# A novel algorithm for a rotation invariant template matching*

**LEI Ming****, and **ZHANG Guang-jun**

*School of Instrumentation Science and Optoelectronics Engineering, Key Laboratory of Precision Optomechatronics Techonology, Ministry of Education, Beijing University of Aeronautics and Astronautics, Beijing 100083, China*

(Received 23 May 2008)

A novel algorithm for a rotation invariant template matching is proposed when the fluctuating scope of the rotation angle is limited within the region of [-20°,20°]. The matching candidates are selected using a computationally low cost improved correlation algorithm. "AND" operation is adopted to reduce the computational cost. Therefore the algorithm improves the matching speed consumedly. The simulation results verify the efficiency of the proposed method. Moreover, when the size of reference image is fixed, the advantage of this time-saving algorithm is more obvious as the increase of the size of the real time image. The matching speed of the proposed method is over 20 times faster than the speed of the two-level pyramid decomposing accelerating method.

**CLC numbers:** TP391.4 **Document code:** A **Article ID:** 1673-1905(2008)05-0379-5

**DOI** 10.1007/s11801-008-8043-1

The template matching process involves cross-correlation between the template and the scene image[1,2]. Since the evaluation of the correlation is computationally expensive, there is a need for low-cost correlation algorithms for real-time processing. A large number of correlation-type algorithms have been proposed[3,4]. One of the approaches is to use an image pyramid for both the template and the scene image, and to perform the registration by a top-down search[5]. Other fast matching techniques use two pass algorithms: use a subtemplate at a coarsely spaced grid in the first pass, and search for a better match in the neighborhood of the previously found positions in the second pass[6].

However, when objects in the image are rotated with respect to each other, the methods described above cannot be used, and a set of templates at different orientations is to be used. This procedure is hardly practical for real-time processing when the rotation angle is arbitrary or unconstrained. In such applications, rotation invariant moment-based methods can be used[7,8]. The moment-based methods, however, require too much computation for practical purposes and are sensitive to noise. In this paper, we propose a novel fast rotation invariant template matching method when the fluctuating scope of the rotation angle is limited within the region of [-20°,20°].

The convolution algorithm is often used in digital image processing, which is simple and efficient. The convolution algorithm and the correlation algorithm are all based on the template matching. Meanwhile, they bear partial similarities in terms of algorithm (seeing Eq.(2) and the numerator in Eq.(3)). Inspired by it, a novel and simplified algorithm similar to the convolution algorithm is proposed, which has not only reduced the complexity of the correlation algorithm, but also withstanded rotation within certain angle. The theories of the convolution algorithm and the correlation algorithm are as follows:

Based on 2D image, we will only explain the theory of 2D convolution. The convolution of the 2D continuous function is similar to the 1D convolution, the only difference is two independent variables, $x$ and $y$. The 2D convolution expression is as follows:

$$h(x,y)=f*g=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(u,v)g(x-u,y-v)\,\mathrm{d}u\,\mathrm{d}v\ . \quad (1)$$

The convolution of the digital image is similar to that of the continuous function, the only difference is that the value of the independent variable is integer. The double integration is transformed into double sum. Therefore, we get Eq.(2) for a digital image,

$$H=F*G \text{ and } H(i,j)=\sum_m\sum_n F(m,n)G(i-m,j-n)\ , \quad (2)$$

where $G$ is convolution kernel, $F$ denotes digital images, and $H$ is the result of the convolution. Because the values of $F$ and $G$ are nonzero only within a limited range, we only need

* This work has been supported by the preparing Fund for defence equipment (No.6140517)

** E-mail: lmeagle01@163.com.

to calculate the sum of the field which is overlapped with the nonzero part.

The normalized cross-correlation theory is as follows:

Let us define a template to be $\boldsymbol{G}_s$ with the size of $M_s \times N_s$ pixels and an image $\boldsymbol{G}_r$ with the size of $M_r \times N_r$ pixels, respectively, with $M_s < M_r, N_s < N_r$. Define the top left corner of the image as the origin, and the sub-image as $\boldsymbol{G}_r(u,v)$ in the image with the size of $M_s \times N_s$ pixels, then the normalized cross-correlation coefficient $\boldsymbol{\rho}(u,v)$ between $\boldsymbol{G}_s$ and $\boldsymbol{G}_r(u,v)$ is defined as[9]

$$\rho(u,v)= \frac{\sum_{i=1}^{M_s}\sum_{j=1}^{N_s}[G_r(i+u,j+v)-\overline{G_r(u,v)}]\times[G_s(i,j)-\overline{G_s}]}{\sqrt{\sum_{i=1}^{M_s}\sum_{j=1}^{N_s}[G_r(i+u,j+v)-\overline{G_r(u,v)}]^2}\sqrt{\sum_{i=1}^{M_s}\sum_{j=1}^{N_s}[G_s(i,j)-\overline{G_s}]^2}}, \tag{3}$$

where the grayscale average of $\boldsymbol{G}_r(u,v)$ and $\boldsymbol{G}_s$ is respectively represented with $\overline{G_r(u,v)}$ and $\overline{G_s}$, and $\rho(u,v)$ is in the scope of [-1, 1]. The best matching location is determined by the maximum cross-correlation coefficient corresponding to the correlation peak, and the matching error is determined by the location of the correlation peak, and the size and position of the template.

Enlightened by the convolution algorithm, we consider to simplify the cross-correlation algorithm so that it will have the algorithmic mode of the convolution algorithm. In this way, the computation will be greatly reduced. However, we must point out that this change just takes the convolution algorithm as reference. Although they seem similar to convolution in appearance, they are different in meaning, therefor, the novel algorithm does not have the operational properties of the convolution. The realization of the simplification is as follows.

In Eq.(3) above mentioned, we presume

$$\boldsymbol{G}_{S_{i,j}}=\sqrt{\sum_{i=1}^{M_s}\sum_{j=1}^{N_s}[\boldsymbol{G}_s(i,j)-\overline{\boldsymbol{G}_s}]^2}, \quad \text{and}$$

$$\boldsymbol{G}_{R_{i,j}}=\sqrt{\sum_{i=1}^{M_s}\sum_{j=1}^{N_s}[\boldsymbol{G}_r(i+u,j+v)-\overline{\boldsymbol{G}_r(u,v)}]^2},$$

then, Eq.(3) can be transformed into Eq.(4) as follows:

$$\rho(u,v)=\frac{\sum_{i=1}^{M_s}\sum_{j=1}^{N_s}[\boldsymbol{G}_r(i+u,j+v)-\overline{\boldsymbol{G}_r(u,v)}]\times[\boldsymbol{G}_s(i,j)-\overline{\boldsymbol{G}_s}]}{\boldsymbol{G}_{R_{i,j}}\boldsymbol{G}_{S_{i,j}}}. \tag{4}$$

Supposing that we only study the binary image, obviously, to remove the denominator in Eq.(4) and make it similar to the mode of the convolution, we must focus on the denominator in Eq.(4). After studying Eq.(4), we find that $\boldsymbol{G}_{S_{i,j}}$ is a fixed value while others are all variable, which increase the complexity of the computation. This is because $\rho(u,v)$ is mainly about the degree of the cross-correlation. So, we just consider the correlation coefficient based on the template and abandon the computation of $\boldsymbol{G}_{R_{i,j}}$ in Eq.(4), reducing the computational cost.

The method is as follows:

The pixel values of the binary image are usually 0 and 255 or 0 and 1. It is Supposed that the pixel values of the binary image are only 0 and $r$. Since it is based on the template, the sum of square of each pixel in the template can be used as the denominator, instead of the denominator in Eq.(4). So the correlation coefficient $\rho(u,v)$ can be simplified as

$$\rho(u,v)=\frac{\sum_{i=1}^{M_s}\sum_{j=1}^{N_s}\boldsymbol{G}_r(i+u,j+v)\times\boldsymbol{G}_s(i,j)}{\sum_{i=1}^{M_s}\sum_{j=1}^{N_s}\boldsymbol{G}_s^2(i,j)}. \tag{5}$$

Obviously, the denominator of Eq.(5) is a constant. Supposing it is $N$, Eq.(5) can be further simplified into Eq.(6)

$$\rho(u,v)=\frac{1}{N}\left(\sum_{i=1}^{M_s}\sum_{j=1}^{N_s}\boldsymbol{G}_r(i+u,j+v)\times\boldsymbol{G}_s(i,j)\right). \tag{6}$$

Apparently, Eq.(6) is like the mode of the convolution algorithm, with $0\le\rho\le1$. Besides, there are only two pixel values (0 and $r$). If $r$=1, then, the operation in the brackets of Eq.(6) is actually equal to the sum of the "AND" operation of the correspondent pixels in the two images. Otherwise, the operation will be equal to the sum of squares of the "AND" operation. Because it is a binary image, no matter the former or the latter, their computational performances are the same with that of the "AND" operation. But the latter offers $\boldsymbol{r}^2$, which is also a fixed constant. Therefore, Eq.(6) can be changed into Eq.(7)

$$\begin{cases}\rho(u,v)=\frac{1}{N}\left(\sum_{i=1}^{M_s}\sum_{j=1}^{N_s}\boldsymbol{G}_r(i+u,j+v)\bullet\boldsymbol{G}_s(i,j)\right), \\ \boldsymbol{G}_r(i+u,j+v)\bullet\boldsymbol{G}_s(i,j)=\begin{cases}0\\1\end{cases} \quad \text{if } r=1, \\ \boldsymbol{G}_r(i+u,j+v)\bullet\boldsymbol{G}_s(i,j)=\begin{cases}0\\r^2\end{cases} \quad \text{otherwise}.\end{cases} \tag{7}$$

In Eq.(7), "●" denotes "AND" operation. Undoubtedly, it is of great value to make the computation more efficient and the hardware implementation of the algorithm is more realizable. On the other hand, there might be mismatching that makes $\rho = 1$ where all the pixel values are $r$ in some region of the reference image. Therefore, these regions must be removed by forcedly making $\rho = 0$ there. The sum of the pixel values in the sub-image covered by the template is Supposed to be $\Psi$, with $M < \Psi \leqslant \Phi$ , $\Phi$ is the sum of the pixel values in the template. $M$ can be adapted according to the need of precision. Here $M = \Phi / 2$. When the template is moved, it just needs to compute the sum of the pixel values in the sub-image covered by the template within this range. As a result, it can not only accelerate the operational speed, but also remove a large amount of false matching positions.

Eq.(3) and Eq.(7) are used to the same image and template, respectively. Fig.1 shows distributions of correlation peaks. In the illustration, the pane stands for the place where the value of $\rho$ is highest, and the location of the pane in Fig.1(a) is the factual location of the template in the reference image.

Experiments were carried out to measure the performance of our algorithm in terms of speed, matching precision and matching accuracy. As proposed in this paper, methods of computation for selecting candidates, the hierarchical Zernike method[10] and the circular projection method of computation in the frequency domain[11] are compared. The hierarchical Zernike method utilizes the pyramid structure of the image and results in comparable to ours in terms of the accuracy.

The proposed algorithm has been tested with four types of variations such as rotation, brightness change, contrast change and added Gaussian noise.

Fig.2 shows a set of templates masked off from the original image, and Fig.3(a)-(d) are the rotated images at arbitrary angles in the scope of [-20°, 20°].

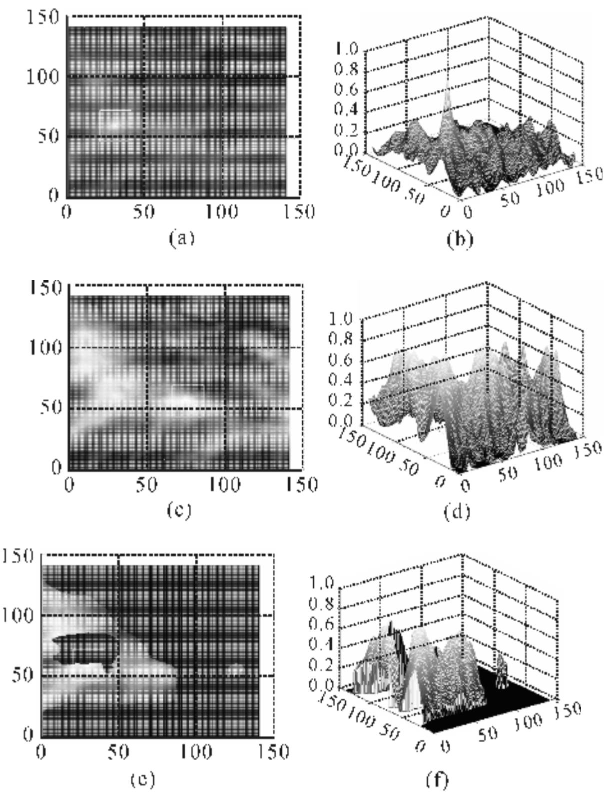


**Fig.1 Correlation peaks. (a) plan view of the correlation peak generated from Eq.(3), (b) correlation peak generated from Eq.(3), (c) plan view of the correlation peak generated from Eq.(3)(image rotates by 30°), (d) correlation peak generated from Eq.(3)(image rotates by 30°), (e) plan view of the correlation peak generated from Eq.(7)(image rotates by 30°), (f) correlation peak generated from Eq.(7)(image rotates by 30°).**

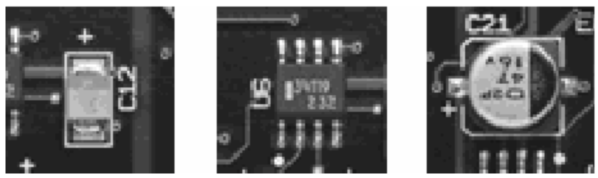

**Fig.2 A set of templates.**

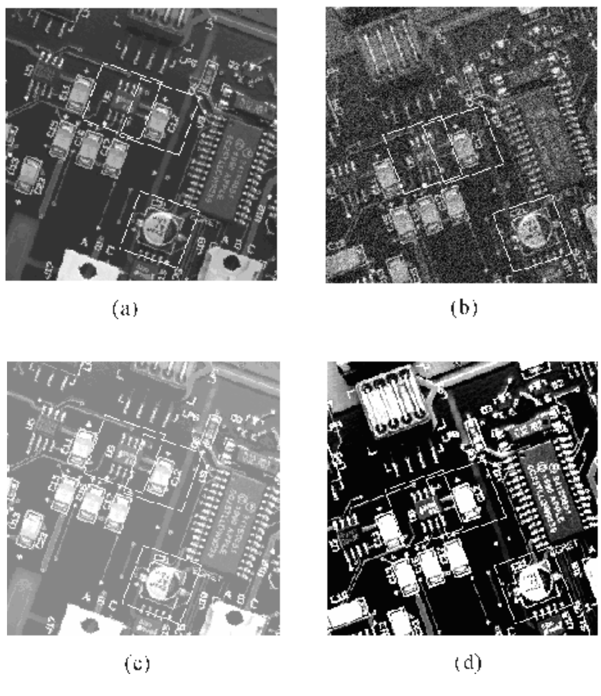


**Fig.3 Test images (a circuit board). (a) a rotated image, (b) rotated image with Gaussian noise, (c) rotated image with brightness variation, (d) rotated image with contrast variation.**

To verify the matching performance of our proposed algorithm in terms of speed, several experiments were carried out with various sizes of templates for a set of 256 × 256 images. Template sizes varied from 31×31 to 71×71. The proposed method was also compared in terms of the speed to the existing algorithm, the hierarchical Zernike method and the circular projection method of computation for selecting candidates in the frequency domain. The speed test was performed on a 2.4 GHz Pentium(R) 4 PC. The result is shown in Tab.1.

The speed comparison of several algorithms shows that the proposed algorithm is the fastest.

**Tab.1 Speed of matching (*ms*)**

| size of template | Proposed method | Circular projection | Hierarchical Zernike |
|---|---|---|---|
| 31 × 31 | 0.0032 | 0.0310 | 0.1120 |
| 41 × 41 | 0.0063 | 0.0288 | 0.1412 |
| 51 × 51 | 0.0082 | 0.0295 | 0.1885 |
| 61 × 61 | 0.0093 | 0.0302 | 0.2352 |
| 71 × 71 | 0.0109 | 0.0298 | 0.2971 |

For the accuracy of our proposed algorithm, a matching ratio was also compared. The matching ratio is defined as the number of correct matches to the number of trials. For this accuracy testing, 30 similar images of electronic circuit boards were captured, and 15 templates of interested area were selected manually from each image, resulting in 450 templates in total. Various types of noise such as Gaussian noise, and contrast/brightness changes have been added to each image. Then each template has been searched for from the corresponding image. The matching ratio indicates how well the queried template is retrieved at the precise location from the corresponding image. Obviously, the matching ratio depends on the number of candidates for each trial. The ratio becomes higher as the number of candidates for each template increases, but the computation takes longer.

Fig.4 shows the relationship between the accuracy and the rotation angles. The rotation matching test was performed with each rotation of 5°.The transverse error $\Delta x$ and the longitudinal error $\Delta y$ are required to satisfy the constraint: $(N-\Delta x)(N-\Delta y)\geqslant 80\%N^2$ in order to have 80% of the overlap between the template and the sub-image with the same size of the template, which can guarantee to obtain the accurate scaled and rotated factors by using Fourier-Mellin transform. Where $N$ is the size of the template, $N^2$ is the area of the template.

Additive white Gaussian noise was added to each image. The plot in Fig.5 shows the relationship between the accuracy and the number of candidates. As evidenced by the experiment, when 40 candidates were selected for each

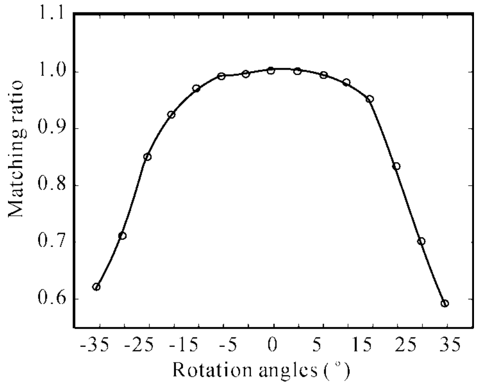


**Fig.4 Matching ratio vs rotation angles.**

template, the accuracy ratio ranked highest. The accuracy degrades only by a small portion as the SNR decreases because the binarization reduces the effect of Gaussian noise.

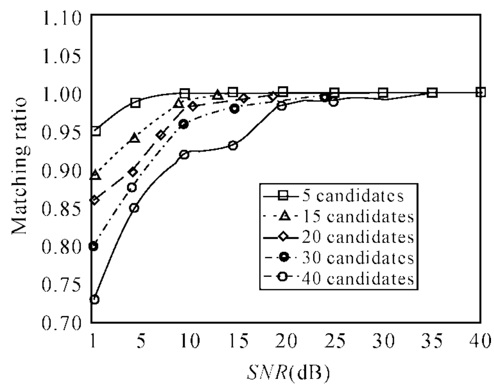


**Fig.5 Matching ratio vs *SNR* at Gaussian noise.**

Fig.6 shows the matching accuracy when the brightness of the image changes due to the lighting conditions. The proposed algorithm survives by more than 90% even with severe changes (-40-40 in gray values) in brightness. It is because the binarization eliminates the effect of the overall brightness change. On the other hand, the correlation algo-

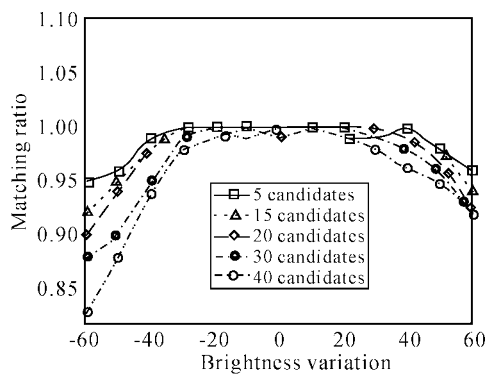


**Fig.6 Matching ratio vs brightness variation.**

rithms are not sensitive to brightness variation. Fig.7 shows the matching accuracy when the contrast of the image changes. The proposed algorithm is robust to contrast change, too.

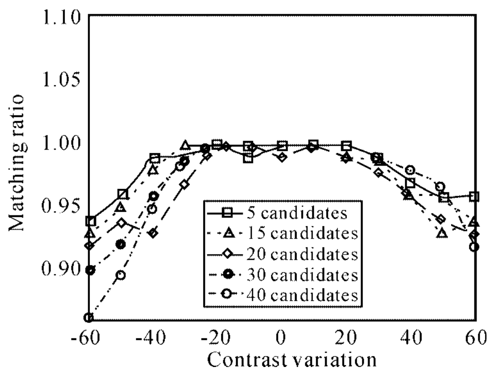


**Fig.7 Matching ratio vs contrast variation.**

Eqs.(8) and (9) were used to change the brightness and the contrast of the images. The abscissa in Figs.6 and 7 indicates the value of $d$ in the equations.

$$f_b = f + 255 \times \frac{d}{100} \quad , \tag{8}$$

$$f_c = f + (f - f_{avg}) \times \frac{d}{100} \quad . \tag{9}$$

In this paper, we proposed a fast and robust algorithm for template matching of an arbitrarily rotated image when the fluctuating scope of the rotation angle is limited within the region of [-20°,20°]. The proposed method employed to select candidates has efficiently reduced the order of feature space using the "AND" operation. In the experiment, we showed that it was computationally simple and fairly reliable for selecting candidates. In addition, the time complexity for the candidate selection process increased very slowly compared with the template size. But it is necessary to point out that the excellent performance of the proposed method employed to select candidates is achieved at the expense of matching accuracy. So the proposed method is more fitted for the coarse matching process.

## References

[1] Baker E S, and Degroat R D, IEEE Trans on Signal Proceesing, **46** (1998), 3112.

[2] Brown L G. ACM Computing Surveys, **24** (1992), 325.

[3] Cahnvon Seelen U M, and Bajcsy R. University of Pennsylvania: GRASP Laboratory Technical Report, 1996.

[4] R.G Caves, P.J Harley, and S.Quegan. IEEE Transactions on geoscience and remote sensing, **30** (1992), 680.

[5] Wong R Y, and Hall E L, IEEE Trans on Computer, **27** (1978), 359.

[6] A. Rosenfeld, and A. Kak. Digital Image Processing, 2nd Edition, Academic Press, Orlando, 1982.

[7] R.J. Prokop, and A.P. Reeves. CVGIP Graph. Models Image Process. **54** (1992), 438.

[8] Sang-Woo Lee, and Whoi-Yul Kim. Proc KITE, **19** (1996), 475.

[9] Luigi Di Stefano, Maschine Vision and Applications, **13** (2003), 213.

[10] Sun-Gi Kim, and Whoi-Yul Kim. Proceedings KSPC, **96** (1996), 1335.

[11] Min-Seok Choi, and Whoi-Yul Kim. Pattern Recognition, **35** (2002), 119.