Comparative analysis between different font types and letter styles using a nonlinear invariant digital correlation

Ángel Coronel-Beltrán, and Josue Álvarez-Borrego

Abstract

We present, in this paper, a comparative analysis of the letters in Times New Roman (TNR), Courier New (CN) and Arial (Ar) font types in plain and italic style and the effects of five foreground/background color combinations using an invariant digital correlation system with a nonlinear filter with $k=0.3$. The evaluation of the output plane with this filter is given by the peak-to-correlation energy (PCE) metric. The results show that the letters in TNR font have a better mean PCE value when compared with the CN and Ar fonts. This result is in agreement with some studies on text legibility and for readability where the reaction time (RT) of some participant individuals reading a text is measured. We conclude that the PCE metric is proportional to $1/\text{RT}$.

Keywords: image recognition; algorithms and filters; pattern; nonlinear correlators; pattern recognition; rotation-invariant pattern recognition

1. Introduction

Since the first appearance of the classical matched filter (CMF) [1], many advances have been made in optical and digital pattern recognition, particularly in correlation methods and the design of new filters. At the beginning of these studies, efforts to achieve an optical device for the identification and counting of objects were unfortunately frustrated because the output correlation peak, using the CMF, is drastically degraded with the geometrical distortion sensitivity, especially to small-scale and rotation changes, of the target. To overcome these obstacles, a polar transformation and the Mellin transform were introduced [2] to make the system invariant to scale and rotation.

The brain has the capability to detect and recognize the form of objects, independently of size and orientation. The perception of the three-dimensional world via the visual system, in humans and some animals, is a natural process, but the cognitive mental processes involved are unique in humans. Light coming from a visual scene is focused by the cornea and the crystalline lens onto the retina, producing a chemical reaction in the rod and cone photoreceptors and, after processing the light in the retina, the photons are converted into electric signals and transmitted through the optic nerve until they reach the primary visual cortex (also called, from different forms, the striate cortex, Brodmann’s area 17, or visual area 1, abbreviated as V1) [3].

Some studies have been performed on animals [4], looking for an analytic treatment to describe the transformation of the visual scene from the retina to the brain cortex in, for example, cats. The first mathematical functions describing this mapping were provided by Schwartz [5] studying the striate cortex in primates. These studies showed that between the retina and the visual cortex there exists a complex logarithmic polar mapping and that this processing is governed by the same analytical functions used in optical and digital pattern recognition [6]. This discovery was very important for the developments of new optical and digital pattern recognition methods. At the same time, some studies in optical pattern recognition [7], showed that the rotation and size-invariant properties of the cortical map were essential to any cross-correlation basis for stereopsis [5]. Some years later a direct visual cortex mapping in humans was realized using positron tomography [8]. We can say that these studies were crucial for new developments in the field of pattern recognition systems such as optical, digital and hybrid systems.

*Corresponding author. Email: josue@cicese.mx

ISSN 0950–0340 print/ISSN 1362–3044 online
© 2010 Taylor & Francis
DOI: 10.1080/09500340903511695
http://www.informaworld.com
Digital images can be obtained in different environmental conditions using an optical–digital sensor, but this causes several problems in the identification and characterization of the object using computerized vision systems, hence this study is limited to static objects and bi-dimensional images. In this work, we obtain some results with a new invariant computational algorithm for the recognition of alphabetic letters with different fonts and styles and the effects of five color combinations. We emphasize here the introduction of the Scale transform [9], instead of the well known Mellin-Fourier transform [7], because it is very sensitive to scale variation. A kth law nonlinear filter is introduced to realize the digital invariant correlation that gives us information on the similarity between different objects. This kind of filter has advantages compared with the phase only filter (POF) [10] and other linear filters, due to its great capacity to discriminate objects, the maximum value of the correlation peak is well localized, narrow, with small sidelobes and the output plane is less noisy.

In other works, a kth law joint transform correlator is used [11–13], instead of a kth law nonlinear filter, in an optical coherent system [14], with better results, where the nonlinear characteristics are used to generate the nonlinear filter with an electrically addressed Spatial Light Modulator (SLM) located in the Fourier plane of the optical correlator. It is shown that both types of nonlinear filter produce better correlation performance compared with linear filters [13]. In our case, we used the kth law nonlinear filter because we are treating an invariant digital system and not an optical system. However, in an optical system it is difficult to implement the Scale transform in order to have the invariances to scale and rotation. In this paper, the nonlinearity affects the Fourier magnitude of both the input signal and the reference signal.

The principal differences in the characteristics of letter fonts can be found the two main types of fonts, namely serif and sans serif. The Arial font is simpler in shape, it has a uniform stroke width, without serif, and the letter height is larger; Times New Roman, on the other hand, has a variable stroke width, with serif, and the letter height is smaller. The Courier New font is a mono-spaced slab serif typeface. The behavior of these letters has been studied for many years [15], with many individuals measuring the reaction time (RT) in the readability and legibility of a written text in different fonts and styles. In this paper, we compare the font Arial (sans serif), Courier New (slab serif) and Times New Roman (serif) font types, in plain and italic styles, with five different combination colors and we investigate the relationship between the RT and the mean PCE values applying the invariant digital correlation method to these alphabetic letters.

2. Invariant digital correlation system

The principal elements used in our invariant digital correlation system are the following: (a) a kth law nonlinear filter (NLF), and (b) a phase only filter (POF). The NLF has the characteristic that its output response is a nonlinear function of the input response, and it is expressed as [16]

\[ NLF = |H(u, v)|^k e^{-j\phi(u,v)}, \]

where \(|H(u,v)|\) is the absolute value of the Fourier transform of the object to be recognized, \(j = \sqrt{-1}\), \(k\) is the nonlinear strength factor that takes values \(0 < k < 1\) and \(\phi(u,v)\) is the phase of the Fourier transform. Changing the \(k\) values in this interval, one can manipulate the discriminate capacities of the nonlinear processor and therefore determine the best \(k\) values of the nonlinear filters. For different \(k\) values in this interval, we have \(k = 1\) (classical matched filter), \(k = 0\) (phase only filter), and \(k = -1\) (inverse filter). So, the POF has the form [10]

\[ POF = e^{-j\phi(u,v)}. \]

These types of filters have been widely described in other works [13,17].

For a better evaluation of the plane correlation, the criteria used is the PCE metric that is defined by [18]:

\[ PCE = \frac{|E[c(0,0)]|^2}{E[|c(x,y)|^2]}, \]

where the numerator is the peak correlation intensity expected value and the denominator is the mean energy expected value in the correlation plane.

2.1. The Scale transform

The Scale transform is a special case of the Mellin transform, which is defined as [19]

\[ M_f(p) = \int_0^\infty f(t)t^{p-1}dt, \]

in the complex variable \(p = -jc + \beta\), with the fixed parameter \(\beta \in \mathbb{R}\) and the independent variable \(c \in \mathbb{R}\). This family of transforms is called the \(\beta\)-Mellin transform. The real part of the complex variable \(p\) is parameterized; for the case \(\beta = 1/2\) we have the Scale transform. Other values for \(\beta\) are possible, for \(\beta = 0\) we have the compression/expansion invariant transform, and for \(\beta = -1\) we have the form invariant transform. Therefore, the Scale transform is a restriction of the Mellin transform on the vertical line \(p = -jc + 1/2\).
We introduce the Scale transform due to its property of invariance to size changes. The Scale transform and its inverse are given by [20]

$$D_f(c) = (2\pi)^{-1/2} \int_0^{\infty} f(t) \exp((-jc - 1/2) \ln t) dt$$

and

$$f(t) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} D_f(c) \exp((jc - 1/2) \ln t) dc.$$ (6)

We used the 2-D Scale transform in polar coordinates \((r, \theta)\) with the log of the radial coordinate \(\lambda = \ln r\) that is expressed as [21]

$$D(c, c_0) = (2\pi)^{-1/2} \int_0^{\infty} \exp(\lambda/2)f(\lambda, \theta) \times \exp(-j(\lambda c_0 + \theta c_0)) d\lambda d\theta,$$ (7)

where the non-separable Scale transform implementation has been utilized in order to maintain the invariance to rotation [19].

3. The nonlinear invariant correlation method

The steps to obtain the nonlinear correlation are shown in Figure 1. First, we have the original image or target (step 1) denoted by a bi-dimensional function, \(f(x,y)\), where \(x\) and \(y\) are spatial coordinates in the Cartesian plane and the value of this function is a non-zero and finite scalar quantity, i.e. \(0 < f(x,y) < \infty\). The next step is to calculate the Fast Fourier Transform (FFT); we obtain (step 2) the modulus of the Fourier transform denoted as \(|F(w_x, w_y)|\), where \((w_x, w_y)\) are the spatial frequency coordinates in the Fourier plane, and in this way the displacement of the input image is not affected in the Fourier plane. After this, we realized a high-pass filtering using a parabolic filter applied to the modulus of the Fourier transform [22]. This kind of filter attenuates low frequencies while passing high frequencies that enhance sharp details. Next, a scale factor, given by \(\sqrt{r}\), is introduced where \(r\) is the radial spatial frequency. Note that it is important to differentiate the Scale transform from the Mellin transform. Cartesian coordinates are mapped on to polar coordinates to obtain the rotation invariance and a bilinear interpolation is done to the first data of the coordinate conversion to reduce the aliasing due to the log-polar sampling. A logarithmic scaling is made in the radial part in polar coordinates (step 3). And taking the FFT in this step, we obtain the nonlinear filter (step 4), where \(|T|\) is the absolute value of the Fourier transform of the object to be recognized when, in the step 1 we have the target information or we choose to take step 5 when we have in step 1 the problem image to recognize, where \(|G|\) is the absolute value of the Fourier transform of the problem image. \(k\) is the nonlinear strength factor (steps 4 and 5). Thus, the nonlinearity affects the Fourier magnitude of both the input signal and the reference signal.

Finally, steps 4 and 5 show the procedure to obtain the invariant digital correlation (step 6).

4. Experimental results

In our digital experiments, we used the \(k\) nonlinearity strength factor value for the nonlinear filter given by \(k = 0.3\) [23], and upper case letters with zero rotation degree and which are not scaled. The results were obtained with 95% level of confidence.

The results of this invariant system for the Arial and TNR font letters are shown in Figure 2. The images are auto-correlated, taking each one of the alphabetic letters. From this figure, we observe better results using a NLF than a POF for the Ar and TNR fonts. In this case, all the alphabetic letters were black/white. The difference in the PCE values between these filters was remarkable: NLF was better than the POF by approximately a factor of ten. That was the reason we choose the NLF instead of the POF in our invariant system. In general, using the NLF, the PCE values were better for the TNR font than for the Ar font. We can observe that for the letters C, O, Q, S and W, the PCE values for the Ar font were slightly greater.
than for the TNR font. And for the letters A, G and V, the PCE values were the same for both fonts.

Figure 3 shows the mean PCE values obtained versus the letters in Arial (Ar), Courier New (CN) and Times New Roman (TNR) fonts in italic and plain styles, with five foreground/background color combinations; black-on-white (BK/W), green-on-yellow (GN/Y), red-on-green (R/GN), white-on-blue (W/BL) and yellow-on-blue (Y/BL), using an invariant digital correlation system with a nonlinear filter with $k = 0.3$. The foreground/background colors were selected with color coordinates RGB (r,g,b), normalized to [0–255] (bytes), as follows; red: (255,0,0), green: (0,128,0), blue: (0,0,255), yellow: (255,255,0), black: (0,0,0) and white: (255,255,255). The letters are 512 × 512 pixels in size and in bitmap file format with 256 colors. The results show that the letters in TNR font have a better mean PCE when compared with the Ar and CN fonts.

For the italicized style, we can observe from this graph that the TNR font had a greater mean PCE value compared with Ar and CN fonts. All of them in this style had the same performance and the color combinations were irrelevant.

For the plain style we can observe that the TNR font also had a greater mean PCE value compared with Ar and CN fonts. However, contrary to the italicized style, in the plain style we have found that the color combinations affect the mean PCE value of the font type letters. The black-on-white and the white-on-blue color combinations had the higher mean PCE values. For the green-on-yellow, red-on-green and yellow-on-blue color combinations, there were no significant changes in the mean PCE values, and like in the italic case, their behavior was the same, and the color combinations were irrelevant and their mean PCE was statistically not significant.

In order to compare Ar and TNR fonts, we choose the letters of both fonts in plain style. The box graph in Figure 4 shows that the TNR font had greater mean PCE values compared with the Ar font.

The next task was to compare all the Ar and TNR fonts in italic style. The box graph in Figure 5 shows a better mean PCE value for the all TNR fonts in italic style compared with the Ar font. And finally, in Figure 6, we compare all Ar and TNR fonts in both italic and plain styles, and, in this case, for the set of all the letters of Ar and TNR, the italic style had a better mean PCE value.

5. Discussion
The invariant digital correlation method used in this work is more general, and is used for correlation invariant to changes in rotation and scale of the object. However, in our work the letters have no changes of rotation and scale, because in the vast majority of written text, websites, emails, etc., there are no rotation or scale changes for the letters. Hence, in this case, the problem is more easily solved with our method, where only the auto-correlation, with zero degree of rotation and no difference in scale, of each letter was taken into account.

The effect of text-background color combinations, font, word style, pixel size, and others variables, are
very important in text legibility and for readability. Using the RT of some individuals reading a text, the effects of six color combinations, three font types (Arial, Courier New and Times New Roman) in italic and plain styles were studied [15] and the results showed that for green-on-yellow, in italic, the TNR font had the better RT values than the Arial font. For all color combinations in italic and plain style, the TNR font had a smaller RT than the Ar font, except for black-on-white and red-on-green in italic, and for black-on-white and white-on-blue in plain style. In general, it was found that plain word style had a better RT than italic word style. Additionally, legibility performance and subjective preference for color combinations in text has been studied [24], and no evidence has been found for an influence of luminance polarity or chromaticity on performance. How the color combinations can improve the legibility and pleasantness of texts on a computer screen or projected on a wider surface has also been studied [25], and these results showed that is better to have a dark text on a light background. Other studies [26] related color combinations and other variables, such as ambient illumination, luminance contrast, and stimulus type, to subjective preference of a video display terminal (VDT) target, such as aesthetic appearance, legibility and visual comfort. These studies have shown that black-on-white and blue-on-white were the most preferred color combinations, with higher luminance contrast.

In our results, it is significant that in plain and italic styles, the TNR had a better performance with greater mean PCE values than Ar (Figures 3–5). For our analysis we choose only the Ar and TNR fonts (Figures 4–6), because they are, in general, the most commonly used on the Web and in reading texts. The black-on-white and the white-on-blue were the most 'preferred' color combinations in our digital system. It was very surprising for us that our results, based on one analytical mathematical method, are
in agreement with some others subjective studies. We have assumed that the output, given in PCE value, is inversely proportional to the reaction time of the reader, i.e. PCE \( \propto 1/RT \). We based this consideration on the investigations realized in the retina-cortical mapping in humans [8] and specifically in connection with Gestalt psychology, because the studies realized in humans have to do with the log polar mapping in the visual cortex, which is a prerequisite for obtaining invariant optical and digital pattern recognition. This means dealing with perceptions that are Gestalten phenomena rather with sensations. Some notable characteristics of perception are the following: three-dimensionality, curvature, rotation, form, and contour, among others. Legibility and readability are directly related, legibility is inversely proportional to the RT, and reading speed is a criterion used for legibility. Our results may give a better and reliable explanation objectively to the current findings on the process of reading.

6. Conclusions
In this work, we presented a new analytical method — using a digital invariant nonlinear correlation for position, rotation and scale — for comparison and
analysis of different font and style letters with five color combinations. Our results showed a better output for the TNR font in comparison with the Ar font in italic and plain styles. Generally, the italicized letter style had greater PCE values than the plain style letter. These results were in accord with some other studies realized with subjective methods. Our results can be useful for studying the effects of background color combinations on the legibility of text displayed on a computer screen, from the Web, email, or other texts written in books, newspapers, magazines, etc, and the improvements in readability. It would be interesting to investigate the relationship between our method, based on analytical mathematical functions, and those other subjective empirical methods used by psychologists and Web designers, among others.

Acknowledgement

This study is based on work partially supported by CONACYT under Grant No. 102007.

References