Multi-channel single-output color pattern recognition with distortion tolerant capability

Chih-Chieh Ting, Chung-Cheng Lee, Chulung Chen
Department of Electrical Engineering, Yuan Ze University, Taiwan, R.O.C.

ABSTRACT

We present a novel color pattern recognition technology based on non-zero order joint transform correlator (NOJTC) system in this paper. In this method, each of the color target image is transformed to a grayscale image by using encoding technique. We also use minimum average correlation energy (MACE) approach to design an optimized synthetic reference function. When the input plane is gray-scaled and monochromatic, the function can be displayed in the liquid crystal spatial light modulator (LCSLM) to achieve real-time operation. Furthermore, we apply a joint transform power spectrum (JTPS) subtraction method to remove the zero-order terms and the desired peaks can be easily detected.

Keywords: color pattern recognition; non-zero order joint transform correlator; encoding technique; liquid crystal spatial light modulator

1. INTRODUCTION

In 1966, Weaver and Goodman [1] proposed the joint transform correlators for pattern recognition. Before that, most of the optical pattern recognition systems were based on the 4-f VanderLugt correlator (VLC) [2]. The Optical pattern recognition can be performed easily by a lens to perform a Fourier transform in parallel and real-time, which is carried out based on the correlation. The basic distinction between these techniques is that the VLC depends on Fourier domain matched filtering, whereas JTC depends on spatial domain filtering. Compared with VLC, JTC does not have the stringent filter-alignment problem. Therefore, it is more suitable for real-time processing and is more robust in term of environment disturbances. Because of this, JTC has become a hot research field in recent years. According to the principle of conventional JTC (CJTC), the joint transform power spectrum (JTPS) has a spatially varying average which leads to a large zero order term. The zero order almost overshadows the desired cross-correlation peaks, and often saturates the output detector. Thus, numerous efforts have been to improve the performance of JTC [3-5]. The joint transform power spectrum subtraction strategy was utilized [6] to remove all unneeded peaks at the output plane in this paper. Recently, in order to reduce the correlation sidelobes as low as possible for a rotated image, Chen et al. [7] adopted the minimum average correlation energy (MACE) method [8] based on the Lagrange multipliers to design a
reference function that can yield sharp correlation peaks.

In this paper, we propose an image encoding technique for performing color pattern recognition. The concept of image encoding led to the evolution of optical pattern recognition. In order to obtain a novel encoding pattern, we utilize recombination of red, green, and blue grayscale images in the input image. Besides, the encoding pattern will be synthesized into the reference function by using MACE filter. On the other hand, the multi-channel single-output NOJTC system has many advantages over other conventional systems especially with distortion tolerant capability. The technique also can result sharp correlation peaks.

2. THEORY

In this paper, we propose a novel space arrangement method with image encoding technique for color image. In the method each colorful target image is transformed to a grayscale image of red, green, and blue color components. Then, these grayscale images are recombined in a new encoding image, as shown in Fig. 1, which is displayed on a single, monochromatic, input plane. The process is listed as follows:

1. Let \( r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}, \quad g = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix}, \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \) \hspace{1cm} (1)

where \( r, g, \) and \( b \) represent the matrices of red, green, and blue images, respectively; \( n \) denotes the number of the maximum row vector. Let \( r_i, g_i, \) and \( b_i \) indicate the \( i \)-th component of the \( r, g, \) and \( b, \) respectively.

2. Insert one by one respectively each \( r, g, \) and \( b \) row vector to a matrix.

\[
\begin{bmatrix}
    r_1 \\
    g_1 \\
    b_1 \\
    r_2 \\
    g_2 \\
    b_2 \\
    \vdots \\
    r_n \\
    g_n \\
    b_n
\end{bmatrix} = E
\] \hspace{1cm} (2)

Therefore, the new matrix \( E \) is an encoding image, which includes all color elements. This method is different from multi-channel single-output JTC. The encoding image can represent the combination of red, green, and blue grayscale images at the input plane. The system could be smaller since the channel number of input image could be reduced to a single channel.

The NOJTC system with interlaced scanned color components contains two liquid crystal spatial light modulators (LCSLMs) and three CCD cameras. The schematic diagram of our system is shown in Fig. 2. The computer transfers grayscale images of red, green, and blue color components of both the reference image and the target image in the
input plane through LCSLM 1. Three separate CCD cameras are used to capture the following: colorful butterfly image information, the joint transform power spectrum, and the distribution of correlation output. A laser beam illuminates the joint input image of the LCSLM 1 at the joint input image, which is exactly at the front focal plane of the Fourier lens at first. Afterward, a CCD 2 locating at the back focal plane of the first Fourier lens can capture the joint transform power spectrum. A computer records the digital signal of the CCD2, uses the JTPS subtraction strategy to remove the zero order term, and displays the JTPS on LCSLM 2. Finally, we can get the correlation between the target and reference patterns at the back focal plane of the second Fourier transform lens. We can obtain this correlation output plane without zero order diffraction. The LCSLM 1 can display the grayscale images of color components for the reference image and the encoding image, which are denoted by \( r(x, y) \) and \( e(x, y) \), respectively. Therefore, the input of the LCSLM 1 can be expressed as

\[
f(x, y) = r(x + \alpha, y) + e(x - \alpha, y),
\]

(3)

where \((\alpha, 0)\) and \((-\alpha, 0)\) are the positions of the input objects.

A collimated coherent beam illuminates the LCSLM 1, and it pass through the first Fourier lens. The Fourier lens performing the Fourier transform of \( f(x, y) \) is given by

\[
F(f_x, f_y) = R(f_x, f_y) \exp\left[i(2\pi f_x \alpha)\right] + E(f_x, f_y) \exp\left[i(-2\pi f_x \alpha)\right],
\]

(4)

where \( R(f_x, f_y) \), \( E(f_x, f_y) \) are the Fourier transforms of \( r(x, y) \), \( e(x, y) \), respectively; \( f_x \) and \( f_y \) are mutually independent frequency domain variables.

Thus, the intensity distribution corresponds to Eq. (2) is called the JTPS. The JTPS which is obtained by CCD 2 can be expressed as

\[
I(f_x, f_y) = |F(f_x, f_y)|^2 = |R(f_x, f_y)|^2 + |E(f_x, f_y)|^2 + R^*(f_x, f_y)E(f_x, f_y)\exp(-i2\pi f_x(2\alpha)) + R(f_x, f_y)E^*(f_x, f_y)\exp(i2\pi f_x(2\alpha)).
\]

(5)

We can put the JTPS on the LCSLM 2 by the computer. After the second Fourier Lens, the JTPS can achieve an inverse Fourier transform and yield the correlation output. It can be detected by a CCD 3. The output is

\[
a_1(x, y) = r(-x, -y) \ast r(-x, -y) + e(-x, -y) \ast e(-x, -y) + c(-x, -y) \otimes \delta(x + 2\alpha, y) + c^*(x, y) \otimes \delta(x - 2\alpha, y),
\]

(6)

where \( c(x, y) = r(x, y) \ast e(x, y) \), which is the complex cross-correlation between \( r(x, y) \) and \( e(x, y) \); \( \ast \) and \( \otimes \) denote the correlation and convolution operations, respectively.
Obviously, the distribution of correlation output depends on the locations of objects at the input plane. Among these cross-correlation terms are some desired parts in the areas around \((2\alpha,0)\) and \((-2\alpha,0)\), as shown in Fig. 3, which are expected to yield sharp peaks for targets. However, the strong zero order peak almost overshadows the desired cross-correlation peaks and is hard to detect at the output plane.

Hence, the unneeded peak can be removed by the JTPS subtraction method [9]. The resultant power spectrum is expressed as follow:

\[
I_{d}(f_x, f_y) = R'(f_x, f_y)E(f_x, f_y)\exp(-i2\pi f_x(2\alpha)) + R(f_x, f_y)E'(f_x, f_y)\exp(i2\pi f_x(2\alpha)).
\] (7)

Finally, the field distribution without auto-correlation term at the output correlation plane will be

\[
o_{xy}(x, y) = c(-x, -y) \otimes \delta(x + 2\alpha, y) + c^\ast(x, y) \otimes \delta(x - 2\alpha, y).
\] (8)

Besides, we can minimize the average correlation energy for all training images to achieve optimal discrimination ability. We assume that there are \(N\) centered training images spanning the investigated distortion-invariant feature range. Let \(e_k\) indicate the \(k\)-th training image with \(z\) pixels present at the input plane. From the preceding understanding, \(E_e\) and \(R\) be the responding Fourier transforms of the encoding image \(e_k\) and the reference image \(r\), respectively. Since the basic correlation property states that \(r \ast e_k\) and \(R E_e\) forms a Fourier transform pair as follows:

\[
c_k(x, y) = r \ast e_k \Leftrightarrow R E_e(f_x, f_y),
\] (9)

where \(c_k(x, y)\) is the correlation output when the \(k\)-th training image is input.

The desired correlation peak value for each training image is given by

\[
c_k(0, 0) = \int \int R'(f_x, f_y)E_k(f_x, f_y)df_xdf_y, \quad k = 1, 2, \cdots, N.
\] (10)

The above equation can be expressed by matrix multiplication as follows:

\[
E^T R_e = \begin{bmatrix} c_1(0,0) & c_2(0,0) & \cdots & c_N(0,0) \end{bmatrix} = P.
\] (11)

where \(E\) is a \(z \times N\) data matrix is obtained by scanning \(E_k(f_x, f_y)\) from top to bottom and from left to right gradually; similar scanning of \(R(f_x, f_y)\) leads to a column vector \(R_e\); \(P\) is the correlation peak requirement vector.

198 Proc. of SPIE Vol. 5642
of size $N$ with $c_k(0,0)$ as entries, which can be determined by the user as the same constant to yield equal correlation peaks in response to all training images.

This states that the total power in the spatial domain is the same as the total power in the Fourier domain by Parseval’s theorem. In either half-output plane, the cross-correlation energy for each training image would be

$$
\iint |k_x(x,y)|^2 dx dy = \iint R^*(f_x, f_y) |E_x(f_x, f_y)|^2 R(f_x, f_y) df_x df_y.
$$

(12)

We minimize the average correlation energy for all training images, and constrain the correlation peak to be a desired value $P$. By using the Lagrange multipliers method, minimizing the average energy in Eq. (12) under the constraints in Eq. (10) yields the optimum solution:

$$
R_c = D^{-1} E \left[ (E^*)^T D^{-1} E \right]^{-1} P^*,
$$

(13)

where $D$ is a real-valued diagonal matrix of size $z \times z$. Each diagonal entry is given by calculating $\sum_{k=1}^N |E_k(f_x, f_y)|^2 / N$ and then by scanning the average lexicographically pixel by pixel. The solution in Eq. (13) is a vector representation in the frequency domain. As the column vector $R_c$ is rearranged as a square matrix $R(f_x, f_y)$. Finally, $r(x, y)$ can be obtained by inverse Fourier transform of $R(f_x, f_y)$.

3. RESULTS

In this work, computer simulations are performed to investigate the correlation performance of the NOJTC using a 2-D fast Fourier transform. The correlation output results are plotted using a 3-D plotting. We chose the colorful butterfly image with $64 \times 64$ pixels in the beginning. Furthermore, we utilize the combination of red, green, and blue grayscale images to form an encoding image with $192 \times 64$ pixels. The multi-channel single-output NOJTC system with encoding technique was used to yield correlation output for each of the encoding images. The training images set contains 61 color images with different angles of in-plane rotation, and the rotational distortion range from $-60^\circ$ to $60^\circ$ is considered (in steps of $2^\circ$).

Therefore, we synthesize the reference image from training images set with MACE filter. In order to test the validity of the proposed configuration, we use the Matlab software package to simulate the execution of the proposed system. Compared with monochromatic, polychromatic JTC has large input image plane. However, the encoding image is composed of red, green, and blue grayscale images. The size of the input scene is $512 \times 512$ pixels. The joint input image is shown in the Fig. 4, in which the right plane contains the encoding image, while the left plane includes the
reference image. We have used JTPS subtraction method as in previous statement. Finally, we obtained the desired correlation peaks in the output plane, as shown in Fig. 5. From the three-dimension output profile of polychromatic NOJTC result, we can see that the desired sharp correlation peaks can be seen clearly.

4. CONCLUSIONS

To better utilize the large space-bandwidth product of the JTC input plane and to use the parallel nature of optics, many researchers have proposed a JTC setup capable of performing several elementary correlations simultaneously. In this paper, we have introduced a new type of NOJTC based on the input image of encoding method. Compared to multi-channel JTC in the joint input plane, the proposed multi-channel single-output NOJTC of image encoding could be easier and the system could be smaller since the channel of input image could be reduced to a single-channel. For real-time processing, the joint input plane is displayed on the LCSLMs. Therefore, the simulation results of the method and the overall performance are good on recognition ability in our system.

Figure 1: The image encoding process.
Figure 2: The NOJTC system with interlaced scanned color components.

Figure 3: Locations of the auto-correlation and cross-correlation terms at the output plane.
Figure 4: The joint input plane of polychromatic NOJTC.

Figure 5: The 3D output profile of polychromatic NOJTC.
ACKNOWLEDGMENT

The research was supported in part by National Science Council, R. O. C. under Grant NSC 93-2215-E-155-001.

REFERENCES