Infrared Physics & Technology 53 (2010) 464–468

Contents lists available at ScienceDirect

Infrared Physics & Technology

journal homepage: www.elsevier.com/locate/infrared

Structural similarity determines search time and detection probability

Alexander Toet *

TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
Intelligent System Laboratory Amsterdam, Faculty of Science, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 29 May 2010
Available online 21 September 2010

Keywords:
Clutter
Structural similarity index (SSIM)
Target structure similarity (TSSIM)
Detection time
Detection probability

ABSTRACT

The recently introduced TSSIM clutter metric is currently the best predictor of human visual search performance for natural images (Chang and Zhang [1]). The TSSIM quantifies the similarity of a target to its background in terms luminance, contrast and structure. It correlates stronger with experimental mean search times and detection probabilities than other clutter metrics (Chang and Zhang [1,2]). Here we show that it is predominantly the structural similarity component of the TSSIM which determines human visual search performance, whereas the luminance and contrast components of the TSSIM show no relation with human performance. This result agrees with previous reports that human observers mainly rely on structural features to recognize image content. Since the structural similarity component of the TSSIM is equivalent to a matched filter, it appears that matched filtering predicts human visual performance when searching for a known target.

1. Introduction

It is well known that visual targets that are similar to their local background or to details in other parts of the scene are harder to find than targets which are highly distinct. This obscuring effect, which is generally known as clutter, determines human visual search and detection performance in electro-optical image to a large extent. Many attempts have been made to quantify the effects of clutter by means of digital clutter metrics. However, since the concept is inherently elusive, attempts to model clutter have only been partly successful [3–6,6–17].

Visual search experiments have shown that detection performance depends mainly on the energy contrast between a target and its local background, whereas recognition depends mainly on the structural dissimilarity between a target and its surround [18,19]. For complex scenes, the spatial relationships (shape and relative location) of features in an image can have a greater effect on detection than the relative luminance of the features [3]. Higher overall contrast may even reduce the amount of perceived clutter because confusing objects are more readily recognized for what they are – nontarget scene elements. An effective clutter metric should account for this type of cognitive screening.

Wang and Bovik introduced the structural image similarity index (SSIM) which measures the similarity between images in terms of luminance, contrast and structure [20–24]. The SSIM has successfully been deployed to model human visual perception of image distortions and modifications in a wide range of different imaging applications (for an overview see [22]). Chang and Zhang [1,2] recently introduced the TSSIM clutter metric, which deploys the SSIM to quantify the similarity of a target to its background in terms luminance, contrast en structure. They showed that the TSSIM correlates significantly with mean search time and detection probability [1,2]. However, it is not immediately obvious to what extent each of the three TSSIM components contributes to this correlation.

Here we analyze the predictive performance of each of the three TSSIM components, and we show that it is predominantly the structural similarity component which determines human visual search performance, whereas the luminance and contrast components of the TSSIM show no relation with human performance. The rest of this paper is organized as follows. In Section 2 we show how rewriting the TSSIM in its full form allows the assessment of the contribution of the luminance, contrast and structural similarity components to the overall clutter metric. In Section 3 we describe how the performance of the TSSIM was evaluated by deployment to a set of natural images for which human observer data are available. The results of this experiment are presented in Section 4. Finally, the conclusions of this study are presented in Section 5.

2. Clutter metrics

2.1. The structural similarity (SSIM) index

Let $x = \{x_i | i = 1, 2, \ldots, N\}$ and $y = \{y_i | i = 1, 2, \ldots, N\}$ represent two discretely sampled grayscale image patches that need to be compared. Let μ_x, μ_y, σ_x, σ_y, σ_{xy} respectively be the mean of x, the

* Address: TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands. Tel.: +31 346 356237; fax: +31 346 353977.

E-mail address: lex.toet@tno.nl

1350-4495/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.infrared.2010.09.003

mean of y, the standard deviation of x, the standard deviation of y, and the covariance of x and y, defined as:

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{1}$$

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2\right)^{\frac{1}{2}},$$
$$\sigma_y = \left(\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\mu_y)^2\right)^{\frac{1}{2}} \tag{2}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y) \tag{3}$$

The mean signal intensity and its standard deviation (the square root of variance) can be regarded as rough estimates of respectively local image luminance and contrast. The covariance of x and y (normalized by their respective variances) reflects the tendency of the two signals to vary together, and can therefore be adopted as a measure of the structural similarity between the two signals.

The similarity of the local patch luminances is then defined as

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2+\mu_y^2+C_1} \tag{4}$$

where C_1 is a small constant given by $C_1 = (K_1L)^2$, which is introduced to stabilize the computation of (4) when the denominator becomes small, L is the dynamic range of the pixel values ($L = 255$ for 8 bits/pixel grayscale images), and $K_1 \ll 1$ is a scalar constant (typically 0.01). The dynamic range of l is $\langle 0, 1\rangle$. The maximum value 1 is approached when both image patches have the same luminance: $\mu_x = \mu_y$.

The similarity of the local patch contrasts is defined as

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2+\sigma_y^2+C_2} \tag{5}$$

where C_2 is a small constant given by $C_2 = (K_2L)^2, K_2 \ll 1$ (typically 0.03). The dynamic range of σ is $\langle 0, 1\rangle$. The maximum value 1 is approached when both image patches have the same contrast: $\sigma_x = \sigma_y$.

The structural similarity between the image patches is defined as

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{6}$$

with $C_3 = C_2/2$. The dynamic range of s is $\langle -1, 1\rangle$. The maximum value 1 is approached when $y_i = ax_i + b$ for all $i = 1, 2, \ldots, N$, where a and b are constants and $a > 0$.

The overall structural similarity index SSIM between signals x and y is defined as:

$$\mathrm{SSIM}(x,y) = |l(x,y)|^{\alpha}\cdot|c(x,y)|^{\beta}\cdot|s(x,y)|^{\gamma} \tag{7}$$

where α, β and γ are parameters that define the relative importance of the three components.

Setting $\alpha = \beta = \gamma = 1$ and substitution of Eqs. (4)–(6) in Eq. (7) results in

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{8}$$

which is the form in which the SSIM is typically used in the literature [21–25].

When comparing two images, the SSIM index is computed locally within a sliding window that moves pixel-by-pixel across the image, resulting in a SSIM map. The SSIM score of the entire image is then computed by pooling the SSIM map, e.g., by simply averaging the SSIM values across the image. SSIM has successfully been applied in a large number of different applications (for an overview see [22]).

2.2. The target structural similarity (TSSIM) clutter metric

Chang and Zhang [1,2] adapted the SSIM for use as a clutter metric, and introduced the target structure similarity metric (TSSIM). Their approach is as follows:

The image for which the clutter metric has to be calculated is divided into N blocks. The blocks are twice the apparent size of the typical search target in each dimension.

Let $T = \{T_i|i = 1, 2, \ldots, N\}$ and $B_j = \{B_{ji}|i = 1, 2, \ldots, N\}$ represent respectively the discretely sampled grayscale target block (i.e. the part of the images which contains the target support area and a local background area around the target) and the jth background image block.

Substitution of T and B for x and y in Eqs. (4)–(6) and neglecting the stabilizing constants yields

$$l(T,B_j) = \frac{2\mu_T\mu_{B_j}}{\mu_T^2+\mu_{B_j}^2} \tag{9}$$

$$c(T,B_j) = \frac{2\sigma_T\sigma_{B_j}}{\sigma_T^2+\sigma_{B_j}^2} \tag{10}$$

$$s(T,B_j) = \frac{\sigma_{TB_j}}{\sigma_T\sigma_{B_j}} \tag{11}$$

Substituting Eqs. (9)–(11) in (7), with $\alpha = \beta = \gamma = 1$, and adopting a single constant C to avoid instabilities, yields the TSSIM metric [1,2]:

$$\mathrm{TSSIM}(T,B_j) = \frac{4\mu_T\mu_{B_j}\sigma_{TB_j} + C}{(\mu_T^2+\mu_{B_j}^2)(\sigma_T^2+\sigma_{B_j}^2)+C} \tag{12}$$

Chang and Zhang used both $C = 0.2$ [2] or $C = 0$ [1]. Here we also adopt $C = 0$ since we observed no instabilities for the image set used in our experiments.

The overall image TSSIM is then calculated in two ways: both as the root mean square of TSSIM_j (TSSIM_{rms}) and the arithmetic mean of TSSIM_j (TSSIM_{am}).

$$\mathrm{TSSIM}_{rms} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\mathrm{TSSIM}_j^2} \tag{13}$$

$$\mathrm{TSSIM}_{am} = \frac{1}{N}\sum_{j=1}^{N}\mathrm{TSSIM}_j \tag{14}$$

where TSSIM_j is the structure similarity measure in the jth image block, and N is the total number of image blocks. The rationale of this metric is the fact that observers will need more time to inspect the image when it contains more details similar to the target. Details similar to the search target can also distract and confuse the observer, and may result in false alarms, thus degrading the detection probability. Thus, a higher TSSIM value corresponds to more clutter in the image, leading to longer search (inspection) times and lower detection probability.

2.3. The TSSIM in its full form

Using Eqs. (9)–(11), and setting all stabilization constants to zero, we rewrite Eq. (12) in its original full form (Eq. (7)), which allows the assessment of the individual contributions of luminance, contrast and structural similarity to the overall TSSIM clutter metric: