the robustness of the algorithm.



Result for two different images

Maximum value of the matrix = 0.00936518 that is of the order of 10^{-3}. No distinct peak. This same technique can also be used with considerable success in case of face recognition under various conditions with the difference in the maximum being approximately 10 times as before.



Results obtained for face recognition (single peak on the left (see magnitudes))

System Setup

Conclusion

Much scope is still there for future improvements. This algorithm can successfully solve the problems like fingerprint matching under various conditions as well as object detection. So it can be used commercially, like in malls, for automated object detection as well as for security applications .Several other usages are also possible. The user friendliness and the several inbuilt VI functionalities in LabVIEW 8.5 greatly reduce the development time of any algorithm.

References

1. D. I. Bamea and H. F. Silverman, “A class of algorithms for fast digital registration,” *IEEE Trans. Computers,* vol. C-21, pp. 179-186, 1972.
2. C. D. Kuglin and D. C. Hines, “The phase correlation image alignment method,” in *Proc. IEEE 1975 Int. Conf Cybernetics Society,,* New York, NY, pp. 163-165.
3. J. L. Horner and P. D. Gianino, “Phase-only matched filtering,” *Applied Optics,* vol. 23, no. 6, pp. 812-816, 1984.
4. Theja Tulabandhula, Amit Patra, Nirmal B Chakrabarti, “The design of a Two Dimensional PRSI Processor”, 11th Euromicro Conference on Digital System Design: Architecture, Methods and Tools, Parma, Italy pp197-205, 2008.