Intelligent Systems for Advanced Manufacturing, Newton, Massachusetts, USA, 28–31 October 2001, Proc. SPIE 4572

Strategies to detect non-linear similarities by means of correlation methods

F. Puente León and M. Heizmann

Institut für Mess- und Regelungstechnik, Universität Karlsruhe (TH),
Postfach 6980, D-76128 Karlsruhe, Germany

ABSTRACT

Correlation methods represent a well-known and reliable approach to detect similarities between images, signal patterns, and stochastic processes. However, by means of the cross-correlation function only linear similarities are registered. Unfortunately, often it is not possible to avoid non-linearities in the characteristics of the sensors used or, as in image processing, in the interaction between illumination and the scene to be captured. Thus, in such cases correlation methods may yield poor results. In this paper, we describe alternative strategies to enhance the performance of correlation methods even when the statistical connection between the signals is non-linear. To reduce the impact of non-linearities on the signals to be analyzed, a preprocessing is performed in which certain properties affecting first-order statistics are manipulated. This step impresses the same histogram to the signals to be compared, so that typically higher correlation coefficients are obtained as compared to if no preprocessing methods were used. The performance of our approach is demonstrated with two different tasks. First, a preprocessing strategy is proposed for signals obtained from train-based sensors to enable an identification of rail switches. Finally, a method for comparing striation patterns in forensic science is presented. To investigate the benefit of this approach, a large database of toolmarks is used.

Keywords: pattern recognition, correlation methods, histogram transforms, generalized similarities, preprocessing, energy minimization methods, cross-entropy analysis

1. INTRODUCTION

The detection of similarities between signals is a common task in signal processing. Typical applications of recognition strategies for one- and two-dimensional data are widely spread.

To deal with this objective, many strategies have been developed to enable a sophisticated and efficient comparison. Here, one of the best known approaches is the application of correlation methods.[10] The benefits of correlation techniques comprise the following points:

- Due to the widely spread use of the correlation calculus, these methods are accessible to extensive theoretical investigations. In so far, experimental detection results can be verified by means of values expected theoretically.
- Since the calculation of the correlation function can be performed by means of a multiplication in the frequency domain, efficient implementations like those based on the FFT algorithm[2] speed up the calculation time required for the comparison.

Unfortunately, correlation methods often suffer from the precondition of a linear similarity between the signals in question. Since for many reasons, the exclusive detection of linear connections is not sufficient, other approaches have to be followed. On the other hand, the exorbitant computational expense of many alternative strategies often prohibit their implementation.

In order to escape from this dilemma, a novel approach is presented in this paper that allows to use correlation methods even in the case of signals showing a non-linear similarity. To this end, preprocessing strategies are proposed which are

Further author information:
Email: {puente, heizmann}@mrt.mach.uni-karlsruhe.de; WWW: http://www-mrt.mach.uni-karlsruhe.de;
Telephone: +49-721-608-{2334, 2338}; Fax: +49-721-661874

based on the manipulation of the signal statistics. By this means, detection results can be improved significantly while preserving the favorable properties of correlation methods concerning the computation time required.

The basis of correlation methods is presented in Sect. 2. For the application of alternative strategies, generalized types of similarity are introduced; see Sect. 3. Based on these extended definitions of similarity, detection strategies are developed. Following, a promising approach to adjust the first-order statistics of two given signals to each other is described in Sect. 4. Finally, the performance of the proposed strategy is demonstrated with two examples: the detection of rail switches, which can be used for a train-based positioning system, and the enhancement of recognition methods used for the identification of striation marks in forensic science.

2. DETECTION OF LINEAR SIMILARITIES

The empirical *cross-correlation function* (CCF)

$$k_{12}(\tau) := \tilde{q}_1(\xi) \circledast \tilde{q}_2(\xi) = \int_{-\infty}^{\infty} \tilde{q}_1(\xi)\, \tilde{q}_2(\xi + \tau)\, \mathrm{d}\xi \tag{1}$$

provides a reliable approach to detect similarities between two signals $q_1(\xi)$ and $q_2(\xi)$ of finite energy, where

$$\tilde{q}_1(\xi) := \frac{q_1(\xi) - m_{q_1}}{s_{q_1}} \qquad \text{and} \qquad \tilde{q}_2(\xi) := \frac{q_2(\xi) - m_{q_2}}{s_{q_2}} \tag{2}$$

denote the signals centered around their mean values m_{q_i} and normalized to have a standard deviation of 1:

$$s_{q_i} := \sqrt{\mathrm{var}\{q_i(\xi)\}}\,. \tag{3}$$

By means of this normalization, the CCF becomes independent of a global offset as well as of a global scaling parameter. The location of the maximum of the CCF $k_{12}(\tau)$ indicates the shift τ_0 leading to the best possible correspondence between the two signals:

$$\tau_0 := \arg\max_{\tau}\{k_{12}(\tau)\}\,. \tag{4}$$

Furthermore, the maximum value of the CCF specifies the similarity between both signals quantitatively:

$$\rho_{12} := \max\{k_{12}(\tau)\}\,. \tag{5}$$

In Fig. 1, the proposed strategy is illustrated with an example from forensic science—the comparison of firearm bullets based on signals extracted from gray level images of them. In the upper area, the signals $q_1(\xi)$ and $q_2(\xi)$ extracted from the images $r_1(\mathbf{x})$ and $r_2(\mathbf{x})$ of the circumferential surface* of two different bullets are shown.[7] In the center of the figure, the CCF $k_{12}(\tau)$ is depicted, the maximum of which indicates the displacement τ_0 of both signals. At the bottom, the corresponding original images from which the signals $q_1(\xi)$ and $q_2(\xi)$ were extracted are represented, shifted by the distance τ_0. Even a layman can recognize at once the great similarity of both bullets, which indeed were fired from the same gun.

To apply this method to discrete signals, Eq. (1) can be implemented very efficiently in the frequency domain by means of the FFT algorithm.[2] The periodic continuation of the signals inherent to the discrete Fourier transform does not impair the result of the computation, because in this case the signals $\tilde{q}_1(\xi)$ and $\tilde{q}_2(\xi)$ are actually cyclic. An important advantage of the proposed methodology is its low computational expense: the computing time required for the correlation of two signals $\tilde{q}_1(\xi)$ and $\tilde{q}_2(\xi)$ on a conventional PC is presently about $12\,\mathrm{ms}$, if signals of up to 8192 samples are assumed. Moreover, this time could be drastically reduced by employing dedicated correlation hardware.

However, it should be emphasized that by means of the CCF only the linear similarity between two signals or processes can be registered. For this reason, in the next section several approaches will be discussed that allow the investigation of arbitrary types of statistical dependence, as well as detecting similarities in case that deformations—such as a local scaling of the signals—have taken place.

*The images $r_1(\mathbf{x})$ and $r_2(\mathbf{x})$ itself were generated by means of a concatenation of several pictures recorded at different rotary positions of the bullets, as described in Ref. 12.

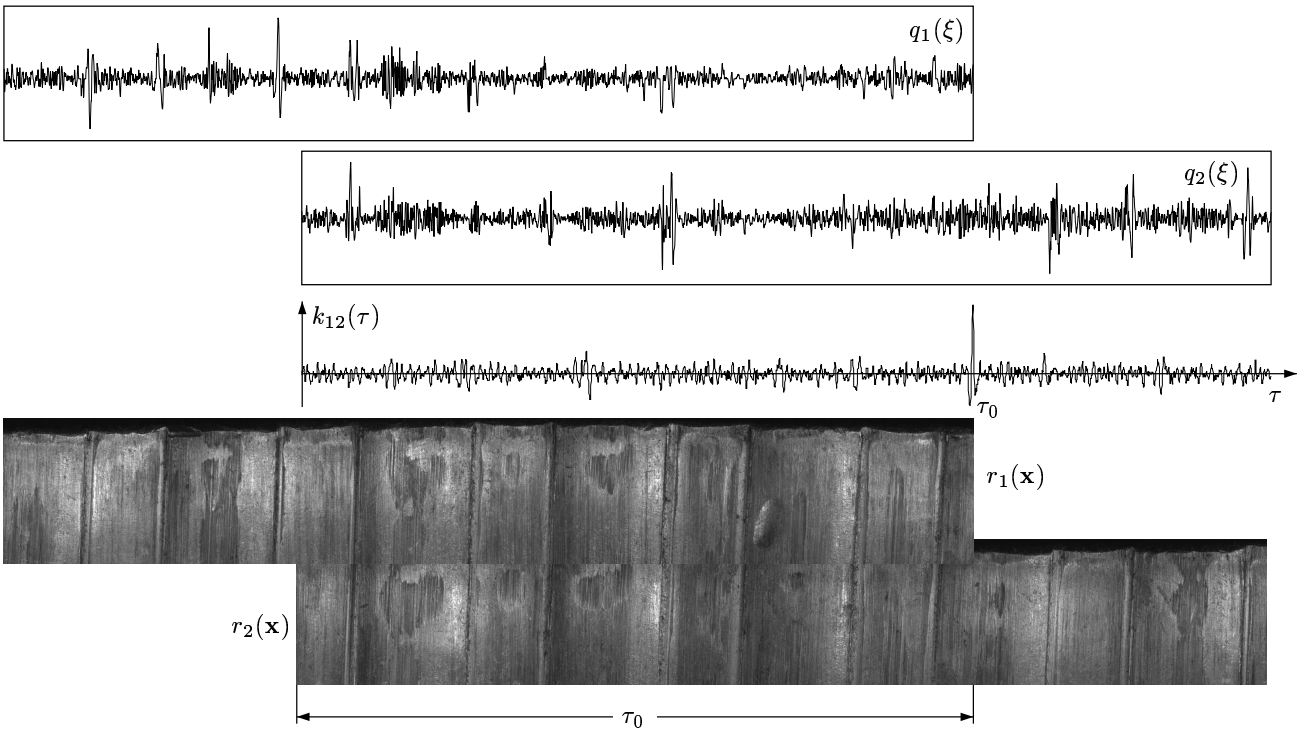

Figure 1. Detection of similarity by means of the CCF: comparison of two firearm bullets fired from the same gun.

3. GENERALIZED APPROACHES

In many cases, more general tools than the CCF are required to detect similarities between two signals. The reasons may arise from different sources:

- The use of the CCF assumes that the signals compared show linear similarities, which implies that they are represented by means of a metric scale. However, in some areas like e.g. automated visual inspection, it is not always sufficient to detect linear similarities. Non-linear transfer functions of the system in question are common and have to be considered. As an example, the characteristics of CMOS sensors can be approximated by a logarithmic curve.

- If system properties can be described only by ordinal scales, additional difficulties arise. In such cases, not the CCF is the means of choice for detecting similarities. Rather, rank correlation methods should be preferred to a transform of the ordinal data to a metric scale.

- In case that system properties cannot be represented even by means of an ordinal scale—e.g. when no order relation between two different events can be established—the signal cannot be associated to a stochastic variable. Since both CCF and rank correlation assume a rank order in the value range of the signal, the transform to a suitable scale may complicate the interpretation of the detection results and should thus be avoided.

Consequently, alternative comparison strategies should be taken into consideration when the process of interest and its resulting signals do not meet the requirement of linear similarity.

3.1. Types of similarity to be detected

Although the expression of “similarity” is often used synonymously for “linear similarity,” more generalized types of likeness may be more suitable to describe connections between signals:

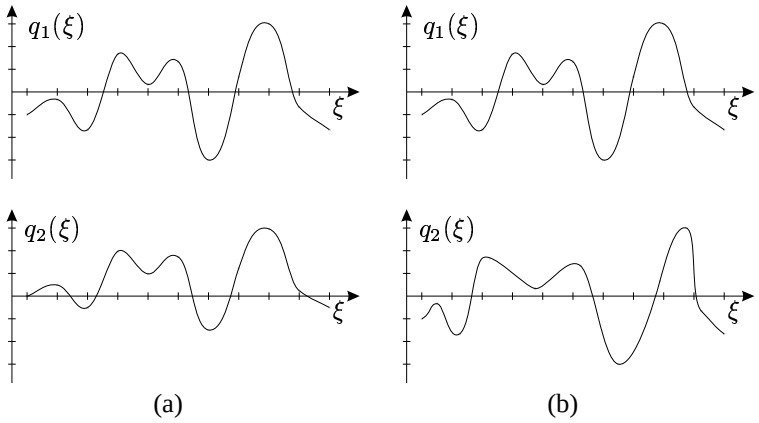

Figure 2. Similar signals concerning different types of similarity: (a) non-linear amplitude scaling; (b) non-uniform abscissae scaling.

(a) The first generalization concerns the amplitude values of the signals. Here, similarity does not necessarily mean that the signals are linearly dependent, i.e. in general they are non-linearly distorted, leaving the order of the amplitude values unchanged. The abscissae of both signals are scaled identically; see Fig. 2(a).

A typical example for the occurrence of such signal distortions is the acquisition of digital images, where the unchanged optical setup provides for a constant spatial magnification, whereas the reflectance properties of the objects recorded or the sensor characteristics may cause non-linear distortions.

(b) Secondly, the abscissae values of the signals may show linear or non-linear distortions; see Fig. 2(b). Such signal properties are typical for time dependent processes, where the time base is established e.g. by a non-uniform motion. In addition, a non-linear amplitude scaling may occur.

(c) A highly general approach for defining a similarity dispenses with an ordinal scale to describe the events emerging from the process of interest. Here, the realizations of the process need not be arranged in a hierarchic order. Instead, a suitable description uses only the probability of each possible outcome. In contrast to the generalizations mentioned above, where extensions of correlation methods may lead to suitable detection strategies, in this case only cross-entropy methods are applicable.

Considering the conciseness of the three definitions presented, the degree of similarity decreases from (a) to (c). On the other hand, the need for general detection issues increases.

3.2. Energy minimization methods

Energy minimization methods provide a powerful tool to formulate an optimization problem, even in the case of multiple constraints.[3, 11] This fact makes them highly interesting to face ill-posed problems, such as the detection of generalized similarities. The optimization result is obtained by minimizing a generalized energy function

$$E = \sum_i \lambda_i E_i \;\rightarrow\; \min\,, \tag{6}$$

where the desired properties of the result are represented by the energy terms E_i and weighted with their respective factors λ_i. Therefore, the energies are to be chosen such that they reflect a priori knowledge on the optimal result as well as on the constraining conditions. In order to enable a sensible minimization, the energy terms must have monotonically decreasing values towards more desirable results.

The application of energy minimization to detect similarities is demonstrated with the second example discussed in Subsect. 3.1, where the abscissae dimension of the two signals is scaled non-uniformly. A possible approach to treat such similarities could be to apply an abscissae scaling $\eta(\xi)$ to the second signal $q_2(\xi)$. Figure 3 shows the signal $q_2(\xi)$ of the example depicted in Fig. 2(b) as well as the application of a suitable abscissae scaling function $\eta(\xi)$ which yields the

corrected signal $q_2(\xi + \eta)$. The scaling alters the sampling period locally and leads to a linear similarity which can be evaluated e.g. with the strategy presented in Sect. 2. For stability reasons and to avoid artifacts, the scaling function has to be a smooth curve. In addition, small values of $\eta(\xi)$ should be favored. These desired properties can be expressed by means of an energy function consisting of three terms:

$$\begin{aligned} E &= E_{\text{dist}}(\eta) + \lambda_1 E_{\text{smooth}}(\eta) + \lambda_2 E_{\text{sim}}(q_1, q_2) \\ &= \sum_{\xi} \eta^2(\xi) + \lambda_1 \sum_{\xi} \left(\frac{\mathrm{d}}{\mathrm{d}\xi}\, \eta(\xi) \right)^2 + \lambda_2 \, (-1) \, \left[q_1(\xi) \circledast q_2(\xi + \eta(\xi)) \right] , \quad \lambda_i > 0 \, . \end{aligned} \tag{7}$$

The first energy term, $E_{\text{dist}}(\eta)$, performs a quadratic rating of the distance of the altered signal $q_2(\xi + \eta)$ from the original by penalizing high values. The second term, $E_{\text{smooth}}(\eta)$, evaluates the smoothness of the scaling function $\eta(\xi)$. Finally, $E_{\text{sim}}(q_1, q_2)$ assesses the linear similarity between the altered signal $q_2(\xi + \eta)$ and $q_1(\xi)$. Here, other strategies to detect the similarity could also be applied. The factor (-1) provides for decreasing values of $E_{\text{sim}}(q_1, q_2)$ when the detected similarity grows.

In order to suitably weight the energy components, the regularization parameters λ_i have to be chosen properly. Ideally, the optimization leads to the minimum value of E, which yields the actual measure of deformation invariant similarity $E_{\text{sim}}(q_1, q_2)$ as well as the appropriate deformation correction $\eta(\xi)$.

However, although energy minimization methods establish flexible means for the formulation of optimization tasks, the application of such strategies causes considerable problems:

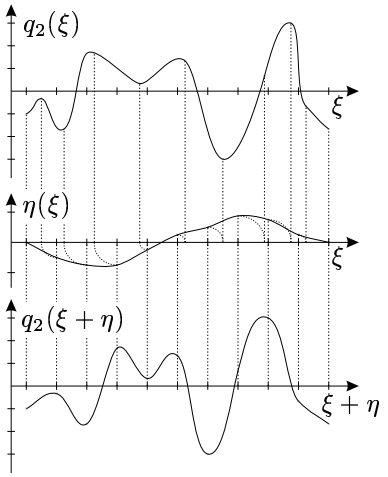

Figure 3. Abscissae scaling.

- The minimization of the energy function Eq. (6) has to be calculated iteratively. Thus, a stable solution cannot always be expected. To deal with instability, a suitable choice of the regularization parameters λ_i should be made. In the example of Eq. (7), the contribution of the smoothing energy component $E_{\text{smooth}}(\eta)$ is of great importance. On the other hand, however, smoothing may decrease the values of the similarity measure $E_{\text{sim}}(q_1, q_2)$.

- The result of the minimization depends on the choice of the regularization parameters λ_i. Since these can be specified arbitrarily within the stability limits, the minimization result represents only the optimum with respect to the corresponding parameter set.

- From a practical point of view, the evaluation and optimization of the extensive energy function leads to an enormous computational expense. Therefore, many approaches to deal with energy minimization methods do not explicitly calculate the energy function of Eq. (6). Instead, they perform an approximation based on a separate optimization of the energy components.

3.3. Cross-entropy analysis

A substantially more general approach, which stems from information theory and allows the investigation of arbitrary types of statistical dependence, is based on the computation of the so-called *cross-entropy function* (CEF).[5, 8, 13] The entropy H is defined as a quantitative measure of uncertainty about a signal generating process Q:

$$H(Q) := -\int_{-\infty}^{\infty} p_q(q) \, \log_2(p_q(q)) \, \mathrm{d}q \, , \tag{8}$$

where $p_q(q)$ denotes the probability density function (PDF) of $q(\xi)$.

When two processes Q_1 and Q_2 are given, the joint probability $p_q(q_1, q_2)$ for all possible combinations of the events $q_1(\xi)$ and $q_2(\xi)$ can be estimated. Accordingly, a joint entropy can be defined as

$$H(Q_1, Q_2) := -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_q(q_1, q_2)\, \log_2(p_q(q_1, q_2))\, \mathrm{d}q_1\, \mathrm{d}q_2\,. \tag{9}$$

If the processes Q_1 and Q_2 are statistically independent, i.e. if

$$p_q(q_1, q_2) = p_q(q_1) \cdot p_q(q_2)\,, \tag{10}$$

the joint entropy becomes due to the addition property of the logarithm

$$H(Q_1, Q_2) = H(Q_1) + H(Q_2)\,. \tag{11}$$

It can be shown that Eq. (11) establishes an upper limit for the joint entropy.[5, 13] Thus, the deviation of the joint entropy from this limit can be introduced as a measure of similarity between Q_1 and Q_2. That way, the common definition of the cross-entropy function[†] (CEF) $I(Q_1; Q_2)$ is obtained:

$$I(Q_1; Q_2) := H(Q_1) + H(Q_2) - H(Q_1, Q_2)\,. \tag{12}$$

In particular, the CEF shows the following properties:

- The CEF is symmetric with respect to the processes Q_1 and Q_2:

$$I(Q_1; Q_2) = I(Q_2; Q_1)\,. \tag{13}$$

- The values of the CEF range from 0 to the smaller value of $H(Q_1)$ and $H(Q_2)$. The CEF equals 0 if and only if the processes are statistically independent:

$$0 \leq I(Q_1; Q_2) \leq \min(H(Q_1); H(Q_2))\,, \tag{14}$$

$$I(Q_1; Q_2) = 0 \iff p_q(q_1, q_2) = p_q(q_1) \cdot p_q(q_2)\,. \tag{15}$$

For the task of detecting similarities, the CEF can be considered as a measure of statistical dependence. Here, it is of great importance that the application of the CEF only requires little preconditions. The probabilities $p_q(q_1)$ and $p_q(q_2)$ must be defined, and the event spaces of Q_1 and Q_2 have to be complete and disjointed:

$$\int_{-\infty}^{\infty} p_q(q_1) = 1\,, \qquad \int_{-\infty}^{\infty} p_q(q_2) = 1\,. \tag{16}$$

However, it is not necessary to set up an ordinal scale, which is equivalent to the statement that the event spaces of the processes need not have an hierarchical order. That way, the CEF represents a general tool to evaluate the similarities mentioned in Subsect. 3.1 (item (c)).

The main drawback of the cross-entropy analysis method to compare two signals is that with this approach, the existing order relation between the amplitudes of the signal $q(\xi)$—e.g. the degree of distinctness of individual signal peaks—is not taken into consideration. Therefore, the similarities indicated by the CEF must by no means necessarily correspond with empirically perceptible connections. Thus, the price of the generality of this alternative approach is that by analysis of the CEF, necessarily a lower information gain is obtained than in the case of the CCF, if the statistical dependence is essentially linear. Especially for image data, the similarities indicated by the CEF do not need to be established by visually prominent features.

[†]The cross-entropy function is also referred to as transinformation, mutual information, synentropy or average relevant information.

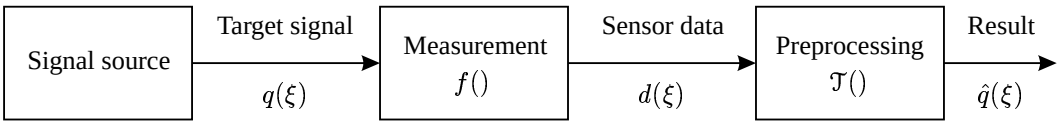

Figure 4. Preprocessing for manipulation of first-order statistics.

3.4. Non-linear signal transforms

In this section, an alternative strategy is proposed to enhance the performance of the methods according to Sect. 2 when statistical dependence is essentially non-linear. To explain the basic idea of this step, a parameter estimation model is shown in Fig. 4. A signal source emits the signal $q(\xi)$ to be analyzed by means of correlation methods. The target signal $q(\xi)$ may either be a measurable signal or an abstract construct that does not necessarily have to exist physically. The next block models the measurement system, which maps the signal $q(\xi)$ non-linearly onto the sensor data $d(\xi)$:

$$d(\xi) = f(q(\xi))\,. \tag{17}$$

To reduce the impact of the non-linearities related to the data acquisition stage, a non-linear signal transform $\mathcal{T}\{\,\}$ is performed in a preprocessing step. Ideally, this transform should invert the mapping according to Eq. (17):

$$\hat{q}(\xi) = \mathcal{T}\{d(\xi)\} \stackrel{!}{=} f^{-1}(d)\,. \tag{18}$$

However, in many cases the knowledge on the mapping $f(\,)$ is not sufficient to define a suitable transform $\mathcal{T}\{\,\}$, so that the preprocessing approach proposed in Eq. (18) will be impracticable. In such cases, an alternative strategy consisting in manipulating certain properties affecting first-order statistics of the signals to be compared can be employed instead;[‡] see Sect. 4.

4. MANIPULATION OF FIRST-ORDER STATISTICS

Even in case of a perfect correlation between two different signals $q_1(\xi)$ and $q_2(\xi)$ emitted by the information source depicted in Fig. 4, the non-linearities inherent to the data acquisition process may influence the signals such that the cross-correlation of the resulting signals $d_i(\xi)$ according to Eq. (1) is much lower than before the mapping $f(\,)$. Essentially, a linear mapping would have an impact on the shift as well as the width (and of course the amplitude) of the probability density function (PDF) $p_d(d)$ of the signals $d_i(\xi)$; see Figs. 5(a) and (b). In the general case that the signal has not a Gaussian PDF, an additional distortion of the PDF may occur. However, in case of non-linearities of the measurement stage, the shape of the PDF's $p_d(d)$ could be modified significantly—see Figs. 5(a) and (c)—leading thus to much lower values of the empirical CCF.

In general, if no knowledge concerning the non-linear mapping $f(\,)$ is available, the transform of Eq. (17) cannot be inverted. However, in some particular cases where the mapping $f(\,)$ is described by a monotonic curve, a histogram transform can be applied to the signals to assure that they both have the same PDF and, consequently, the same first-order statistics. In such cases, higher correlation values can be expected after the transform, if the signals being compared have a sufficient degree of likeness.

In the next subsections, two approaches allowing an equalization of the first-order statistics of two different signals are discussed. Both procedures are based on the impression of certain statistical properties to the signals to be compared. Whereas the first method performs a shift invariant transform of the global histogram, the second one represents a refinement of this strategy that aims to equalize the local PDF's of the signals to a given global PDF.

[‡]Since in this paper we focus on similarity between realizations of stochastic processes—i.e. between given signals $q_i(\xi)$—, the consideration of the two-dimensional joint probability density function is not possible.

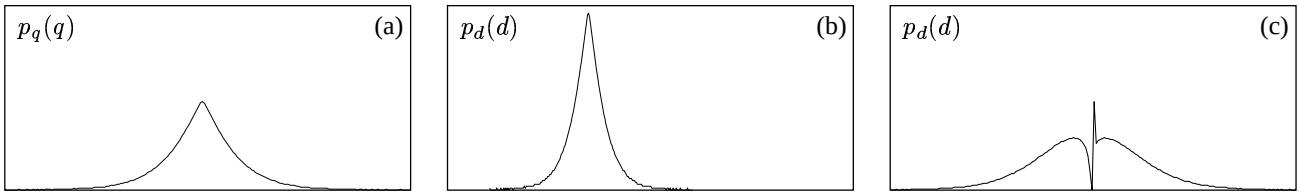

Figure 5. Examples of probability density functions (PDF's): (a) PDF $p_q(q)$ of a target signal $q(\xi)$; (b) PDF $p_d(d)$ of the signal $d(\xi)$ resulting from a linear transform of $q(\xi)$; (c) PDF $p_d(d)$ of the signal $d(\xi)$ resulting from a non-linear transform of $q(\xi)$.

4.1. Shift invariant transforms

To perform a shift invariant histogram transform, the amplitudes d of the sensor signal are to be mapped onto values $\hat{q}(d)$ such that $\hat{q}(d)$ is distributed according to $p_{\hat{q}}(\hat{q})$. Since the cumulative distribution functions (CDF's) $P_{\hat{q}}(\hat{q})$ and $P_d(d)$ must be equal at $\hat{q}(d)$, we obtain the desired mapping as[6]

$$\begin{aligned} P_{\hat{q}}(\hat{q}(d)) &= \int_{-\infty}^{\hat{q}(d)} p_{\hat{q}}(\beta)\,\mathrm{d}\beta \stackrel{!}{=} \int_{-\infty}^{d} p_d(\beta)\,\mathrm{d}\beta = P_d(d) \\ \Rightarrow \quad \hat{q}(d) &= P_{\hat{q}}^{-1}(P_d(d))\,, \end{aligned} \tag{19}$$

where $P_d(d)$ performs a transformation of d onto a random variable uniformly distributed on $[0,1]$, which $P_{\hat{q}}^{-1}(\,)$ then transforms onto a random variable distributed according to $p_{\hat{q}}(\hat{q})$; see Fig. 6.

To apply Eq. (19) to digital signals, we must consider that the signal values as well as the locations ξ are discrete. Furthermore, we will assume an ergodic signal generating process, so that the PDF can be estimated by means of a histogram $h(d)$ and a cumulative histogram $H(d)$:

$$h(d) = \frac{1}{|\mathcal{D}|}\sum_{\xi\in\mathcal{D}} \delta^d_{d(\xi)}\,, \tag{20}$$

$$H(d) = \sum_{\gamma=d_1}^{d} h(\gamma)\,, \tag{21}$$

$$d \in \{d_1, d_2, \ldots, d_D\}, \qquad \delta^b_a = \begin{cases} 1 & \text{for } a = b \\ 0 & \text{for } a \neq b \end{cases},$$

where $\mathcal{D}$ denotes the discrete support set of the signal $d(\xi)$ and consists of $|\mathcal{D}|$ elements. An approximate realization of the transform given in Eq. (19) is

$$\mathcal{T}\{d(\xi); H_{\text{target}}(d)\} = \arg \min_{\gamma\in\{d_1,\ldots,d_D\}} \{|H_{\text{target}}(\gamma) - H(d)|\}\,, \tag{22}$$

where $H_{\text{target}}(d)$ denotes an arbitrary global cumulative histogram to which the global intensity distribution is to be equal. Ideally, the choice of a proper target cumulative histogram $H_{\text{target}}(d)$ should be made taking the statistics of the signal $q(\xi)$ emitted by the source into account. However, since knowledge on the properties of $q(\xi)$ is not always available, a practicable alternative consists in transforming e.g. the signal $d_2(\xi)$ to adopt the histogram $H(d_1)$ of $d_1(\xi)$ while leaving the signal $d_1(\xi)$ unaltered:

$$\hat{q}_1(\xi) = d_1(\xi)\,, \qquad \hat{q}_2(\xi) = \mathcal{T}\{d_2(\xi); H(d_1)\}\,, \tag{23}$$

and to compute then the cross-correlation according to the Eqs. (1)–(2). An example of the latter strategy is given in Subsect. 4.3.

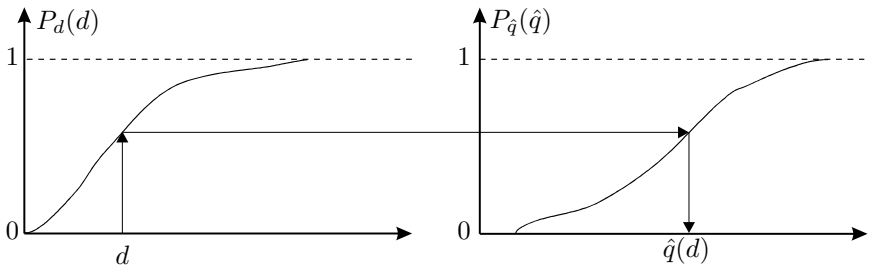

Figure 6. Mapping of the signal amplitudes d onto $\hat{q}(d)$.

4.2. Shift variant transforms

In case of performing shift variant histogram transforms, it is desirable that the first-order statistics are independent of ξ. This is equivalent to the requirement that all local PDF's $p_d(d;\xi)$ are equal to a global PDF $p_{\hat{q}}(\hat{q})$ not depending on ξ. Therefore, the histogram transform can be achieved by equating all local PDF's.

In contrast to the procedure described in the former section, the shift variance of the transform has to be taken into account. Here, the signal amplitudes d are to be mapped onto values $\hat{q}(d;\xi)$ such that $\hat{q}(d;\xi)$ is distributed according to $p_{\hat{q}}(\hat{q})$. The CDF's $P_{\hat{q}}(\hat{q})$ and $P_d(d;\xi)$ must be equal at $\hat{q}(d;\xi)$, so that in this case the following mapping is obtained:

$$\begin{aligned} P_{\hat{q}}(\hat{q}(d;\xi)) &= \int_{-\infty}^{\hat{q}(d;\xi)} p_{\hat{q}}(\beta)\,\mathrm{d}\beta \stackrel{!}{=} \int_{-\infty}^{d} p_d(\beta;\xi)\,\mathrm{d}\beta = P_d(d;\xi) \\ \Rightarrow \quad \hat{q}(d;\xi) &= P_{\hat{q}}^{-1}(P_d(d;\xi))\,. \end{aligned} \tag{24}$$

To apply this equation to digital signals, not only discrete signal values as well as locations ξ are to be considered, but also local ergodicity has to be postulated for the signal generating process. To this end, a neighborhood set $\mathcal{U}$ of the ξ origin is defined such that the local fluctuations of the statistical properties of the process are small enough that they can be assumed to be approximately constant within every neighborhood set:

$$\mathcal{U}(\xi) := \{\chi | \chi = \xi + \epsilon,\, \epsilon \in \mathcal{U}\}\,. \tag{25}$$

On the other side, $\mathcal{U}$ should be significantly larger than the largest details of the signal to avoid a deterioration of the signal of interest. Due to the local ergodicity assumed, the local PDF's can be estimated with local histograms $h(d;\xi,\mathcal{U})$ and local cumulative histograms $H(d;\xi,\mathcal{U})$:

$$h(d;\xi,\mathcal{U}) = \frac{1}{|\mathcal{U}|}\sum_{\epsilon\in\mathcal{U}} \delta^d_{d(\xi+\epsilon)}\,, \tag{26}$$

$$H(d;\xi,\mathcal{U}) = \sum_{\gamma=d_1}^{d} h(\gamma;\xi,\mathcal{U})\,, \tag{27}$$

where $|\mathcal{U}|$ denotes the size of the neighborhood set $\mathcal{U}$. For a discrete realization of the transform, the following approximation is obtained:

$$\mathcal{T}\{d(\xi); H_{\text{target}}(d),\mathcal{U}\} = \arg \min_{\gamma\in\{d_1,\ldots,d_D\}} \{|H_{\text{target}}(\gamma) - H(d;\xi,\mathcal{U})|\}\,, \tag{28}$$

where $H_{\text{target}}(d)$ denotes the arbitrary global cumulative histogram to be impressed.

The problems described in Subsect. 4.1 regarding a suitable choice of the target cumulative histogram $H_{\text{target}}(d)$ are also applicable to the case of a shift variant histogram transform, so that an analogous strategy can be employed here, if no explicit knowledge concerning the statistics of the target signal is available:

$$\hat{q}_1(\xi) = d_1(\xi)\,, \qquad \hat{q}_2(\xi) = \mathcal{T}\{d_2(\xi); H(d_1),\mathcal{U}\}\,. \tag{29}$$

The next section presents a successful application of this method to solve a forensic task—the automated comparison of striation patterns such as grooves on firearm bullets or toolmarks.

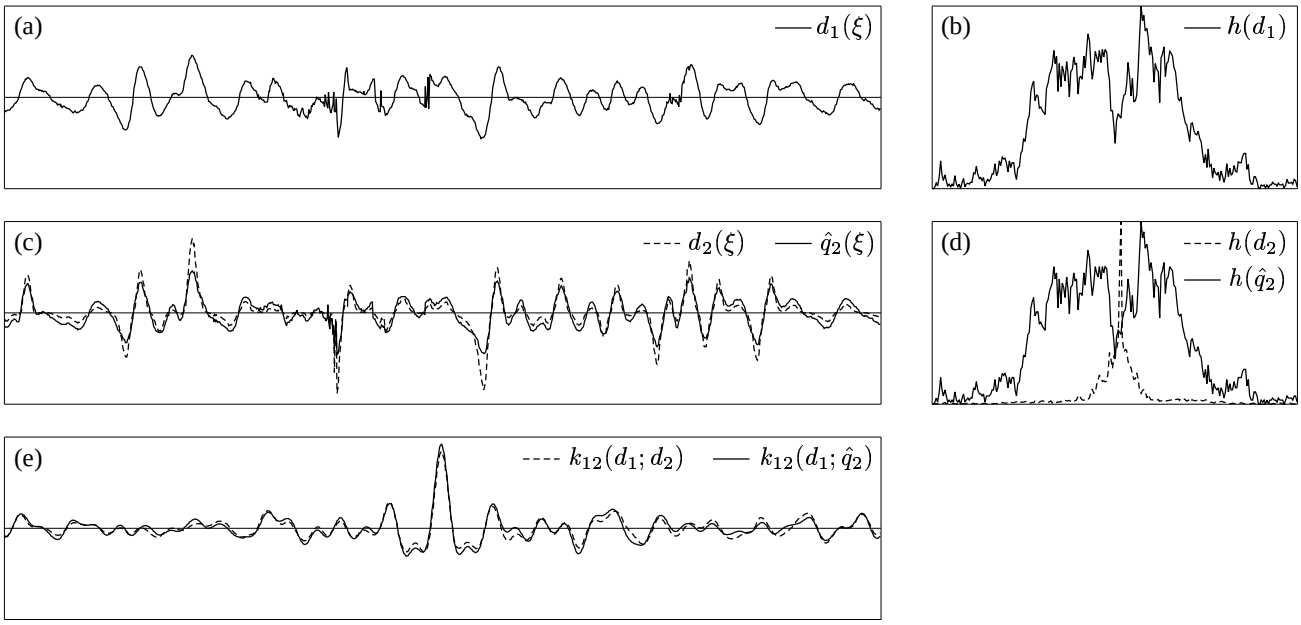

Figure 7. Detection results of a rail switch: (a) entry $d_1(\xi)$ of the digital road map, recorded with $z_1 = 120\,\mathrm{mm}$; (b) histogram $h(d_1)$; (c) sensor data $d_2(\xi)$ acquired with $z_2 = 80\,\mathrm{mm}$ and transformed signal $\hat{q}_2(\xi)$; (d) histograms $h(d_2)$ and $h(\hat{q}_2)$ of the sensor data and the transformed signal (normalized to the maximum value); (e) correlation results: $\rho_{12}(d_1; d_2) = 0.84$, $\rho_{12}(d_1; \hat{q}_2) = 0.92$.

4.3. Experimental results

The methodology proposed in Subsect. 4.1 is demonstrated with an application to determine the position of rail vehicles. A promising approach to enable a reliable positioning is to record certain characteristics of the rail track like switches, crossings etc. by means of train-based sensors and to compare them with a database called "digital road map" containing all characteristics of the route.[9] For the detection of characteristic metallic parts in the area of the rail track, eddy current sensors are placed in the vehicle at a defined distance z above the rail edge.[4] Due to the sensor characteristics, minor deviations from the calibrated sensor position result in a non-linear transform of the signal amplitudes, among other influences. Hence, to provide for a robust recognition of switches, a non-linear similarity between the sensor data acquired and the entries of the digital road map has to be tolerated.

To enable a fast detection, the empirical cross-correlation function is evaluated. In the common case of varying sensor distances z to the rail edge, the detection results can be improved significantly by applying a shift invariant transform according to Subsect. 4.1 to the sensor signal recorded. In Fig. 7, the principle of the switch recognition is illustrated together with the improvement achieved by the proposed strategy. Whereas the signal of the switch $d_1(\xi)$ stored in the digital road map has been recorded at a sensor distance of $z_1 = 120\,\mathrm{mm}$, the signal registered from the train when passing over the same switch has been acquired at a reduced distance of $z_2 = 80\,\mathrm{mm}$. To perform the shift-invariant transform, the sensor signal $d_2(\xi)$ is mapped onto the transformed signal $\hat{q}_2(\xi)$, so that its histogram $h(\hat{q}_2)$ equals the histogram $h(d_1)$ of the entry $d_1(\xi)$ of the digital road map. The enhancement obtained can be noticed in the maximum value of the CCF ρ_{12}, which is increased by 9%.

Finally, to demonstrate the performance of the proposed approach in practice, an application of the method based on the shift variant histogram transforms will be presented—the forensic comparison of striation patterns, such as grooves on bullets or toolmarks. For that purpose, a database consisting of 54 real toolmarks from 6 different tools was used. First of all, a high quality image of each mark was recorded by using an image acquisition system for forensic applications developed at the University of Karlsruhe.[12] Then, a meaningful one-dimensional signal called "signature" was obtained for each mark from the corresponding image by applying the strategies described in Ref. 7. An example of the signature extraction is depicted in Fig. 1 for the analogous case of striation marks on bullets. The signature contains all the relevant information on the mark and can thus be used for an efficient database search by means of the methods proposed in this paper. To evaluate the quality of the different measures of similarity, each mark is considered as an exhibit and compared

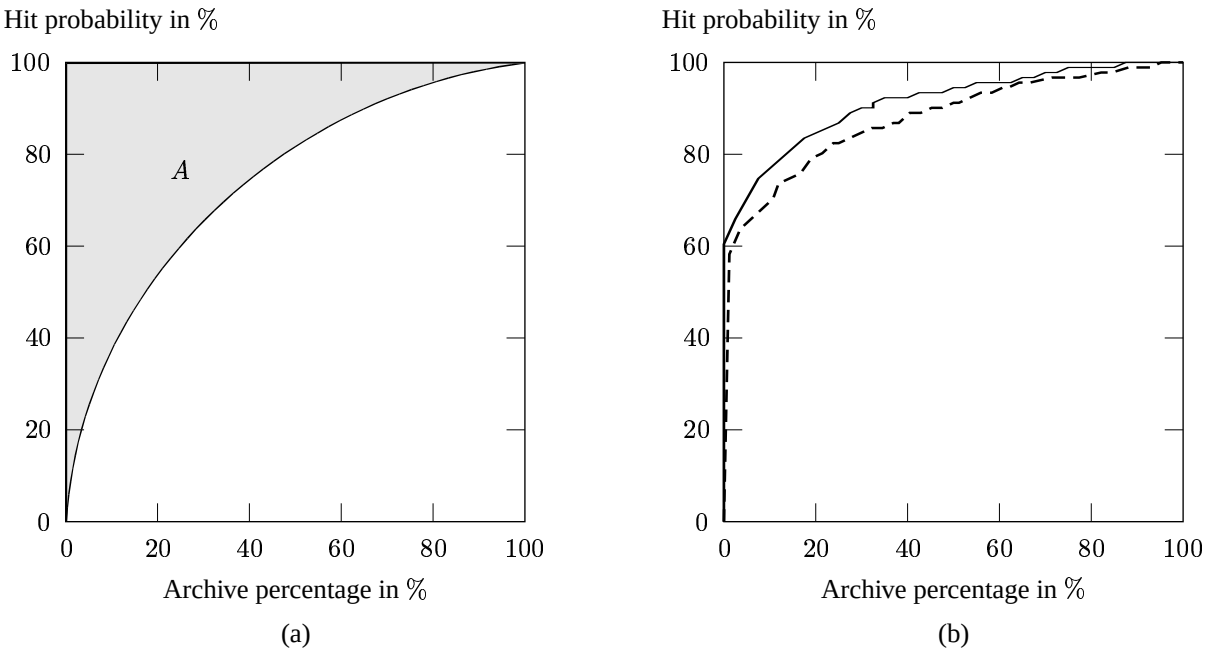

Figure 8. Results of the automated comparison of toolmarks: (a) assessment principle based on a concentration curve; (b) experimental results with (solid curve) and without (dashed curve) the method of the shift variant histogram transforms.

with all remaining marks of the database. Based on the results of each individual comparison, a sorted list is created, which provides a descending order of the different marks with respect to the similarity measure.

By evaluating all sorted lists generated by this procedure, a so-called *concentration curve* can be computed by plotting the cumulative probability of finding an actually existing match (i.e. a mark generated by the same tool as the exhibit) over the percentage of the sorted archive to be searched (i.e. the relative position of the hit list). Figure 8(a) presents an example of a concentration curve, where the grayed area A above the curve denotes the percentage of the archive to be examined on the average until the existing match is found. Thus, the measure A provides a means of evaluating the performance of the comparison of marks globally.

To investigate the benefit of the approach proposed, firstly the concentration curve as well as the global measure A have been calculated for the case of a direct cross-correlation of the signatures according to Sect. 2. The resulting concentration curve is plotted in Fig. 8(b, dashed curve) and the value of the area above the corresponding curve is $A = 14.7\%$. Thereafter, the statistics of the signal patterns were manipulated with the method described in Subsect. 4.2 to obtain a distribution function suitable for the comparison. Here, a generalized Gaussian distribution $p_x(x)$ was impressed locally to each signature:

$$p_x(x) = \frac{1}{C} \exp\left(\frac{x^n}{2\sigma^n}\right), \tag{30}$$

where C denotes a normalization factor, and an exponent of $n = 0.8$ was chosen. To avoid the influence of areas not containing marks, the size of the neighborhood set $\mathcal{U}$ was selected to 400 samples. Then, to measure the similarity of each combination of two signals, the maximum value of the CCF was computed. These parameters result in the concentration curve depicted in Fig. 8(b, solid curve), and they yield an area above the curve of $A = 11.6\%$. Obviously, since this represents a reduction of the comparison effort of more than 21% compared to the case of the purely linear procedure, this result is by all means satisfactory, and it proves that the conservative strategy of transforming non-linearly the signals to be analyzed can lead to a significant enhancement even in case of using large databases that could induce numerous false alarms.

5. CONCLUSIONS

In this paper, strategies for the detection of non-linear similarities have been presented. For the suitable description of such similarities, generalized types of likeness have been introduced. Several approaches based on energy minimization methods, cross-entropy analysis and non-linear signal transforms have been evaluated to deal with generalized similarities.

For the application of non-linear signal transforms, methods for the manipulation of first-order statistics have been proposed. By means of such transforms, non-linearities related to the data acquisition as well as the signal processing chain can be eliminated. This conservative approach is advantageous—especially when false alarms arising from a too weak concept of similarity are to be avoided.

To demonstrate the convenience of this approach, experimental results achieved for two different applications have been presented. First, a non-linear preprocessing has been applied to signals obtained from train-based sensors for the identification of rail switches. In this case, the non-linearities induced by changes of the sensor distance to the rail edge can efficiently be compensated by a shift invariant histogram transform.

The second practical example concerns the forensic comparison of striation marks. Here, characteristic signals are extracted from gray level images of the marks to be compared. Due to the non-linearities inherent to the interaction between illumination and surface reflectance as well as the strategy used to extract the signals of interest, a linear evaluation of similarity is not able to exploit all the information contained in these signals. However, the local impression of certain properties affecting first-order statistics by means of a non-linear histogram transform enables a significant increase of the recognition rate of the marks to be evaluated and thus represents an important contribution to the automation of forensic identification issues.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Tom Engelberg for making available the signals recorded with the eddy current sensors. Furthermore, they acknowledge the support of the Bundeskriminalamt as well as the LKA Berlin PTU for providing the archive of toolmarks used to investigate the performance of the presented methods. In addition, they would like to thank Christian Berger and Prof. Franz Mesch for helpful discussions and valuable comments on a draft version of this paper.

REFERENCES

1. J. Beyerer and F. Puente León, "Suppression of Inhomogeneities in Images of Textured Surfaces," *Opt. Eng.* **36** (1), 85–93, 1997.
2. E.O. Brigham, *The fast Fourier transform and its applications*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.
3. J.J. Clark and A.L. Yuille, *Data Fusion for Sensory Information Processing Systems*, Kluwer Academic Publishers, Boston, 1990.
4. T. Engelberg and F. Mesch, "Eddy current sensor system for non-contact speed and distance measurement of rail vehicles." In: *Computers in Railways VII*, pp. 1261–1270, WIT Press, Southampton, 2000.
5. R.M. Fano, *Transmission of Information*, MIT Press, New York, 1961.
6. R.C. González and P. Wintz, *Digital Image Processing*, Addison-Wesley, Reading, 1977.
7. M. Heizmann and F. Puente León, "Model-based analysis of striation patterns in forensic science." In: *Enabling Technologies for Law Enforcement and Security*, Proceedings of SPIE, Vol. 4232, pp. 533–544, 2001.
8. S. Kutscha, *Statistische Bewertungskriterien für die Entropieanalyse dynamischer Systeme*, VDI Verlag, Düsseldorf, 1989.
9. F. Mesch, F. Puente León and T. Engelberg, "Train-based location by detecting rail switches." In: *Computers in Railways VII*, pp. 1251–1260, WIT Press, Southampton, 2000.
10. A. Papoulis, *Probability, random variables and stochastic processes*, McGraw-Hill, New York, 1991.
11. M. Pelillo and E.R. Hancock (eds.), *Energy Minimization Methods in Computer Vision an Pattern Recognition*, Springer-Verlag, Berlin, 1997.
12. F. Puente León, *Automatische Identifikation von Schußwaffen*, VDI Verlag, Düsseldorf, 1999.
13. C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois, Urbana, 1949.