2. G.J. Cowle and D.Yu. Stepanov, Multiple wavelength generation with Brillouin/Erbium fibre lasers, IEEE Photon Technol Lett, 8 (1996), 1465–1467.
3. M.K. Abd-Rahman, M.K. Abdullah, and H. Ahmad, Multiwavelength, bidirectional operation of twin-cavity Brillouin/erbium fibre laser, Opt Comm, 181 (2000), 135–139.

CONSTRAINED OPTIMIZATION FOR COLOR PATTERN RECOGNITION WITH A NONZERO ORDER JOINT TRANSFORM CORRELATOR

Chih-Sung Wu, Chulung Chen, and Jian-Shuen Fang
Department of Electrical Engineering
Yuan Ze University
Taoyuan 320, Taiwan

Received 21 November 2001

ABSTRACT: *In this article, we apply a novel method on color pattern recognition with the nonzero order joint transform correlator. By the special arrangement of separated RGB channels of objects and reference images at the input plane, we can predict the location of correlation outputs. Furthermore, by the use of power-spectrum subtraction strategy, the zero-order part can be removed. The test pattern is a frog. The desired correlation peak is found to be quite distinctive and sharp.* © 2002 Wiley Periodicals, Inc. Microwave Opt Technol Lett 33: 385–388, 2002; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/mop.10330

Key words: *joint transform correlator; color pattern recognition*

1. INTRODUCTION

Optical correlation for pattern recognition has many advantages over digital images, e.g., real-time and parallel processing capabilities. The most frequently used methods have been the VanderLugt optical correlator [1] and the joint transform correlator (JTC) proposed by Weaver and Goodman [2]. However, the VanderLugt setup requires an accurate alignment with the optical axis. On the other hand, in the JTC setup the reference and the target are displayed side by side at the input plane, reducing the alignment requirement. The correlation operations performed by these two correlators are shift-invariant. The conventional JTCs suffer from large correlation sidelobes, large autocorrelation bandwidth, low light efficiency, and poor discrimination ability. A number of efforts have been made to improve JTC performance by pioneers. To alleviate this imperfection, Zhong et al. [3] presented a binary joint transform correlator with differential processing in the joint transform power spectrum (JTPS) that can eliminate the dc component and sharpen the correlation peak but with lower complexity of computation. Yu et al. [4] realized the nonzero order joint transform correlators (NOJTCs) with a phase-shifting technique [5]. But owing to the heavy dc content of the input target, it also produces relatively broad correlation profiles, which hinder its application to multitarget detection. To improve this, Chen et al. [6, 7] designed a reference function that can produce sharp correlation peaks for distortion invariant pattern recognition.

In addition, for color pattern recognition the classical multichannel system was implemented for the JTC by spatial separation of the spectra for each color channel in the Fourier plane with a dispersive element located at the input plane. In fact, it could be possible to find an arrangement of objects for all color channels at the input plane that would lead to the coherent addition of the three cross-correlation terms between the corresponding RGB color channels of the target and the input scene. To achieve this, Deutsch et al. [8] introduced a multichannel single-output JTC. However, their concept led to massive zero-order parts composed of several autocorrelation terms that suffer from poor detection efficiency. In the present article, we introduce a constrained optimization method for performing color pattern recognition with a NOJTC.

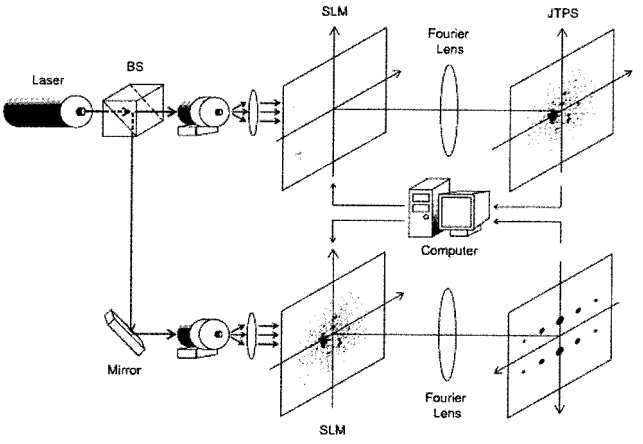

Figure 1 An experimental system of the JTC

2. ANALYSIS

In the beginning, we review the basic theory of the JTC, which has been described in detail elsewhere [9]. The proposed optical implementation of a JTC with two spatial light modulators (SLMs) is shown in Figure 1. A He-Ne laser illuminates the SLM, which records the joint images at the input plane. First, the Fourier lens performs a Fourier transform. Then a CCD camera can receive the joint transform power spectrum (JTPS), which are transferred to an electronically addressed SLM for the inverse Fourier transform. The JTPS can be modified by the subtraction of the test-image-only power spectrum and the reference-image-only power spectrum. The subtraction process can be done optically or using a computer. The subtracted JTPS image is then transferred through the computer to the SLM for the inverse Fourier transform. Finally, the correlation output can be obtained at the back focal plane of the second lens.

Let us assume the input object to the JTC is given by:

$$t(x, y) = \sum_{\alpha=1}^{n} g_\alpha(x - a_\alpha, y - b_\alpha), \tag{1}$$

where (a_α, b_α) is the position of the object $g_\alpha(x, y)$ at the input plane. For simplicity, $n = 6$. Since the reference function and test image are separated into R, G, B channels, the corresponding JTPS can be shown as:

$$\begin{aligned} I(u, v) &= \left| \sum_{\alpha=1}^{6} G_\alpha(u, v)\exp[-i(ua_\alpha + vb_\alpha)] \right|^2 \\ &= \sum_{\alpha=1}^{6} |G_\alpha(u, v)|^2 + \sum_{\alpha=1}^{6} \sum_{\substack{\beta=1 \\ \beta\neq\alpha}}^{6} G_\alpha(u, v)G_\beta^*(u, v) \\ &\quad \times \exp[iu(a_\beta - a_\alpha) + iv(b_\beta - b_\alpha)], \end{aligned} \tag{2}$$

where the asterisk denotes the complex conjugate. $G_\alpha(u, v)$ and $G_\beta(u, v)$ are the Fourier spectra of the object $g_\alpha(x, y)$ and $g_\beta(x, y)$, while (u, v) are the spatial coordinate system of the joint-spectrum plane.

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

$$O(x', y') = \sum_{\alpha=1}^{6} C_{\alpha\alpha}(x', y') + \sum_{\alpha=1}^{6} \sum_{\substack{\beta=1 \\ \beta\neq\alpha}}^{6} C_{\alpha\beta}[x' - (a_\beta - a_\alpha), y' - (b_\beta - b_\alpha)], \quad (3)$$

The correlation $C_{\alpha\beta}$ between the objects g_α and g_β is defined as:

$$C_{\alpha\beta}(x', y') = \iint g_\alpha^*(x, y)\, g_\beta(x + x', y + y')\, dxdy. \quad (4)$$

We further note that the first term on the right-hand side in Eq. 3 represents the sum of all the autocorrelation distributions of the input objects; it appears at the origin of the output plane. The second term is the cross-correlation term between any pair of input objects. The distribution of cross-correlation terms at the output plane depends on the locations at the input plane. Hence, overlapping of these cross-correlation distribution may occur. In order to ensure nonoverlapping cross-correlation distribution at the output plane, we require:

$$D_X \geq S_X \quad \text{and} \quad D_Y \geq S_Y, \quad (5)$$

where D_X is the horizontal separation between the color channels of the input scene and D_Y is the vertical separation between the input color channels and the target color channels. S_X and S_Y are the horizontal and vertical sizes of the color target and the color input scene, respectively.

To solve the problem of a strong zero-order peak appearing at the output plane that might make the desired correlation signal hard to detect, the joint transform power spectrum (JTPS) subtraction strategy [10] can be used. The zero-order spectrum removal procedure will be done as follows:

1. Precalculate and store the power spectra of the references in the computer.
2. Capture the power spectra of the input scenes and store it in the computer.
3. In the joint transform operation cycle, send the captured JTPS to the computer for the zero-order removal.
4. Subtract the recorded JTPS by the power spectra of the reference and the input scene, and we have:

$$I(u, v) = \sum_{\alpha=1}^{3} \sum_{\beta=4}^{6} G_\alpha(u, v) G_\beta^*(u, v) \times \exp[iu(a_\beta - a_\alpha) + iv(b_\beta - b_\alpha)] + \sum_{\beta=1}^{3} \sum_{\alpha=4}^{6} G_\alpha(u, v) G_\beta^*(u, v) \times \exp[-iu(a_\beta - a_\alpha) + iv(b_\beta - b_\alpha)] \quad (6)$$

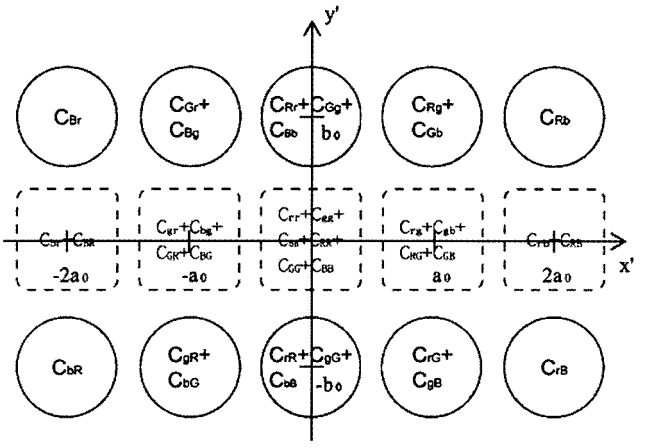

Figure 2 Locations of the correlation terms at the output plane. Capital letters in a correlation term refer to color channels of the reference image; lowercase letters represent those of the input test scene

5. Send the nonzero order JTPS to the SLM and acquire the final correlation output:

$$O(x', y') = \sum_{\alpha=1}^{3} \sum_{\beta=4}^{6} [x' - (a_\beta - a_\alpha), y' - (b_\beta - b_\alpha)] + \sum_{\beta=1}^{3} \sum_{\alpha=4}^{6} [x' + (a_\beta - a_\alpha), y' + (b_\beta - b_\alpha)]. \quad (7)$$

As the zero-order terms are removed, on-axis correlation distribution can be avoided. The intensity distribution recorded by the CCD detector at the output plane comes out to be:

$$I'(x', y') = \sum_{\alpha=1}^{3} \sum_{\beta=4}^{6} |C_{\alpha\beta}(x', y')|^2 \cdot \delta(x' - (a_\beta - a_\alpha), y' - (b_\beta - b_\alpha)) + \sum_{\beta=1}^{3} \sum_{\alpha=4}^{6} |C_{\alpha\beta}(-x', -y')|^2 \cdot \delta(x' + (a_\beta - a_\alpha), y' + (b_\beta - b_\alpha)). \quad (8)$$

The output plane of the JTC is shown in Figure 2, the subscripts in capital letters and lowercase letters represent the input reference and test image, respectively. There are 18 correlation terms grouped in 10 correlation locations, each location consisting of the coherent addition of a few overlapping cross-correlation terms.

The joint input image is composed of two parts: the upper part contains RGB channels from the synthesized reference image and the lower part contains the input test scene. We assume that each reference channel and the corresponding input test channel separated by a distance $2b_0$ along the y axis are situated simultaneously at two different positions in the input array of the JTC, where $k = R,G,B$. The digital images are sent to one SLM at the input plane of the optical setup. Let H_k and F_k be the corresponding Fourier transforms of $h_k(x, y)$ and $f_k(x, y)$, respectively.

We minimize the average correlation energy for all training images in each channel. We assume that there are N centered training target images spanning the desired distortion-invariant feature range. Let f_{kn} denote the n^{th} training image with d pixels

present at the input plane. $h_k \otimes f_{kn}$ and $H^*_k F_{kn}$ form a Fourier transform pair, i.e.:

$$X_{kn}(x, y) = h_k \otimes f_{kn} \leftrightarrow H^*_k F_{kn}(u, v), \tag{9}$$

where $X_{kn}(x, y)$ is the correlation output when the n^{th} training image is the input. The value of the desired correlation peak for each training image is given by:

$$X_{kn}(0, 0) = \int\int H^*_k(u, v) F_{kn}(u, v)\, dudv, \quad n = 1, 2, \ldots N. \tag{10}$$

From Parseval's theorem, which states that the total power in the spatial domain is the same as the total power in the Fourier domain, the cross-correlation energy in either half-output plane for each training image would be:

$$\int\int |X_{kn}(x, y)|^2\, dxdy = \int\int H_k(u, v)|F_{kn}(u, v)|^2 H^*_k(u, v)\, dudv. \tag{11}$$

Equation 10 can be rewritten with matrix multiplication as:

$$\underline{\underline{F}}^T_k \underline{H}^*_{kc} = [X_{k1}(0, 0) \quad X_{k2}(0, 0) \quad \cdots \quad X_{kN}(0, 0)]^T = \underline{P}_k, \tag{12}$$

where $\underline{\underline{F}}_k$ is a $\mathbf{d} \times \mathbf{N}$ data matrix whose n^{th} column is obtained by lexicographically (pixel by pixel) scanning $F_{kn}(u, v)$; similar scanning of $H_k(u, v)$; leads to a column vector $\underline{H}_{kc}$; the superscript T stands for the transpose operator of a matrix and P_k is the correlation peak requirement vector of size N with $X_{nk}(0, 0)$ as entries, which can be specified by the user as the same constant to yield equal correlation peaks in response to all training images. In the same way, Eq. 11 leads to the average cross-correlation energy function given by:

$$E_{ave} = \underline{H}^T_{kc} \underline{\underline{D}}_k \underline{H}^*_{kc}, \tag{13}$$

where $\underline{\underline{D}}_k$ is a real-valued diagonal matrix of size $\boldsymbol{d} \times \boldsymbol{d}$; $\boldsymbol{d}$ is the total pixel number. Each diagonal entry is given by calculating $(1/N)\Sigma_{n=1}^{N}|F_{kn}(u, v)|^2$.

With the help of Lagrange Multipliers, the solution to the optimization problem for each channel is:

$$\underline{H}_{kc} = \underline{\underline{D}}^{-1}_k \underline{\underline{F}}_k [(\underline{\underline{F}}^*_k)^T \underline{\underline{D}}^{-1}_k \underline{\underline{F}}_k]^{-1} \underline{P}^*_k, \quad k = R, G, B. \tag{14}$$

The solution in Eq. 14 is a vector representation in the frequency domain. As the column vector $\underline{H}_{kc}$ is rearranged as a square matrix H_k, $h_k(x, y)$ can be obtained by inverse Fourier transforming on $H_k(u, v)$. The final optimum reference function in the input domain is constructed by:

$$h_k(x, y) = \mathcal{F}^{-1}\{H_k(u, v)\}, \quad k = R, G, B, \tag{15}$$

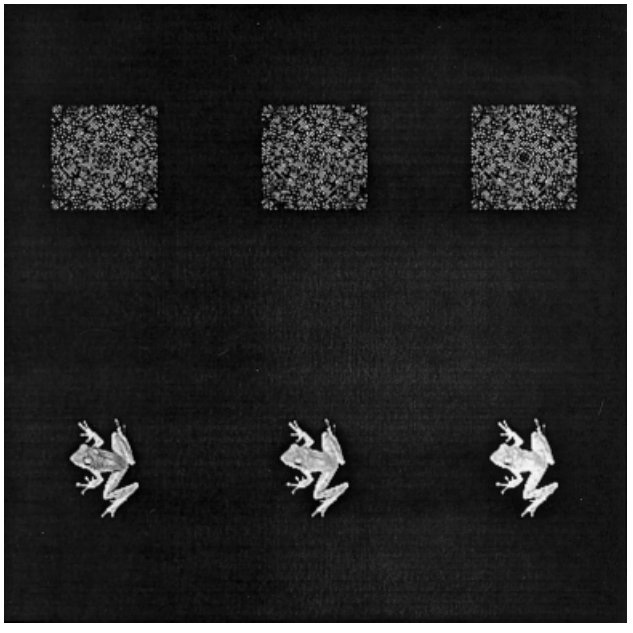

Figure 3 Input image of a JTC (RGB channels from the right to the left)

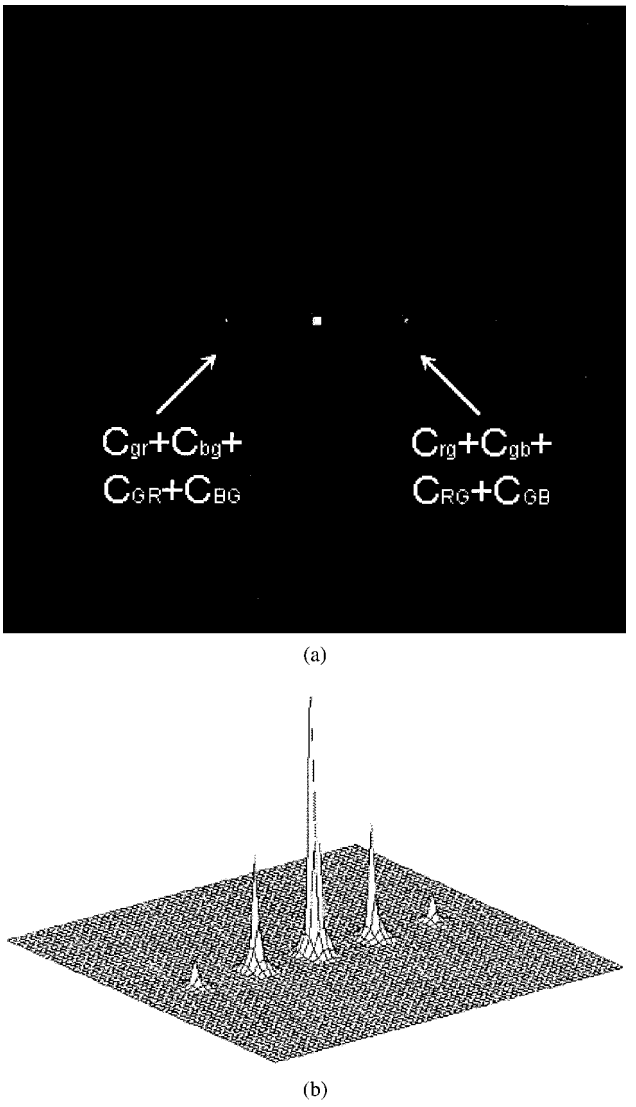

Figure 4 (a) The conventional JTC output. (b) A three-dimensional plot of the correlation output for the CJTC

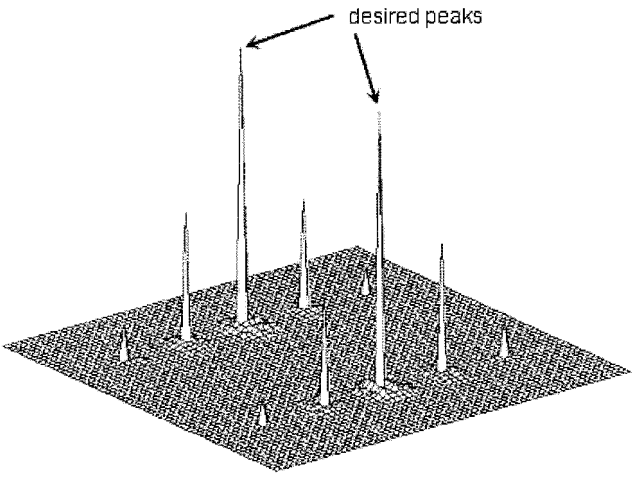

Figure 5 A three-dimensional plot of the correlation output for the NOJTC

where $\mathcal{T}$ denotes the inverse Fourier transform.

3. NUMERICAL RESULTS

The test image for the study is chosen to be a 6×6 pixel color image from a frog. The joint input image is composed of two parts: the upper part contains the synthesized reference images for three color channels and the lower part includes the input test scenes (see Fig. 3). The dimension of the input space is 384×384. A rotational distortion range from 0° to 360° is considered. Each image is selected 30° apart in rotation. In this case, D_X, D_Y, S_X, and S_Y in Eq. 5 are 64, 128, 64, and 64 pixels, respectively. Hence, there are 12 training set color images to construct the reference function. Both parts are in the order of red, green, and blue channel from right to left.

The output of a conventional JTC (CJTC) is shown in Figure 4(a) and the three-dimensional profile is shown in Figure 4(b). From the preceding analysis, the correlation locations of interest are the areas around $(0,-2b_0)$ and $(0,2b_0)$ and they represent the coherent addition of the three cross-correlation terms between the corresponding RGB color channels of the target and the input scene. The correlation peaks are symmetric with respect to the main axes origin, as expected from Figure 2. Apparently, the zero-order term and the other two terms on the x axis are extremely strong and the desired peaks around $(0,-2b_0)$ and $(0,2b_0)$ cannot be observed. On the other hand, for the NOJTC the three-dimensional plot of the correlation output is given in Figure 5, in which the central zero-order part is removed and the desired correlation peak is quite distinctive and sharp. The peaks in the graph have been normalized for clarity. Besides the two desired peaks, one can see some insignificant peaks, which are due to other cross-correlation terms. According to our numerical results in this case, the zero-order term at the origin, which is composed of six autocorrelation terms, is 2,918 times greater than the peak we want in this experiment. The amount of zero-order terms increases by the employment of more channels at the input plane.

4. CONCLUSION

A procedure for color pattern recognition with a nonzero-order joint transform correlator has been demonstrated. It has the advantages of better pixel utilization, higher detection efficiency, and the fact that false alarms due to the inner-object correlation can be avoided. What it costs is the required additional steps in the JTPS subtraction procedure.

This proposed method was demonstrated by use of the RGB color channel separation, but it may be adapted to any multichannel case. By the removal of zero-order terms the distance between the input images can be shorter. The implementation of color pattern recognition with NOJTC could be easier and the system could be smaller since the arrangement of input images could be closer. Further investigation of the removal of unwanted correlation terms is proceeding. Also, we are working on test images with distortion and noise and some exhilarating results have been obtained.

REFERENCES

1. A. VanderLugt, Signal detection by complex spatial filtering, IEEE Trans Inform Theory, IT-10 (1964), 139–146.
2. C.S. Weaver and J.W. Goodman, A technique for optically convolving two functions, Appl Opt 5 (1966), 1248–1249.
3. S. Zhong, J. Jiang, S. Liu, and C. Li, Binary joint transform correlator based on differential processing of the joint transform power spectrum, Appl Opt 36 (1997), 1776–1780.
4. F.T.S. Yu, C. Lu, M. Lu, and D. Zhao, Application of position encoding to a complex joint transform correlator, Appl. Opt. 34 (1995), 1386–1388.
5. G. Lu, Z. Zhang, S. Wu, and F.T.S. Yu, Implementation of a nonzero-order joint-transform correlator by use of phase shifting techniques, Appl Opt 36 (1997), 470–483.
6. C. Chen and J. Fang, Cross-correlation peak optimization on joint transform correlators, Opt Commun 178 (2000), 315–322.
7. C. Chen, J. Fang, and S. Yin, Optimized synthetic aperture radar image detection with nonzero-order joint transform correlators, Microwave Opt Technol Lett 26 (2000), 312–316.
8. M. Deutsch, J. García, and D. Mendlovic, Multichannel single-output color pattern recognition by use of a joint-transform correlator, Appl Opt 35 (1996), 6976–6982.
9. F.T.S Yu and X.J. Lu, A real-time programmable joint transform correlator, Opt Commun 52 (1984), 10–16.
10. C. Li, S. Yin, and F.T.S. Yu, Nonzero-order joint transform correlator, Opt Eng 37 (1), (1998), 58–65.