JOURNAL OF MODERN OPTICS, 2003, VOL. 50, NO. 9, 1353–1364

Taylor & Francis
Taylor & Francis Group

Polychromatic pattern recognition using the non-zero-order joint transform correlator with cross-correlation peak optimization

CHULUNG CHEN* and CHIH-SUNG WU

Department of Electrical Engineering, Yuan Ze University, Taoyuan 320, Taiwan

(*Received 3 April 2002*)

Abstract. A multi-channel polychromatic non-zero-order joint transform correlator (NOJTC) with minimum average cross-correlation energy is introduced. The system is implemented using a general JTC scheme with additional power-spectrum subtraction strategy and the design of an ideal reference function with a Lagrange multipliers technique. As the zero-order part is removed, the desired correlation peaks can be quite distinctive and sharp. The correlator also shows the advantage of better detection ability over monochromatic systems.

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

1. Introduction

The most common techniques used for optical pattern recognition are the joint transform correlator (JTC) [1] and Vander Lugt's filter based system [2]. To reduce the system alignment requirement along with the optical axis in the 4-f Vander Lugt setup, the JTC system arranges the reference and the target side by side at the input plane. Both of these two correlators have the characteristic of shift-invariance that can be used for targets tracking in pattern recognition. Nevertheless, the correlation operations performed by conventional JTCs (CJTCs) result in large correlation sidelobes, large zero-order bandwidth, low light efficiency and poor discrimination. These drawbacks are practically due to the existence of the zero-order power spectrum, which also restrains applications to multi-target detection. Hence, a lot of research into improving JTC performance has been done. For example, Lu *et al.* [3, 4] introduced a non-zero-order JTC (NOJTC) with a phase-shifting technique. Some other methods for eliminating zero-order have also been proposed and demonstrated [5, 6]. In the topic of distortion invariant pattern recognition, Alam *et al.* [7] designed a fringe-adjusted JTC based on the Newton–Raphson algorithm. However, there is no theoretical guarantee that this algorithm will converge to a solution in every case. Recently, Chen *et al.* [8–10] adopted the technique of Lagrange multipliers to design a reference function that can produce sharp correlation peaks for monochromatic images.

* Author for correspondence; e-mail: chulung@saturn.yzu.edu.tw

Journal of Modern Optics ISSN 0950–0340 print/ISSN 1362–3044 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0950034021000011239

Not only shapes but also the colours are essential characteristics for pattern recognition. In most cases, Vander Lugt correlators and JTCs are implemented to handle monochromatic images. In the real world, however, most visual signals are composed of polychromatic information, which drives the development of polychromatic pattern recognition [11–13]. In the mean time, the idea of multi-channel JTC has been proposed [14, 15]. To achieve colour pattern recognition, Deutsch *et al.* [16] demonstrated a multi-channel single-output JTC by the arrangement of objects for all colour channels at the input plane. Similar to CJTCs, their concept led to an even larger zero-order part formed by six auto-correlation terms that result in poor detection efficiency. Recently, to obtain better performance in multi-channel single-output JTC, Alam *et al.* [17] proposed a fringe-adjusted JTC technique where a fringe-adjusted filter (FAF) is applied to the joint power spectrum (JPS). The convergence of the algorithm has to be investigated further.

In this paper, we propose a linearly constrained optimization method for performing polychromatic pattern recognition with an NOJTC, which has many advantages over other monochromatic or conventional systems. We will show the procedure step by step to obtain the desired optimized function. The proposed correlator is found to have superior discrimination ability.

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

2. Analysis

The basic concept of the proposed system is founded on the JTC proposed by Yu [18]. The architecture diagram of the JTC system is shown in figure 1. This opto-electronic hybrid system consists of two spatial light modulators (SLMs) and two CCD cameras. A laser beam illuminates the first SLM comprising the joint input images at the first plane, which is exactly the front focal plane of the first

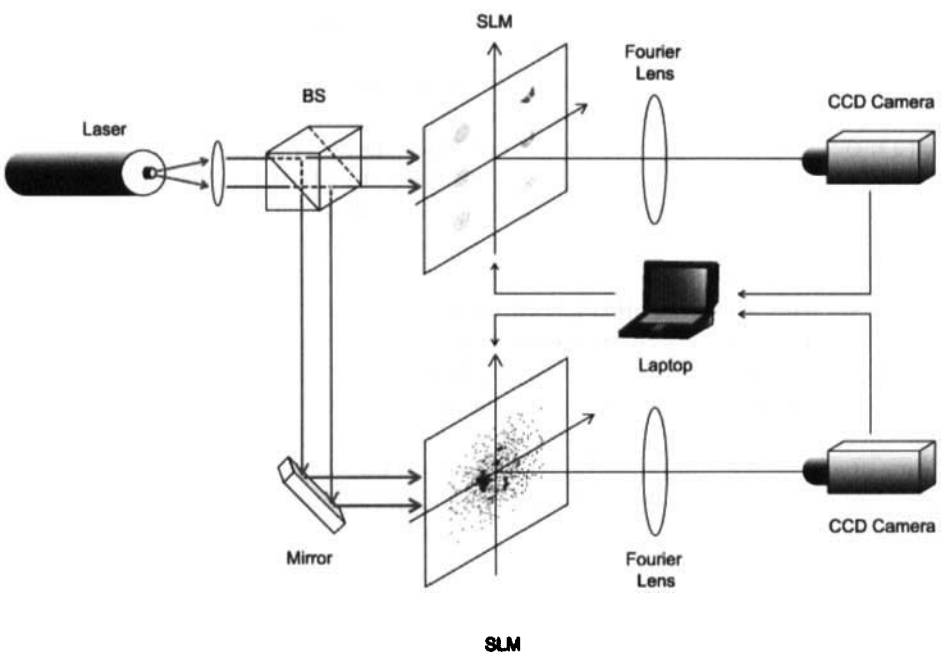

Figure 1. Polychromatic NOJTC architechure.

Fourier lens. A CCD camera located at the back focal plane of the first Fourier lens is utilized to capture the joint transform power spectrum (JTPS), which is actually the absolute magnitude square of the Fourier transform of the input scene. The captured digital information is read by a laptop computer for the subsequent JTPS subtraction strategy to remove the zero-order term. The subtracted JTPS is then transferred through the laptop to the second SLM for inverse Fourier transform. Because of the special property of Fourier transform pairs, the image showing at the back focal plane of the second Fourier lens is in fact a 180-degree rotation of the desired one. Finally, the CCD camera at the back focal plane of the second Fourier lens can obtain the cross-correlation output without zero-order term.

The joint input image consists of two parts: the left part contains RGB channels from the synthesized reference image, and the right part containing the input target scene. For numerical analysis, an input scene with objects locating at the first SLM is described by

$$t(x,y) = \sum_{n=1}^{6} f_n(x - x_n, y - y_n) \tag{1}$$

where (x_n, y_n) is the channel position of either the reference image ($n = 1$, 2, 3 or R, G, B) or the test image ($n = 4$, 5, 6 or r, g, b) at the input domain.

Since both the reference function and test image are separated into R, G, B channels, there are 6 terms in this case. The corresponding JTPS captured by the CCD camera, which is in fact the magnitude squared of the Fourier transform for the joint input scene, can be rewritten as

$$I(\alpha,\beta) = \left| \sum_{n=1}^{6} F_n(\alpha,\beta) \exp[-i(\alpha x_n + \beta x_n)] \right|^2 \tag{2}$$

where $F_n(\alpha,\beta)$ symbolizes the Fourier transform of objects $f_n(x,y)$, while (α,β) is the spatial coordinate system of the joint-spectrum plane.

The correlation distribution at the output plane $O(x',y')$ can be written as

$$O(x',y') = \sum_{n=1}^{6} C_{nn}(x',y') + \sum_{n=1}^{6} \sum_{\substack{m=1 \\ m\neq n}}^{6} C_{nm}[x' - (x_m - x_n), y' - (y_m - y_n)] \tag{3}$$

where the correlation C_{nm} of objects f_n and f_m is defined as

$$C_{nm}(x',y') = \int\int f_n^*(x,y) f_m(x + x', y + y') \mathrm{d}x\mathrm{d}y \tag{4}$$

Let $(x_1,y_1) = (-x_0,y_0)$, $(x_2,y_2) = (-x_0,0)$, $(x_3,y_3) = (-x_0,-y_0)$, $(x_4,y_4) = (x_0,y_0)$, $(x_5,y_5) = (x_0,0)$, and $(x_6,y_6) = (x_0,-y_0)$. The output plane of the NOJTC is shown in figure 2; the subscripts in capital letters and lowercase letters represent the input reference and target image, respectively.

The first term on the right-hand side in equation (3) represents the sum of the six autocorrelation terms appear at the centre of the output plane. This is the annoying part that some researchers have tried to eliminate. The second term is the summation of cross-correlation terms between any pair of input objects. The distribution of this part depends on the locations of objects involved in the correlation process at the input plane. Among these cross-correlation terms are

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

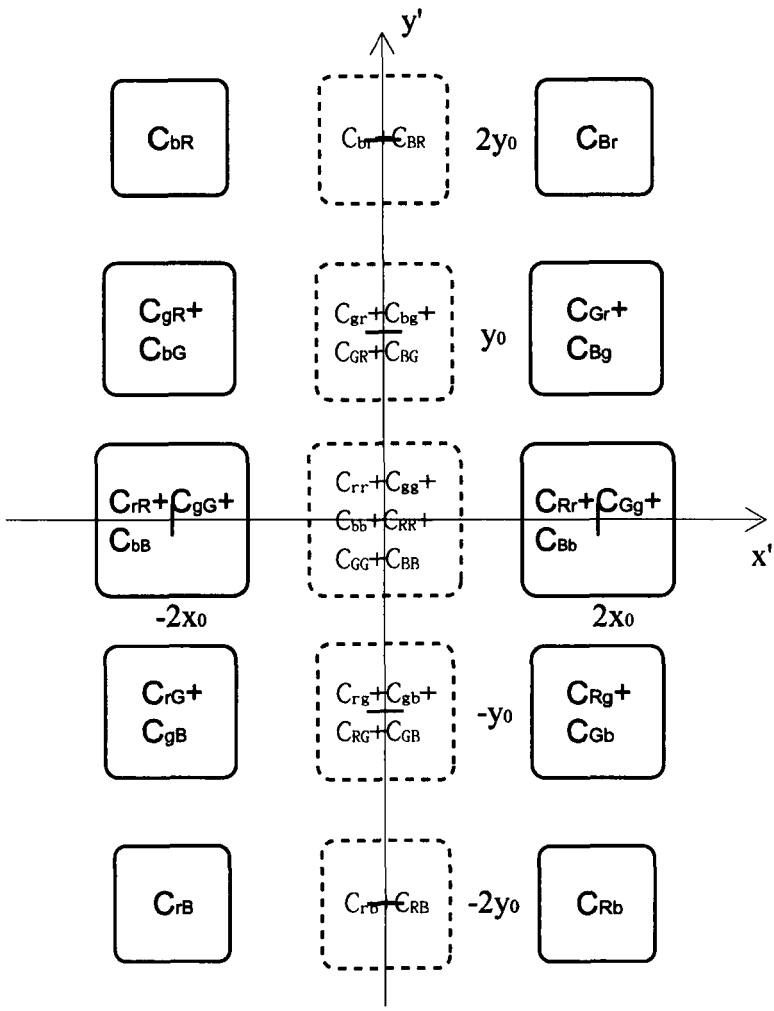

Figure 2. Locations of the correlation distribution at the output plane. Capital letters in a correlation term refer to colour channels of the reference image; lowercase letters represent those of the input test scene. The dash line denotes the removed terms in the JTPS subtraction procedure.

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

some desired terms in the areas around $(0, 2x_0)$ and $(0, -2x_0)$ in figure 2 that yield sharp peaks for targets.

To avoid the problem of a strong zero-order peak appearing at the centre of the output plane, we chose the joint transform power spectrum (JTPS) subtraction strategy [19]. The strategy is based on equation 3. The removal procedure is stated as follows:

1. Divide the six images at the input scene into two parts, one consisting of three reference images and the other of three test images.
2. Capture the power spectra of these two parts individually, and store them in the computer.

3. Capture the JTPS of the joint input scene and subtract it by the power spectra captured in step 1. Therefore, we obtain the non-zero-order JTPS (NOJTPS).
4. Send the NOJTPS to the SLM and acquire the final correlation distribution at the output plane.

Finally, the correlation distribution without zero-order term at the output plane is given by

$$O(x',y') = \sum_{n=1}^{3}\sum_{m=4}^{6} C_{nm}(x' - (x_m - x_n), y' - (y_m - y_n)) + \sum_{m=1}^{3}\sum_{n=4}^{6} C_{nm}^{*}(x' + (x_m - x_n), y' + (y_m - y_n)) \tag{5}$$

The intensity distribution captured by the CCD camera at the output plane is then

$$I'(x',y') = |0(x',y')|^2 \tag{6}$$

In pattern recognition, it is quite important to detect the targets with different geometric distortion, e.g., rotation and scaling, etc. A particular frequency domain filter, namely MACE [8, 20], is able to include all possible distortion information and to reduce correlation sidelobes for monochromatic images. This task can be accomplished by minimizing the average correlation energy of all training images while keeping the desired correlation peaks at a specified height. Trying not to complicate the derivation procedure, we utilize a compact style denotation in the following paragraph.

Let F and H be the corresponding Fourier transforms of the test image f and the reference image h, respectively. From the preceding presumption, these symbols actually refer to any of the R, G and B channels in the colour images. In order to minimize the average correlation energy for all training images in each channel, we assume that there are N centred training target images spanning the expected distortion-invariant feature database. Therefore, $f_i \otimes h$ and $F_i^* H$ form a Fourier transform pair, where f_i denotes the ith training image with d pixels present at the input plane. Since the intensity distribution (see equation (6)) at the output plane is an even function, the values of the desired correlation peak intensity on both sides are identical. The constrained correlation peaks can be written in matrix vector notation as follows

$$\hat{F}^T \hat{H}^* = [p_1, p_2, \ldots, p_N] = \hat{p} \tag{7}$$

where $\hat{F}$ is a matrix with N column vectors in which each column vector represents the sampled Fourier transform of f_i; similar process for h leads to a column vector $\hat{H}$; the superscript T stands for the matrix transpose operator; $\hat{p}$ is the correlation peak requirement vector of size N with p_i as entries, which can be specified as the same to yield equal correlation peaks in response to all training images.

By applying Parseval's identity, the average cross-correlation energy function can be described as follows:

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

$$E = H^{+}AH \tag{8}$$

where the superscript + denotes the Hermitian symmetric transpose while A is a real-valued diagonal matrix whose diagonal entry is the average of $|F_i(p,q)|^2$ with respect to all training images. By rewriting equations (7) and (8) with real parts and image parts, we obtain

$$F^T H^* = (F_r^T + jF_i^T)(H_r - jH_i) = p = p_r + jp_i \tag{9}$$

$$\begin{aligned} E = H^{+}AH &= (H_r + jH_i)^{+} A (H_r + jH_i) \\ &= H_r^T A H_r + H_i^T A H_i + jH_r^T A H_i - jH_i^T A H_r \end{aligned} \tag{10}$$

where the subscripts r and i denote the real part and image part of the associated symbol, respectively. Taking advantage of Lagrange multipliers, the problem of the optimisation of the average cross-correlation energy given in equation (10) while satisfying the constraints in equation (9) for each channel can be solved. By making use of equations (9) and (10), the Lagrangian can be defined as

$$J = H_r^T A H_r + H_i^T A H_i - 2\lambda_r^T (F_r^T H_r + F_i^T H_i - p_r) - 2\lambda_i^T (F_i^T H_r - F_r^T H_i - p_i) \tag{11}$$

The gradients of J with respect to H_r and H_i are

$$\begin{cases} \nabla_{H_r} J = 2AH_r - 2F_r\lambda_r - 2F_i\lambda_i \\ \nabla_{H_i} J = 2AH_i - 2F_i\lambda_r + 2F_r\lambda_i \end{cases} \tag{12}$$

Setting the gradients to zeros, and solving the simultaneous equations gives

$$\begin{cases} H_r = A^{-1}(F_r\lambda_r + F_i\lambda_i) \\ H_i = A^{-1}(F_i\lambda_r - F_r\lambda_i) \end{cases} \tag{13}$$

Applying the results to H^* leads to

$$\begin{aligned} H^* = H_r - jH_i &= A^{-1}[F_r(\lambda_r + j\lambda_i) - jF_i(\lambda_r + j\lambda_i)] \\ &= A^{-1}(F_r - jF_i)(\lambda_r + j\lambda_i) \\ &= A^{-1}F^*\lambda \end{aligned} \tag{14}$$

Substitute H^* as $A^{-1}F^*\lambda$ in equation (9) and obtain

$$\lambda = (F^T A^{-1} F^*)^{-1} p \tag{15}$$

The optimum solution is obtained by combing equation (14) and (15), that is

$$H = A^{-1}F(F^{+}A^{-1}F)^{-1}p^* \tag{16}$$

The solution in equation (16) is a vector representation in the frequency domain. The final optimum reference function in the input domain with a more formal notation is constructed by

$$h_k(x,y) = \Im^{-1}\{\hat{H}_k(u,v)\} \qquad k = \mathrm{R}, \mathrm{G}, \mathrm{B}, \tag{17}$$

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

where $\Im^{-1}$ denotes the inverse Fourier transform.

For further investigation of the performance, we introduce the peak to correlation energy (PCE) ratio, which is the sharpness of the correlation peak measured at the desired area. It is defined as

$$PCE = \frac{\max_{x,y}\{|C(x,y)|^2\}}{\int\int |C(x,y)|^2 \mathrm{d}x\mathrm{d}y} \tag{18}$$

It is expected that the sharper the correlation profile, the higher the accuracy of target detection.

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

3. Numerical results

Numerical analysis of the NOJTC system was carried out using the Matlab software package. The test image was chosen to be a colourful butterfly photo with 64×64 pixels. For simplicity, 18 colour images are selected for the training set in the subsequent simulation. Figure 3 illustrates the training set images of R, G, B channels. An in-plane rotational distortion range from 0° to 360° was considered. Each image is selected 20° apart in rotation. Therefore, there were 18 training images for each channel to construct the associated reference function. Each training image was assigned a desired correlation peak of 255. To avoid overlapping of the correlation distribution, all input scene images used in the simulation were arranged in proper intervals.

The joint input image is composed of three channels, in the order of red, green and blue, from the top to the bottom. Each channel has one reference image on the left half part and one target image on the right half part (see figure 4). The overall size of the input image is 384×384 pixels.

From the preceding analysis, the peaks of interest at the output plane are located around areas $(0, -2x_0)$ and $(0, 2x_0)$. They represent the coherent addition of the three cross-correlation terms between the target and the input scene of the corresponding RGB color channels. Figure 5 shows the corresponding 3D output plot of a CJTC. Because of the existence of the extra large zero-order term, it is hard to tell where the desired peaks are without use of computer simulation. Since the CCD camera detection ability is unknown here, all peaks in figure 5 are normalized according to the highest peaks to enhance comparison. In practical experiment, the 3D profile captured by the CCD camera is expected to be more complicated. The output has 15 correlation peaks regularly arranged, as is shown in figure 2.

Figure 6 illustrates the corresponding 3D output plot of an NOJTC. In contrast to CJTC, NOJTC provides clear and sharp correlation peaks to detect targets. The sharp correlation peak corresponding to the valid butterfly image is quite distinguishable and indicates that the valid image is detected properly. All peaks in figure 6 are normalized according to the desired peaks to enhance the observation. Besides are two desired peaks, there are some insignificant peaks, which are due to other cross-correlation terms. In this case, these insignificant peaks are within size limits that have little influence on observation of the desired peaks.

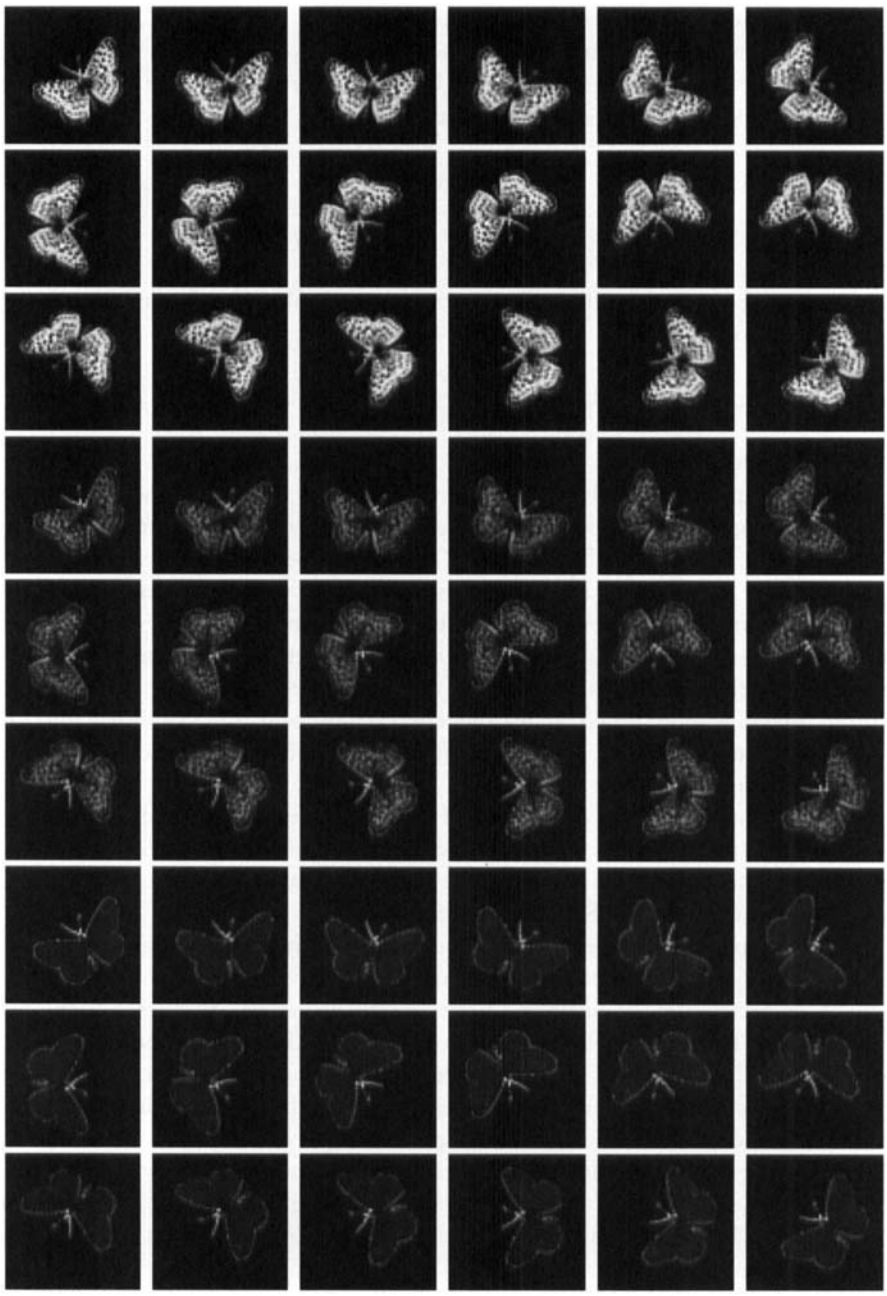

Figure 3. Training image for R, G, B channels.

According to this numerical analysis of output spectra in polychromatic pattern recognition with CJTCs, the zero-order term at the origin, which is composed of six auto-correlation terms, can be about 1102 times higher than the desired term. Moreover, the zero-order content increases with the number of channels at the input plane. Accordingly, by applying extra steps in the JTPS subtraction technique, we can make the recognition capability of JTCs better.

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

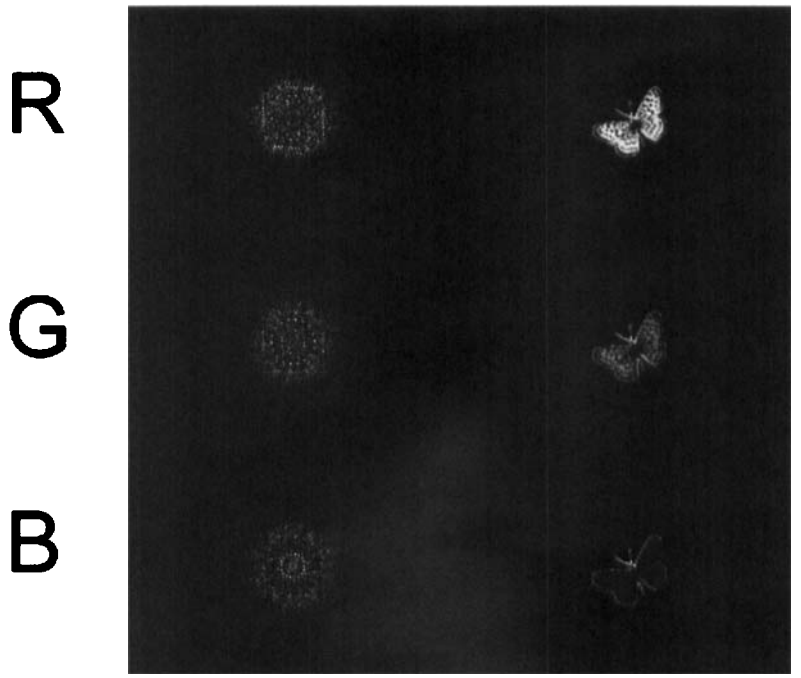

Figure 4. Input plane of a JTC (RGB channels from the top to the bottom) with noise-free test images.

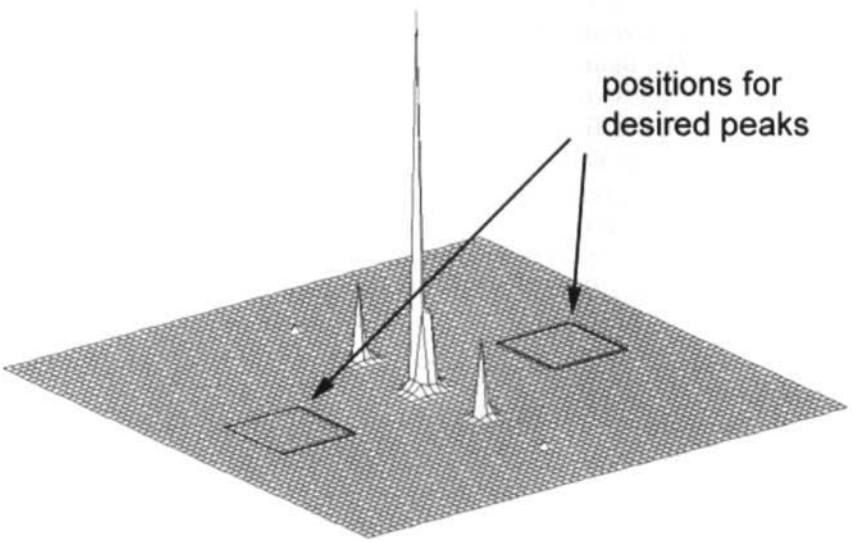

Figure 5. 3D profile for the correlation output of a CJTC for figure 4.

Much research has shown that the performance of a JTC system depends heavily on the number of training set images. The more training set images we use in the procedure to produce the reference image, the greater the recognition

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

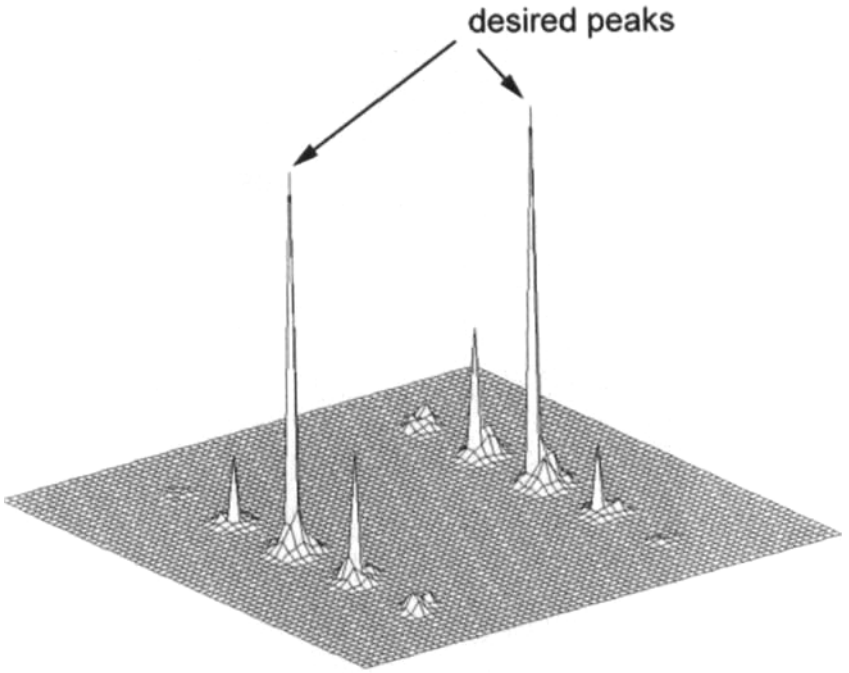

Figure 6. 3D profile for the correlation output of a NOJTC for figure 4.

capacity of the system. However, expanding the number of training set images implies increasing average energy, which results in magnification of sidelobes.

To clarify the behaviour of the monochromatic NOJTC and multi-channel single output NOJTC, figure 7 illustrates the PCE as a function of the number of training images for the two cases. Values along with x-axis are the number of training images, ranging from 10 to 360 training images in increments of 10 images. The tested image is the same as the test image in figure 3. As the number of training set images increases, the constraints increase, thus the ratios decrease due to the increase of the correlation output energy. The figure shows that the multi-channel NOJTC system generates better discrimination performance over the tested training images than does the monochromatic NOJTC.

4. Conclusions

We have introduced a multi-channel polychromatic pattern recognition method with the use of an NOJTC. Computer simulation verified that the proposed technique produces high correlation discrimination. It has been shown that the general algorithm used in monochromatic NOJTC systems can easily be applied to polychromatic cases. We also adopted the PCE to investigate the recognition performance. In the meantime, polychromatic NOJTC also implies superior discriminability over monochromatic JTC in the PCE graph. However, we need more analytical evidence in mathematics to support our hypothesis. The system has improved the performance of a JTC in terms of better pixel utilization, distortion invariance, higher detection efficiency, and the fact that false alarms can be avoided. The proposed system also offers parallel multi-channel and single-

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

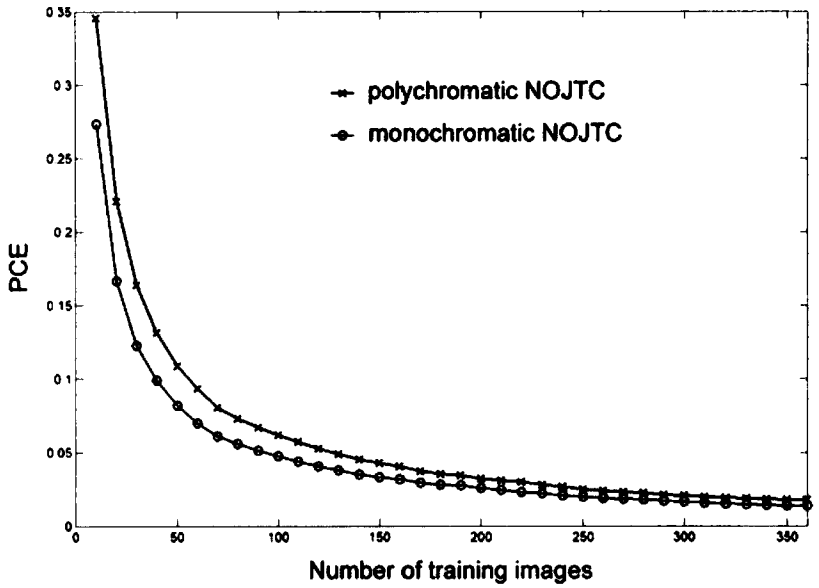

Figure 7. PCE as a function of the number of training images for different correlators.

output processing capability, by which the system can detect simultaneously targets of multiple wavelengths.

However, there are some drawbacks we have to overcome. As we mentioned in the preceding section, those insignificant peaks distributed at the output plane were ignorable in the tests carried out so far. However, in other cases, such as test images with noise, a very large number of training set images, and distorted test images, these unwanted peaks will be larger and will make observation of the desired peaks more difficult. Furthermore, the implementation of multi-channel systems indicates a requirement for a larger SLM, and thus higher cost. Further investigation of system improvement is proceeding.

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011

Acknowledgments

This research has been supported by the National Science Council, Taiwan under Grants NSC 90-2215-E-155-006 and 91-2215-E-155-001.

References

[1] Weaver, C. S., and Goodman, J. W., 1966, *Appl. Opt.*, **5,** 1248.
[2] VanderLugt, A., 1964, *IEEE Trans. Inf. Theory*, **IT 10,** 139.
[3] Lu, G., Zhang, Z., Wu, S., and Yu, F. T. S., 1997, *Appl. Opt.*, **36,** 470.
[4] Yu, F. T. S., Lu, G., Lu, M., and Zhao, D., 1995, *Appl. Opt.*, **34,** 1386.
[5] Jutamulia, S., Storti, G. M., Gregory, D. A., and Kirsch, J. C., 1991, *Appl. Opt.*, **30,** 4173.
[6] Jutamulia, S., and Gregory, D. A., 1998, *Opt. Eng.*, **37**(1), 49.
[7] Alam, M. S., Chen, X., and Karim, M. A., 1997, *Appl. Opt.*, **36,** 7422.
[8] Chen, C., and Fang, J., 2000, *Opt. Commun.*, **178,** 315.
[9] Chen, C., Fang, J., and Yin, S., 2000, *Microw. Opt. Technol. Lett.*, **26,** 312.
[10] Chen, C., and Fang, J., 2001, *J. Mod. Opt.*, **48,** 1329.

[11] WARDE, C., CAULFIELD, H. J., YU, F. T. S., and LUDMAN, J. E., 1984, *Opt. Commun.*, **49,** 241.
[12] LUDMAN, J. E., JAVIDI, B., YU, F. T. S., CAULFIELD, H. J., WARDE, C., and EFRON, U., 1984, *SPIE Proc.*, **465,** 143.
[13] YU, F. T. S., YANG, Z., and PAN, K., 1994, *Appl. Opt.*, **33,** 2170.
[14] FENG, J. H., CHIN, G. F., WU, M. X., YAN, S. H., and YAN, Y. B., 1995, *Opt. Lett.*, **20,** 82.
[15] KERYER, G., and TOCNAYE, J. L. B., 1995, *Opt. Commun.*, **118,** 102.
[16] DEUTSCH, M., GARCIA, J., and MENDLOVIC, D., 1996, *Appl. Opt.*, **35,** 6976.
[17] ALAM, M. S., and WAI, C. N., 2001, *Opt. Eng.*, **40,** 2407.
[18] YU, F. T. S., and LU, X. J., 1984, *Opt. Commun.*, **52,** 10.
[19] LI, C., YIN, S., and YU, F. T. S., 1998, *Opt. Eng.*, **37,** 58.
[20] MAHALANOBIS, A., VIJAYA KUMAR, B. V. K., and CASASENT, D., 1987, *Appl. Opt.*, **26,** 3633.

Downloaded by [Georgia Tech Library] at 16:38 10 December 2011