On a Non-monotonicity Effect of Similarity Measures

Bernhard Moser[1], Gernot Stübl[1], and Jean-Luc Bouchot[2]

[1] Software Competence Center Hagenberg,
Softwarepark 21, A-4232 Hagenberg
{bernhard.moser,gernot.stuebl}@scch.at
[2] Department of Knowledge-Based Mathematical Systems,
Johannes Kepler University, Altenbergerstr. 69, A-4040 Linz
jean-luc.bouchot@jku.at

Abstract. The effect of non-monotonicity of similarity measures is addressed which can be observed when measuring the similarity between patterns with increasing displacement. This effect becomes the more apparent the less smooth the pattern is. It is proven that commonly used similarity measures like f-divergence measures or kernel functions show this non-monotonicity effect which results from neglecting any ordering in the underlying construction principles. As an alternative approach Weyl's discrepancy measure is examined by which this non-monotonicity effect can be avoided even for patterns with high-frequency or chaotic characteristics. The impact of the non-monotonicity effect to applications is discussed by means of examples from the field of stereo matching, texture analysis and tracking.

Keywords: Kernel functions, f-divergence, discrepancy measure, Lipschitz property, stereo matching, texture analysis, tracking.

1 Introduction

This paper is devoted to the question whether similarity measures behave monotonically when applied to patterns with increasing displacement. Misalignment of patterns is encountered in various fields of applied mathematics, particularly signal processing, time series analysis or computer vision. Particularly when dealing with patterns with high frequencies the comparison of the shifted pattern with its reference will show ups and downs with respect to the resulting similarity values induced by commonly used similarity measures. More precisely, let us think of a pattern M as a function $v : X \subseteq \mathbb{R}^n \to \mathbb{R}$. A translational shift by a vector $\mathbf{t}$ induces a displaced pattern $M_{\mathbf{t}}$ represented by $v_{\mathbf{t}}(.) = v(. - \mathbf{t})$. In this paper we study the monotonicity behavior of similarity measures S as function $\Delta_S[v, \mathbf{t}](\lambda) = S(v_{\mathbf{0}}, v_{\lambda\mathbf{t}})$ depending on the displacement factor $\lambda \geq 0$ along the vector $\mathbf{t}$. If $\Delta_S[v, \mathbf{t}](.)$ is monotonically increasing for a class $\mathcal{V}$ of patterns $v \in \mathcal{V}$ for any direction $\mathbf{t}$ we say that the similarity measure S satisfies the monotonicity condition (MC) with respect to the class $\mathcal{V}$. Unless mentioning $\mathcal{V}$ explicitly we restrict to the class of patterns with non-negative entries with bounded support.

M. Pelillo and E.R. Hancock (Eds.): SIMBAD 2011, LNCS 7005, pp. 46–60, 2011.

As main theoretical contribution of this paper a mathematical analysis in Section 2 and Section 3 show how this effect follows from construction principles which neglect any ordering between the elements of the patterns. While Section 2 refers to similarity and distance measures which rely on the aggregation of an element-wise operating function, Section 3 is devoted to the class of f-divergence measures which evaluate the frequencies of single values $v(x)$ of the pattern. For both classes of similarity measure examples are presented that demonstrate the non-monotonicity effect. In Section 4 an alternative construction principle based on the evaluation of partial sums is introduced and recalled from previous work, particularly [Mos09]. Theoretical results show that the non-monotonicity effect can be avoided. Finally, in Section 5 the impact of the non-monotonicity effect to applications in the field of stereo matching, tracking and texture analysis is discussed.

2 Construction Principles of Similarity Measures Induced by the Aggregation of Element-Wise Operating Functions

The analysis of formal construction principles of similarity measures based on the composition of an element-wise operating function and an aggregation operation leads to elucidating counter examples showing that commonly used similarity measures in general are not monotonic with respect to the extent of displacement.

Therefore we will have a look at similarity measures from a formal construction point of view. For example let us consider the elementary inner product $< .,. >$ of Euclidean geometry which is defined as

$$< \boldsymbol{x}, \boldsymbol{y} >= \sum_i x_i \cdot y_i. \tag{1}$$

Formular (1) is constructed by means of a composition of the algebraic product which acts coordinate-wise and the summation as aggregation function. Formally, (1) therefore follows the construction principle

$$\Delta_{[\mathcal{A},\mathcal{C}]}(f,g) := \mathcal{A}_x(\mathcal{C}(f(x), g(x))), \tag{2}$$

where $\mathcal{C}$ and $\mathcal{A}$ denote the coordinate-wise operating function and the aggregation, respectively. f, g refer to vectors, sequences or functions with x as index or argument and the expression $\mathcal{A}_x$ means the aggregation of all admissible x. The elements f, g denote the elements from some admissible space $\Psi \subset \{f : X \to \mathbb{R}\}$ for which the formal construction yields well defined real values. For example, in the case of (1) the n-dimensional Euclidean space $\mathbb{R}^n$ for $n \in \mathbb{N}$ or the Hilbert space of square-integrable sequences l^2 would be admissible.

In the following we draw conclusions about the monotonicity behavior of the induced function (2) by imposing certain algebraic and analytic properties on the coordinate-wise operating function $\mathcal{C}$ and the aggregation $\mathcal{A}$.

Theorem 1. *The construction*

$$\Delta_{[\mathcal{A},\mathcal{S},\mathcal{C}]}(f,g) := \mathcal{S}(\mathcal{A}_x(\mathcal{C}(f(x),g(x)))) \tag{3}$$

induces a function

$$\Delta_{[\mathcal{A},\mathcal{S},\mathcal{C}]} : \Psi \times \Psi \to \mathbb{R}$$

that does not satisfy the monotonicity condition (MC) under the assumption that Ψ *is an admissible space of functions* $f : \mathbb{Z} \subseteq \mathbb{R} \to \mathbb{R}$ *that contains at least the set of pairwise differences of indicator functions of finite subsets of* $\mathbb{Z}$*,* $\mathcal{C} : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ *is a coordinate function that satisfies*

(C1) $\mathcal{C}$ *is commutative,*
(C2) $\mathcal{C}(1,0) \neq \mathcal{C}(1,1)$*,*
(C3) $\mathcal{C}(0,0) = \min\{\mathcal{C}(1,0), \mathcal{C}(1,1)\}$*,*

the aggregation function $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}$ *is*

(A1) *commutative and*
(A2) *strictly monotonically increasing or decreasing in each component, respectively,*

and the scaling function $\mathcal{S} : \mathbb{R} \to \mathbb{R}$ *is strictly monotonically increasing or decreasing in each component, respectively.*

Proof. Without loss of generality let us assume that the aggregation function is strictly monotonically increasing in each component.

We use the notation: $c_{01} := \mathcal{C}(0,1)$, $c_{11} := \mathcal{C}(1,1)$, $c_{00} := \mathcal{C}(0,0)$ and

$$h(.) := 1_{\{0\}}(.) + 1_{\{2\}}(.). \tag{4}$$

Consider

$$\begin{aligned}
\Delta_0 &= \Delta_{[\mathcal{A},\mathcal{S},\mathcal{C}]}(h(.), h(. - 0)) = \mathcal{S}(\mathcal{A}(c_{00}, \ldots, c_{00}, c_{11}, c_{00}, c_{11}, c_{00}, c_{00}, \ldots, c_{00})),\\
\Delta_1 &= \Delta_{[\mathcal{A},\mathcal{S},\mathcal{C}]}(h(.), h(. - 1)) = \mathcal{S}(\mathcal{A}(c_{00}, \ldots, c_{00}, c_{10}, c_{10}, c_{10}, c_{10}, c_{00}, \ldots, c_{00}))\\
\Delta_2 &= \Delta_{[\mathcal{A},\mathcal{S},\mathcal{C}]}(h(.), h(. - 2)) = \mathcal{S}(\mathcal{A}(c_{00}, \ldots, c_{00}, c_{10}, c_{00}, c_{11}, c_{00}, c_{10}, \ldots, c_{00})).
\end{aligned}$$

The case of $c_{10} < c_{11}$ implies $c_{00} = c_{01}$, hence a strictly increasing scaling function entails $\Delta_0 > \Delta_1 < \Delta_2$, and the case $c_{10} > c_{11}$, $c_{00} = c_{11}$ yields $\Delta_0 < \Delta_1 > \Delta_2$ which proves (4) to be a counter-example with respect to the monotonicity condition (MC). An analogous conclusion applies to a strictly decreasing scaling function. □

A direct consequence of Theorem 1 is that a binary operation $\odot : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ that preserves ordering in each argument, or reverses the ordering on both arguments, respectively, yields a further construction that cannot satisfy the monotonicity condition (MC) with respect to the class of patterns with non-negative entries with bounded support.

Corollary 1. *Let*

$$\Delta_1[\mathcal{A}_1, \mathcal{S}_1, \mathcal{C}_1](f,g) := \mathcal{S}_1(\mathcal{A}_{1x}(\mathcal{C}_1(f(x), g(x))))$$
$$\Delta_2[\mathcal{A}_2, \mathcal{S}_2, \mathcal{C}_2](f,g) := \mathcal{S}_2(\mathcal{A}_{2x}(\mathcal{C}_2(f(x), g(x))))$$

be functions following the construction principle (3) then

$$\Delta(f,g) = \Delta_1[\mathcal{A}_1, \mathcal{S}_1, \mathcal{C}_1](f,g) \odot \Delta_2[\mathcal{A}_2, \mathcal{S}_2, \mathcal{C}_2](f,g)$$

does not satisfy the monotonicity criterion (MC), where $\odot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *is an operation that is strictly monotonic of the same type in each component.*

Examples of similarity and distance measures following the construction principles of Theorem 1 or Corollary 1 are listed in Table 1.

Table 1. Examples of kernels and distance measures that follow the construction principles of Theorem 1 or Corollary 1 with summation as aggregation function

fomular	name	remark
$\|f-g\|_p$	Minkowski distance	$\mathcal{C}(a,b) = \|a-b\|^p$, $\mathcal{S}(x) = \sqrt[p]{x}$
$< f,g >= \sum_i f_i \cdot g_i$	inner product	$\mathcal{C}(a,b) = a \cdot b$
$e^{-\frac{1}{\sigma}\sum_i (f_i - g_i)^2}$	Gaussian kernel	$\mathcal{S}(x) = \exp(-x/\sigma)$
$-\sqrt{\|f-g\|^2 + c^2}$	multiquadratic	$\mathcal{S}(x) = -\sqrt{x+c^2}$
$\frac{1}{\sqrt{\|f-g\|^2+c^2}}$	inverse multiquadratic	$\mathcal{S}(x) = (\sqrt{x+c^2})^{-1}$
$\|f-g\|^{2n}\ln(\|f-g\|)$	thin plate spline	ln, x^n as scaling, $\odot(a,b) = a \cdot b$
$< f,g >^d, d \in \mathbb{N}$	polynomial kernel	(1) recursively applied, $\odot(a,b) = a \cdot b$
$(< f,g > +c)^d, d \in \mathbb{N}$	inh. polynomial kernel	(1) recursively applied, $\odot(a,b) = a \cdot b$
$\tanh(\kappa < x,y > +\theta)$	sigmoidal kernel	$\mathcal{S}(x) = \tanh(\kappa x + \theta)$

The following construction principle which does not require strictly monotonicity of the scaling function also leads to similarity measures that do not satisfy the monotonicity condition (MC).

Theorem 2. *The construction*

$$\Delta_{[\mathcal{A},\mathcal{S},\mathcal{C}]}(f,g) := \mathcal{S}(\mathcal{A}_x(\mathcal{C}(f(x), g(x)))) \qquad (5)$$

induces a function

$$\Delta_{[\mathcal{A},\mathcal{S},\mathcal{C}]} : \Psi \times \Psi \to \mathbb{R}$$

that does not satisfy the monotonicity condition under the assumption that Ψ *is an admissible space of functions* $f : \mathbb{Z} \subseteq \mathbb{R} \to \mathbb{R}$ *that contains at least the set of pairwise differences of scaled indicator functions of finite subsets of* $\mathbb{Z}$, $\mathcal{C} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *is a continuous coordinate function that satisfies*

(C'1) *$\mathcal{C}$ is commutative,*
(C'2) *$\mathcal{C}(0,.)$ is strictly monotonic,*
(C'3) *$\forall\alpha : \mathcal{C}(\alpha,\alpha) = 0$,*

the aggregation function $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}$ is continuous and satisfies (A1) and (A2) of Theorem 1 and the scaling function $\mathcal{S} : \mathbb{R} \to \mathbb{R}$ is

(S1) *continuous,*
(S2) *non-trivial in the sense that it is not constant on the range*

$$\mathcal{R} = \{\mathcal{A}(\mathcal{C}(\alpha,0),\mathcal{C}(\alpha,0),0,\ldots,0) \in \mathbb{R}_0^+ : \alpha \in \mathbb{R}\}.$$

Proof. Without loss of generality $0 = \mathcal{A}_x(0) = \mathcal{A}(0,\ldots,0)$. Set

$$h(.) := a \cdot 1_{\{0\}}(.) + b \cdot 1_{\{2\}}(.)$$

and, let us denote $\Theta_t(a,b) = \mathcal{A}_x(\mathcal{C}(f(x), f(x-t)))$.

Then, by applying (A1) we obtain

$$\begin{aligned}
\Theta_0(a,b) &= \mathcal{A}(0,\ldots,0,\mathcal{C}(0,0),\mathcal{C}(0,0),\mathcal{C}(0,0),\mathcal{C}(0,0),0,\ldots,0)\\
\Theta_1(a,b) &= \mathcal{A}(0,\ldots,0,\mathcal{C}(a,0),\mathcal{C}(a,0),\mathcal{C}(b,0),\mathcal{C}(b,0),0,\ldots,0)\\
\Theta_2(a,b) &= \mathcal{A}(0,\ldots,0,\mathcal{C}(a,0),\mathcal{C}(a,b),\mathcal{C}(b,0),\mathcal{C}(0,0),0,\ldots,0).
\end{aligned}$$

Let $\zeta \in \mathcal{R}$, $\zeta > 0$, and note that there is $a_0 > 0$ such that $\Theta_1(a_0, 0) = \zeta$. Observe that

$$\forall a \in [0, a_0]\, \exists\, b_a \in [0, a_0] : \Theta_1(a, b_a) = \zeta.$$

Let $\gamma_\zeta = \{(a, b_a) : \Theta_1(a, b_a) = \zeta\}$. Further, note that

$$\forall (a, b_a) \in \gamma_\zeta, a > 0 \,:\, \Theta_2(a, b_a) < \Theta_1(a, b_a) = \zeta$$

and

$$\underbrace{\lim_{a \to a_0^-} \Theta_2(a, b_a)}_{\Theta_2(a_0,0)} = \underbrace{\lim_{a \to a_0^-} \Theta_1(a, b_a)}_{\Theta_1(a_0,0)}.$$

Then $\forall \varepsilon > 0\, \exists \xi \in (\zeta - \varepsilon, \zeta)\, \exists (a_\xi, b_\xi) \in \gamma_\zeta$ we have

$$\xi = \Theta_2(a_\xi, b_\xi) < \Theta_1(a_\xi, b_\xi) = \zeta. \tag{6}$$

Let $s_0 = \mathcal{S}(0)$. Without loss of generality $s_0 > 0$. As $\mathcal{S}$ is not constant on $\mathcal{R}$, there is a $\zeta \in \mathcal{R}$ such that

$$s = \mathcal{S}(\zeta) \neq \mathcal{S}(0) = s_0.$$

Hence $\zeta > 0$. Without loss of generality $s < s_0$. Let $\xi = \inf\{\zeta > 0 : \mathcal{S}(\zeta) \leq s\}$. The continuity assumption of $\mathcal{S}$ implies $\xi > 0$. By (6) for all $n \in \mathbb{N}$ there is $\xi_n \in (\xi - \frac{1}{n}, \xi)$ for which there is $(a_n, b_n) \in \gamma_\xi$ with

$$\xi_n = \Theta_2(a_n, b_n) < \Theta_1(a_n, b_n) = \zeta.$$

Note that $\forall n \in \mathbb{N}: \mathcal{S}(\xi_n) > \mathcal{S}(\xi)$ and $\lim_n \mathcal{S}(\xi_n) = \mathcal{S}(\xi)$. Therefore, there is a n_0 with $\mathcal{S}(\xi_{n_0}) \in (s, s_0)$. For an illustration of the construction of ξ_{n_0} see Figure 2. By construction, for

$$h_0(.) := a_{n_0} \cdot 1_{\{0\}}(.) + b_{n_0} \cdot 1_{\{2\}}(.)$$

we obtain

$$0 = \Theta_0(a_{n_0}, b_{n_0}) < \Theta_2(a_{n_0}, b_{n_0}) < \Theta_1(a_{n_0}, b_{n_0})$$

which shows that the monotonicity condition (MC) cannot by satisfied, as

$$\mathcal{A}_x(\mathcal{C}(h_0(x), h_0(x-0))) < \mathcal{A}_x(\mathcal{C}(h_0(x), h_0(x-1))) > \mathcal{A}_x(\mathcal{C}(h_0(x), h_0(x-2)))$$

□

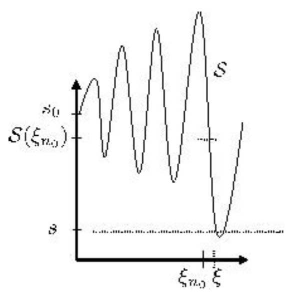


Fig. 1. Illustration of construction of ξ_{n_0}

Examples of similarity measures that meet the conditions of Theorem 5 are translational invariant kernels $\Phi(x, y) = \Phi(\|x - y\|)$ where $\phi : [0, \infty) \to \mathbb{R}$ is a continuous function that results from a Bessel transform of a finite non-negative Borel measure μ on $[0, \infty)$, i.e. $\phi(r) = \int_0^\infty \Omega_s(rt)d\mu(t)$ where $\Omega_1(r) = \cos r$ and $\Omega_s(r) = \Gamma(\frac{s}{2})\frac{s}{2}^{(s-2)/2} J_{(s-2)/2}(r)$, $s \geq 2$ and $J_{(s-2)/2}$ is the Bessel function of first kind of order $\frac{s-2}{2}$. For example there is the Dirichlet kernel $k(x, y) = \Phi_D(x)(\|x - y\|)$ provided by the continuous function $\Phi_D(x) = \sin\left((2n+1) \cdot \frac{x}{2}\right) / \sin\left(\frac{x}{2}\right)$ or the B_n-spline kernels $k(x, y) = B_{2p+1}(\|x - y\|)$ that result from multiple convolution of indicator functions, $B_n = \otimes_{i=1}^n 1_{[-\frac{1}{2}, \frac{1}{2}]}$, where the positive definite kernel property is only satisfied by odd orders.

For details on kernels and particularly translational invariant kernels see e.g. [SS01].

3 *f*-Divergence Measures

In this Section we concentrate on histogram based measures, see e.g. [TJ91]. The most prominent one is the *mutual information*, which for two discrete random variables X and Y can be defined as

$$I(X;Y) = \sum_{x,y} P_{XY}(x, y) \log\left(\frac{P_{XY}(x, y)}{P_X(x)P_Y(y)}\right) \tag{7}$$

where P_{XY} is the joint probability distribution of X and Y, and P_X and P_Y are the marginal probability distribution of X and Y respectively. This measure is commonly used in various fields of applications as for example in registering images, see e.g. [GGL08], [LZSC08]. Equation (7) is a special case of *Kullback-Leibler divergence*, [Kul59],

$$D_{\mathrm{KL}}(P\|Q) = \sum_z P(z) \log \left(\frac{P(z)}{Q(z)} \right) \tag{8}$$

which measures the deviation between the probability distributions P and Q. The mutual information is regained from (8) by setting $z = (x, y)$, $P(x, y) = P_{XY}(x, y)$ and $Q(x, y) = P_X(x)P_Y(y)$. A further generalization is provided by the class of f-divergence measures $D_{\mathrm{f}}(P\|Q)$, see e.g. [DD06, LV06], defined by

$$D_{\mathrm{f}}(P\|Q) = \sum_z Q(z) \mathrm{f} \left(\frac{P(z)}{Q(z)} \right) \tag{9}$$

where $f : [0, \infty] \to \mathbb{R} \cup \{+\infty\}$ is convex and continuous. These measures were introduced and studied independently by [Csi63], [Mor63] and [AS96]. The Kullback-Leibler divergence (8) results from (9) by means of $f(t) = t \log(t)$.

Theorem 3. *Let $f : [0, \infty] \to \mathbb{R} \cup \{+\infty\}$ be a strictly convex and continuous function. For two discrete sequences $A = (a_i)_{i=1}^n \in \mathcal{V}^n$ and $B = (b_i)_{i=1}^n \in \mathcal{V}^n$, $n \in \mathbb{N}$ let*

$$D_{\mathrm{f}}(A\|B) = \sum_{v,w \in \mathcal{V}} P_A(v)P_B(w) f \left(\frac{P_{A\,B}(v, w)}{P_A(v)P_B(w)} \right) \tag{10}$$

where $P_{A\,B}(v, w)$ denotes the joint frequency of occurrence of the pair of values (v, w), and $P_A(v)$, $P_B(w)$ denote the frequencies of v, w in the corresponding sequences A and B, respectively. Then there are sequences $h : \mathbb{Z} \to \mathcal{V}$ such that $\chi : \mathbb{N} \to [0, \infty]$ given by

$$\chi_t = D_f(A_0, A_t)$$

does not behave monotonically with respect to t, where $A_t(.) = 1_{1,\ldots,n}(.) \cdot h(. - t)$.

Proof. Set $\mathcal{V} = \{0, 1\}$, and define $h(.) := \sum_{j=1}^m 1_{\{2 \cdot j\}}(.)$ where $m \in \mathbb{N}$. Set $n = K \cdot m$ with $K \geq 3$. Then

$$P_{A_t}(0) = \tfrac{n-m}{n}, \; P_{A_t}(1) = \tfrac{n-m}{n}$$

for $t \in \{0, 1, 2\}$, further

$$\begin{array}{llll} P_{A_0,A_0}(0,0) = \frac{n-m}{n}, & P_{A_0,A_0}(0,1) = 0, & P_{A_0,A_0}(1,0) = 0, & P_{A_0,A_0}(1,1) = \frac{m}{n}, \\ P_{A_0,A_1}(0,0) = \frac{n-2m}{n}, & P_{A_0,A_1}(0,1) = \frac{m}{n}, & P_{A_0,A_1}(1,0) = \frac{m}{n}, & P_{A_0,A_1}(1,1) = 0, \\ P_{A_0,A_2}(0,0) = \frac{n-m}{n}, & P_{A_0,A_2}(0,1) = \frac{1}{n}, & P_{A_0,A_3}(1,0) = \frac{1}{n}, & P_{A_0,A_2}(1,1) = \frac{m-2}{n}. \end{array}$$

By taking $n = K \cdot m$ into account we get

$$\frac{n^2}{m^2}\chi_0(K,m) = f\left(\frac{K}{K-1}\right)(K-1)^2 + 2f(0)(K-1) + f(K),$$
$$\frac{n^2}{m^2}\chi_1(K,m) = f\left(\frac{(K-2)K}{(K-1)^2}\right)(K-1)^2 + 2f\left(\frac{K}{K-1}\right)(K-1) + f(0),$$
$$\frac{n^2}{m^2}\chi_2(K,m) = f\left(\frac{K}{K-1}\right)(K-1)^2 + 2f\left(\frac{K}{K-1}\frac{1}{m}\right)(K-1) + f\left(K\frac{m-2}{m}\right).$$

Observe that because of the continuity of f for all $K \geq 2$ we obtain

$$\lim_{m\to\infty}(\chi_0(K,m) - \chi_2(K,m)) = 0. \tag{11}$$

As

$$\frac{(K-1)^2 - 2(K-1)}{(K-1)^2} + \frac{2(K-1)-1}{(K-1)^2} + \frac{1}{(K-1)^2} = 1$$

and

$$\frac{(K-2)K}{(K-2)^2}\frac{(K-1)^2 - 2(K-1)}{(K-1)^2} \cdot \frac{K}{K-1} + \frac{2(K-1)-1}{(K-1)^2} \cdot 0 + \frac{1}{(K-1)^2} \cdot K$$

the strict convexity of f implies

$$f\left(\frac{(K-2)K}{(K-2)^2}\right)$$
$$= \frac{(K-1)^2 - 2(K-1)}{(K-1)^2} \cdot f\left(\frac{K}{K-1}\right) + \frac{2(K-1)-1}{(K-1)^2} \cdot f(0) + \frac{1}{(K-1)^2} \cdot f(K)$$

and, therefore, for all $m > 2$ it follows that

$$\chi_0(K,m) - \chi_2(K,m) = \varepsilon_K > 0. \tag{12}$$

Together, formulae (11) and (12) imply that there are indices K_0 and m_0 such that $\chi_0(K_0,m_0) > \chi_1(K_0,m_0) < \chi_2(K_0,m_0)$ which proves the claim. □

Finally let us remark that an analogous proof shows that the claim of Theorem 3 is also true if the histograms P_X and P_Y are compared directly in the sense of definition (9).

4 The Monotonicity Property of the Discrepancy Measure

The concept of discrepancy measure was proposed by Hermann Weyl [Wey16] in the early 20-th century in order to measure deviations of distributions from uniformity. For details see, e.g. [BC09, Doe05, KN05]. Applications can be found in the field of numerical integration, especially for Monte Carlo methods in high dimensions, see e.g. [Nie92, Zar00, TVC07] or in computational geometry, see e.g. [ABC97, Cha00, KN99]. For applications to data storage problems on parallel disks see [CC04, DHW04] and half toning for images see [SCT02].

In the image processing context of registration and tracking, the discrepancy measure is applied in order to evaluate the auto-misalignment between a pattern P with its translated version P_T with lag or shift T. The interesting point about this is that based on Weyl's discrepancy concept distance measures can be constructed that guarantee the desirable registration properties: (R1) the measure vanishes if and only if the lag vanishes, (R2) the measure increases monotonically with an increasing lag, and (R3) the measure obeys a Lipschitz condition that guarantees smooth changes also for patterns with high frequencies. As the discrepancy measure as defined by (13)

$$\|\boldsymbol{f}\|_D := \sup \left\{ \left| \sum_{i=m_1}^{m_2} f_i \right| : m_1, m_2 \in \mathbb{Z} \right\} \tag{13}$$

induces a norm on the space of vectors $\boldsymbol{f} = (f_1, \dots, f_n) \in \mathbb{R}^n$ in the geometric sense we further on refer to it as discrepancy norm. As pointed out in [Mos09] Equation (13) is equivalent to

$$\|\boldsymbol{f}\|_D := \max(0, \max_{1 \le k \le n} \sum_{i=0}^{k} f_i) - \min(0, \min_{1 \le k \le n} \sum_{i=0}^{k} f_i) \tag{14}$$

which is advantageous in terms of computational complexity which amounts to $O(n)$ in comparison with $O(n^2)$ of the original definition (13). Note that the only arithmetical operations in the algorithm are summation, comparisons and inversion which on the one side are fast to compute and on the other side cheap in hardware design. In the context of this paper its dependency on the ordering of the elements is worth mentioning which is illustrated by the examples $\|(1, -1, 1)\|_D = 1$ and $\|(-1, 1, 1)\|_D = 2$. Note that alternating signals like $(-1, 1, -1, \dots)$ lead to small discrepancy values, while reordering the signal e.g. in a monotonic way maximizes it.

As outlined in [Mos09] Equation (13) can be extended and generalized to arbitrary finite Euclidean spaces equipped with some measure μ in the following way:

$$\|f\|_{\mathcal{C}}^{(d)} = \sup_{c \in \mathcal{C}} \left| \int_c f d\mu \right| \tag{15}$$

where $\mathcal{C}$ refers to a set of Cartesian products of intervals. For example, let $\mathcal{B}^d$ denote the set of d-dimensional open boxes $I_1 \times I_2 \times \cdots \times I_d$ with open intervals I_i from the extended real line $[-\infty, \infty]$, and $\widetilde{\mathcal{B}}^d \subset \mathcal{B}^d$ the set of Cartesian products of intervals of the form $]-\infty, x[$, $]x, \infty[$. It can be shown that for all $d \in \mathbb{N}$ and non-negative $f \in \mathcal{L}(\mathbb{R}^d, \mu)$, $f \ge 0$, there holds

$$\|f - f \circ T_{\mathbf{t}}\|_{\mathcal{B}^d}^{(d)} = \|f - f \circ T_{\mathbf{t}}\|_{\widetilde{\mathcal{B}}^d}^{(d)} \tag{16}$$

where $T_{\mathbf{t}} = \mathtt{x} - \mathtt{t}$. Formulae 16 can be expressed by means of integral images and their higher dimensional variants which is crucial in terms of efficient computation. With this definitions the following result can be proven, for details and the proof see [Mos09].

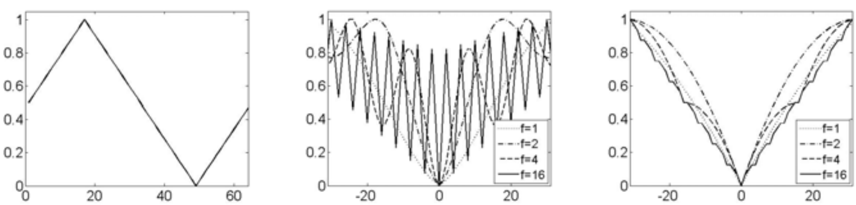
Theorem 4. *Let* $d \in \mathbb{N}$, *let* $f \in \mathcal{L}(\mathbb{R}^d, \mu)$, $f \geq 0$ *and let* $\Delta_{\mathcal{C}}[f](\mathbf{t}) = \|f - f \circ T_{\mathbf{t}}\|_{\mathcal{C}}$ *denote the misalignment function* $\mathbf{t} \in \mathbb{R}^d$. *Further, let*

$$\delta_\mu[f](\mathbf{t}) = \sup_{C \in \mathcal{C}} \max\{\mu(C \backslash T_{\mathbf{t}}(C)), \mu(T_{\mathbf{t}}(C) \backslash C)\}.$$

Then for $\mathcal{C} = \mathcal{B}^d$ *or* $\mathcal{C} = \widetilde{\mathcal{B}}^d$ *we have*

1. *If* f *is non-trivial, i.e.,* $\int |f| d\mu > 0$ *then* $\Delta_C[f](\mathbf{t}) = 0 \Longleftrightarrow \mathbf{t} = 0$
2. *Lipschitz property:* $\Delta_C[f](\mathbf{t}) \leq \delta_\mu[f](\mathbf{t}) \|f\|_\infty$.
3. *Monotonicity:* $0 \leq \lambda_1 \leq \lambda_2 \Longrightarrow \Delta_C[f](\lambda_1 \mathbf{t}) \leq \Delta_C[f](\lambda_2 \mathbf{t})$ *for arbitrary* $\mathbf{t} \in \mathbb{R}^d$.

Figure 2 illustrates the principle difference between the characteristics of the resulting misalignment functions induced by a measure, in this case normalized cross-correlation, that shows the non-monotonicity artefact on the one hand and the discrepancy norm on the other hand.



(a) (b) (c)

Fig. 2. Figure (a) shows a sawtooth function with frequency $\omega = 1$. In the other two figures misalignment functions for this sawtooth function and its variants with higher frequencies, $\omega = 2, 4, 16$ with respect to one minus the normalized cross-correlation, Figure (b), and the discrepancy norm, Figure (c), are shown. With increasing frequencies of the in Figure (b) the In contrast to Figure (b) the discrepancy norm induced misaligment functions in (c) show a monotonic behaviour with bounded slope due to the Lipschitz property.

5 Impact of the Non-monotonicity Effect on Applications

Misalignment is a phenomenon which can be observed in numerous situations in applied mathematics. In this paper we concentrate on examples from image processing in order to illustrate the relevance and impact of the monotonicity and Lipschitz property of the discrepancy measure in comparison to commonly used measures for which these properties cannot be guaranteed.

5.1 Image Tracking

Image tracking aims at identifying and localizing the movement of a pattern along a sequence of images. In this context a commonly used similarity measure is the so-called Bhattacharyya coefficient [Bha43] defined by

$$D_B(P_X, P_Y) = \sum_x \sqrt{P_X(x) P_Y(x)}. \quad (17)$$

See [CRM00] for details in the context of tracking. Note that $-D_B(P_X, P_Y)$ turns out to be a special f-divergence measure by means of $f(u) = -\sqrt{x}$.

Figure 3 depicts the cost functions of a person track on the CAVIAR (Context Aware Vision using Image-based Active Recognition) [1] database based on the discrepancy norm (second row) and the Bhattacharyya coefficient. It interesting to observe the robustness of the discrepancy norm at the presence of massive noise.

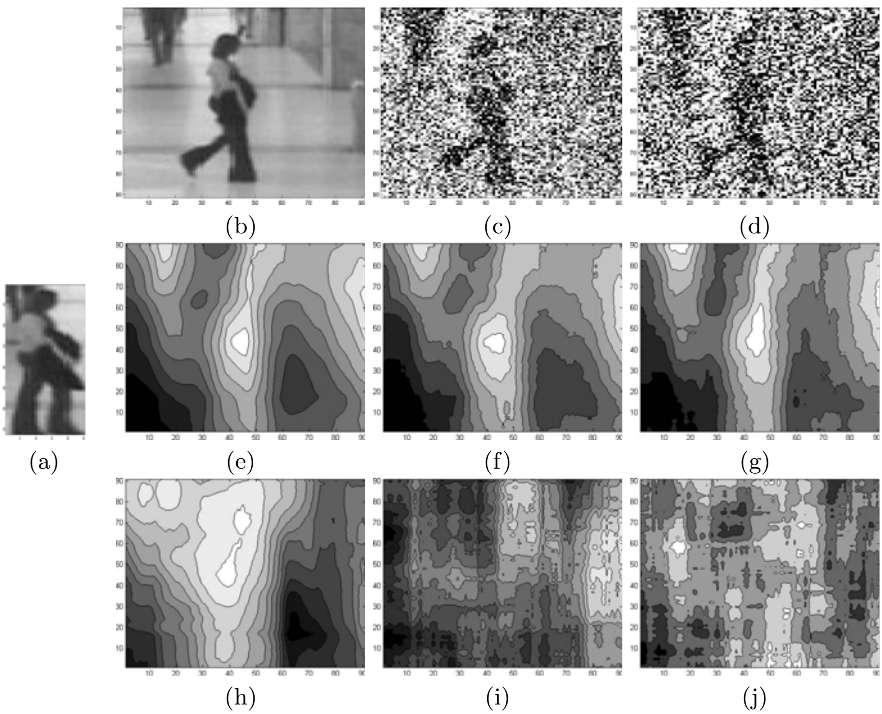


Fig. 3. Tracking of female from Figure 2(a) in a consecutive frame Figure (2b) and the same frame corrupted with additive gaussian noise with $SNR = 3$ in Figure 2(c) and $SNR = 1.5$ in Figure 2(d). Figures 2(e), (f) and (g) depict the corresponding cost function based on the discrepancy norm (DN) as similarity, whereas (h), (i) and (j) refer to the Bhattacharyya coefficient based similarity. The images are taken from frame 697 and frame 705 of the EC Funded CAVIAR project/IST 2001 37540 ("Shopping Center in Portugal", "OneLeaveShop2cor").

[1] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/

5.2 Stereo Matching

Cost estimation in stereo matching is crucial for stereo vision, see [SS02]. Figure 4 illustrates the working principle of a typical stereo matching algorithm: the content of the white window in Figure 4(a) is compared with the windows along the white line in Figure 4(b). Figure 4(c) plots the comparison results with different matching cost functions. The x-value with the lowest dissimilarity is finally taken as disparity from which depth information can be derived.

Typically the sum of absolute distances (SAD), sum of squared distances (SSD) or cross correlation as well as their normalized and zero mean variants (NCC, ZSAD, etc.) are used as dissimilarity measures in this context. However these cost functions follow the construction principle of Equation (5) and suffer therefore from non-monotonic behaviour. Especially when adding white noise to the source images the number of local minima of these matching cost functions increase, whereas the discrepancy norm keeps mainly its monotonic behaviour, see Figure 4(c).

5.3 Defect Detection in Textured Surfaces

In the context of quality control typically reference image patches are compared to image patches which result from a sliding window procedure. For a discussion on similarity and a template matching based approach for detecting defects in regularly textured images see [BSM11]. Here an example is presented that

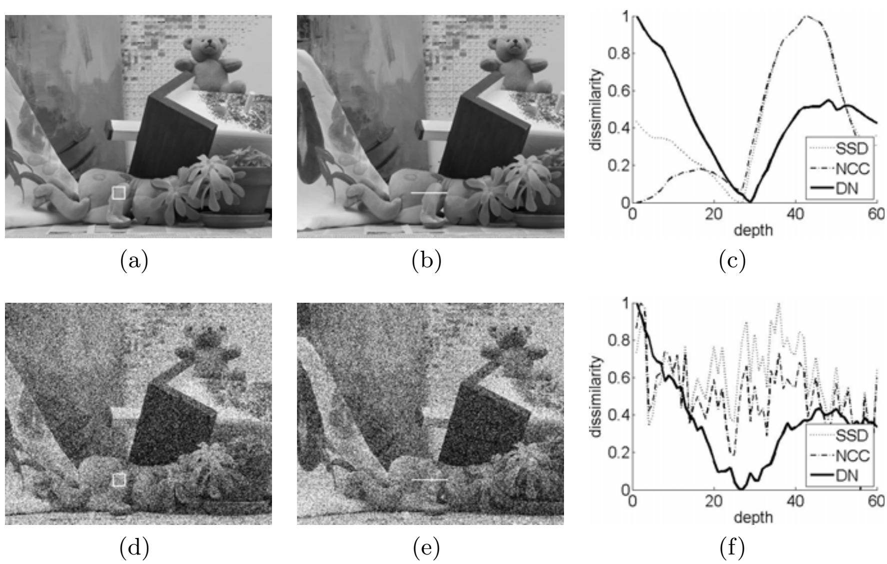


Fig. 4. Matching cost evaluation of sum of squared differences (SSD), normalized cross correlation (NCC) and discrepancy norm (DN) evaluated on the Middlebury Stereo 2003 Dataset [SS03], Teddy Example, at position x=192/y=300 with windowsize 10, depth 60. Figure 4(c) shows the evaluation of the white patch in Figures 4(a) along the white line in Figure 4(b). Whereas the results of Figures 4(d) and 4(e) with $SNR = 6.1$ are shown in Figure 4(f). DN is more robust regarding noise than the other costs.

demonstrates the behaviour of similarity measures showing the non-monotonicity effect versus the discrepancy norm. Fig. 5(b) depicts an example taken from the TILDA database [2]. As the presented texture shows a repetitive pattern it allows to apply a pattern matching approach and to compute the dissimilarity given some translational parameters. A defect-free pattern, depicted in Figure 5(a), is considered as a reference pattern and is then translated along the textured image. Each t_x and t_y displacement induces a dissimilarity value as illustrated in Figure 5(d) where the ordinate refers to the dissimilarity value. Observe the distinct local minumum of the discrepancy norm even in the presence of noise in Figures 5(e) and 5(f).

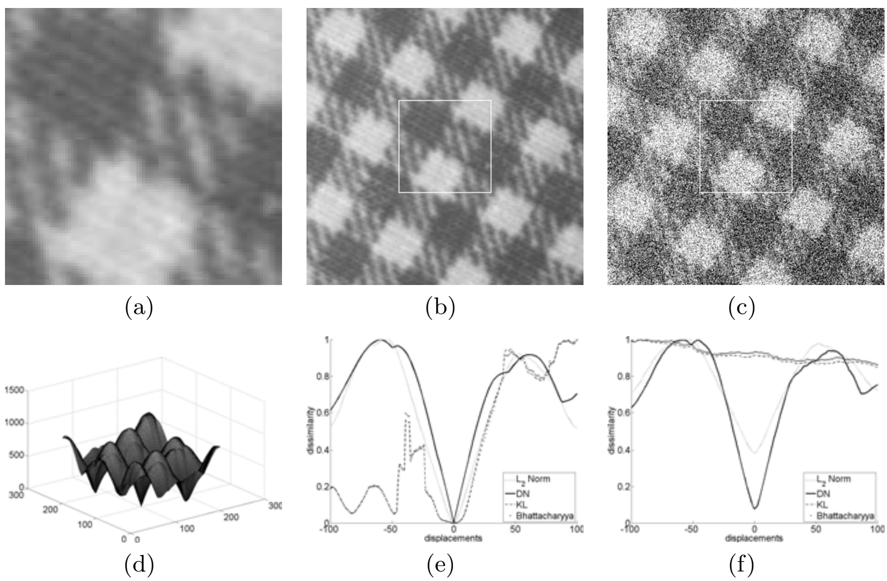


Fig. 5. Template matching example for regularly textured images. A defect-free reference template is shown in Figure (a) with corresponding patches (white square) in a noise free and a corrupted image by added white Gaussian noise, Figure (b) and Figure (c), respectively. Figure (d) plots a surface of dissimilarity values between the reference and the patches of Figure (b). Figures (e) (noise-free) and (f) (gaussian noise) show the behaviour of different cost functions along the x-axis: discrepancy norm (solid), L_2 norm (dotted), Bhattacharyya measure (square plotted) and mutual information (dashed-dotted).

6 Conclusion and Future Work

A non-monotonicity effect of commonly used similarity measures has been examined in the context of misaligned patterns. As it was shown this non-monotonicity

[2] Available from Universität Freiburg, Institut für Informatik, Lehrstuhl für Mustererkennung und Bildverarbeitung (LMB).

effect is caused by certain underlying construction principles. As the application section demonstrates this effect is worth thinking about for example in order to reduce local minima in resulting cost functions e.g. in the context of stereo matching. It remains future work to elaborate alternative similarity concepts as for instance based on Weyl's discrepancy measure to come up with cost functions that avoid the artefacts from the non-monotonicity effect.

Acknowledgement. This work was supported in part by the Austrian Science Fund (FWF) under grant no. P21496 N23 and the Austrian COMET program.

References

[ABC97] Alexander, J.R., Beck, J., Chen, W.W.L.: Geometric discrepancy theory and uniform distribution, pp. 185–207. CRC Press, Inc., Boca Raton (1997)

[AS96] Ali, S.M., Silvey, S.D.: A General Class of Coefficients of Divergence of One Distribution from Another. J. Roy. Statist. Soc. Ser. B 28, 131–142 (1996)

[BC09] Beck, J., Chen, W.W.L.: Irregularities of Distribution. Cambridge University Press, Cambridge (2009)

[Bha43] Bhattacharyya, A.: On a measure of divergence between two statistical populations defined by probability distributions. Bull. Calcutta Math. 35, 99–109 (1943)

[BSM11] Bouchot, J.-L., Stübl, G., Moser, B.: A template matching approach based on the discrepancy norm for defect detection on regularly textured surfaces. Accepted to Quality Control by Artificial Vision Conference, QCAV 2011 (June 2011)

[CC04] Chen, C.M., Cheng, C.T.: From discrepancy to declustering: Near-optimal multidimensional declustering strategies for range queries. J. ACM 51(1), 46–73 (2004)

[Cha00] Chazelle, B.: The Discrepancy Method: Randomness and Complexity. Cambridge University Press, Cambridge (2000)

[CRM00] Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean shift, vol. 2, pp. 142–149 (2000)

[Csi63] Csiszár, I.: Eine informationstheoretische Ungleichung und ihre anwendung auf den Beweis der ergodizität von Markoffschen Ketten. Publ. Math. Inst. Hungar. Acad. 8, 95–108 (1963)

[DD06] Deza, M., Deza, E.: Dictionary of Distances. Elsevier, Amsterdam (2006)

[DHW04] Doerr, B., Hebbinghaus, N., Werth, S.: Improved bounds and schemes for the declustering problem. In: Fiala, J., Koubek, V., Kratochvíl, J. (eds.) MFCS 2004. LNCS, vol. 3153, pp. 760–771. Springer, Heidelberg (2004)

[Doe05] Doerr, B.: Integral approximations. Habilitation thesis, University of Kiel (2005)

[GGL08] Gao, Z., Gu, B., Lin, J.: Monomodal image registration using mutual information based methods. Image and Vision Computing 26(2), 164–173 (2008)

[KN99] Kuipers, L., Niederreiter, H.: Geometric Discrepancy: An Illustrated Guide. Algorithms and combinatorics, vol. 18. Springer, Berlin (1999)

[KN05] Kuipers, L., Niederreiter, H.: Uniform distribution of sequences. Dover Publications, New York (2005)

[Kul59] Kullback, S.: Information Theory and Statistics. Wiley, New York (1959)

[LV06] Liese, F., Vajda, I.: On Divergences and Informations in Statistics and Information Theory. IEEE Transactions on Information Theory 52(10), 4394–4412 (2006)

[LZSC08] Lu, X., Zhang, S., Su, H., Chen, Y.: Mutual information-based multi-modal image registration using a novel joint histogram estimation. Computerized Medical Imaging and Graphics 32(3), 202–209 (2008)

[Mor63] Morimoto, T.: Markov processes and the h-theorem. Journal of the Physical Society of Japan 18(3), 328–331 (1963)

[Mos09] Moser, B.: Similarity measure for image and volumetric data based on Hermann Weyl's discrepancy measure. IEEE Transactions on Pattern Analysis and Machine Intelligence (2009), doi:10.1109/TPAMI.2009.50

[Nie92] Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. Society for Industrial and Applied Mathematics, Philadelphia (1992)

[SCT02] Sadakane, K., Chebihi, N.T., Tokuyama, T.: Discrepancy-based digital halftoning: Automatic evaluation and optimization. In: WTRCV 2002, pp. 173–198 (2002)

[SS01] Schölkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, 1st edn. Adaptive Computation and Machine Learning. The MIT Press, Cambridge (2001)

[SS02] Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Hingham, MA, USA, vol. 47, pp. 7–42. Kluwer Academic Publishers, Dordrecht (2002)

[SS03] Scharstein, D., Szeliski, R.: High-Accuracy Stereo Depth Maps Using Structured Light. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 195–202 (June 2003)

[TJ91] Thomas, C.M., Joy, T.A.: Elements of Information Theory, 1st edn., p. 18. John Wiley & Sons, Inc., Chichester (1991)

[TVC07] Takhtamysheva, G., Vandewoestyneb, B., Coolsb, R.: Quasi-random integration in high dimensions. Image Vision Comput. 73(5), 309–319 (2007)

[Wey16] Weyl, H.: Über die Gleichverteilung von Zahlen mod. Eins. Math. Ann. 77, 313–352 (1916)

[Zar00] Zaremba, S.K.: The mathematical basis of Monte Carlo and Quasi-Monte Carlo methods. SIAM Review 10(3), 303–314 (1968)