Comments on "On approximating Euclidean metrics by weighted t-cost distances in arbitrary dimension"

M. Emre Celebi [a,*], Hassan A. Kingravi [b], Fatih Celiker [c]

[a] Department of Computer Science, Louisiana State University, Shreveport, LA, USA
[b] School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA
[c] Department of Mathematics, Wayne State University, Detroit, MI, USA

ARTICLE INFO

ABSTRACT

Mukherjee [Mukherjee, J., 2011. On approximating Euclidean metrics by weighted t-cost distances in arbitrary dimension. Pattern Recognition Lett. 32, 824–831] recently introduced a class of distance functions called weighted t-cost distances that generalize m-neighbor, octagonal, and t-cost distances. He proved that weighted t-cost distances form a family of metrics and derived an approximation for the Euclidean norm in $\mathbb{Z}^n$. In this note we compare this approximation to two previously proposed Euclidean norm approximations and demonstrate that the empirical average errors given by Mukherjee are significantly optimistic in $\mathbb{R}^n$. We also propose a simple normalization scheme that improves the accuracy of his approximation substantially with respect to both average and maximum relative errors.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The Minkowski ($\mathcal{L}_p$) metric is inarguably one of the most commonly used quantitative distance (dissimilarity) measures in scientific and engineering applications. The Minkowski distance between two vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ in the n-dimensional Euclidean space, $\mathbb{R}^n$, is given by

$$\mathcal{L}_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}. \tag{1}$$

Three special cases of the $\mathcal{L}_p$ metric are of particular interest, namely, $\mathcal{L}_1$ (city-block metric), $\mathcal{L}_2$ (Euclidean metric), and $\mathcal{L}_\infty$ (chessboard metric). Given the general form (1), $\mathcal{L}_1$ and $\mathcal{L}_2$ can be defined in a straightforward fashion, while $\mathcal{L}_\infty$ is defined as

$$\mathcal{L}_\infty(\mathbf{x}, \mathbf{y}) = \max_{1 \leqslant i \leqslant n} |x_i - y_i|.$$

The Minkowski metric enjoys the property of being translation invariant, i.e., $\mathcal{L}_p(\mathbf{x}, \mathbf{y}) = \mathcal{L}_p(\mathbf{x} + \mathbf{z}, \mathbf{y} + \mathbf{z})$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$. Since in many applications the data space is Euclidean, the most natural choice of metric is $\mathcal{L}_2$, which has the added advantage being isotropic (rotation invariant). For example, when the input vectors stem from an isotropic vector field, e.g., a velocity field, the most appropriate choice is to use the $\mathcal{L}_2$ metric so that all vectors are processed in the same way, regardless of their orientation (Barni

et al., 1995). However, $\mathcal{L}_2$ has the drawback of a high computational cost due to the multiplication and square root operations. As a result, $\mathcal{L}_1$ and $\mathcal{L}_\infty$ are often used as alternatives. Although these metrics are computationally more efficient, they deviate from $\mathcal{L}_2$ significantly.

Due to the translation invariance of $\mathcal{L}_p$, it suffices to consider $D_p(\mathbf{x}) = \mathcal{L}_p(\mathbf{x}, \mathbf{0})$, i.e., the distance from the point $\mathbf{x}$ to the origin. Therefore, in the rest of the paper, we will consider approximations to $D_p(\mathbf{x})$ rather than $\mathcal{L}_p(\mathbf{x}, \mathbf{y})$.

Let $\widetilde{D}$, defined on $\mathbb{R}^n$, be an approximation to D_2 (Euclidean norm). We assume that $\widetilde{D}$ is a continuous and absolutely homogeneous function. Recall that $\widetilde{D}$ is called absolutely homogeneous (of degree one) if $\widetilde{D}(\lambda \mathbf{x}) = |\lambda| \widetilde{D}(\mathbf{x}) \; \forall \lambda \in \mathbb{R}, \; \forall \mathbf{x} \in \mathbb{R}^n$.

We note that all variants of $\widetilde{D}$ we consider in this paper satisfy these assumptions. As a measure of the quality of the approximation of $\widetilde{D}$ to D_2 we define the maximum relative error (MRE) as

$$\varepsilon_{\max}^{\widetilde{D}} = \sup_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\left| \widetilde{D}(\mathbf{x}) - D_2(\mathbf{x}) \right|}{D_2(\mathbf{x})}. \tag{2}$$

Using the homogeneity of D_2 and $\widetilde{D}$, (2) can be written as

$$\varepsilon_{\max}^{\widetilde{D}} = \sup_{\mathbf{x} \in S_2^{n-1}} \left| \widetilde{D}(\mathbf{x}) - 1 \right|, \tag{3}$$

where $S_2^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : D_2(\mathbf{x}) = 1\}$ is the unit hypersphere of $\mathbb{R}^n$ with respect to the Euclidean norm. Furthermore, by the continuity of $\widetilde{D}$, we can replace the supremum with maximum in (3) and write

* Corresponding author.
E-mail addresses: ecelebi@lsus.edu (M.E. Celebi), kingravi@gatech.edu (H.A. Kingravi), celiker@math.wayne.edu (F. Celiker).

$$\widetilde{\varepsilon}_{max}^{D} = \max_{\mathbf{x} \in S_2^{n-1}} \left| \widetilde{D}(\mathbf{x}) - 1 \right|. \tag{4}$$

We will use (4) as the definition of MRE throughout.

Mukherjee (2011) recently introduced a class of distance functions called weighted t-cost distances that generalize m-neighbor (Das et al., 1987), octagonal (Das et al., 1987), and t-cost (Das et al., 1992) distances. He proved that weighted t-cost distances form a family of metrics and derived an approximation for the Euclidean norm in $\mathbb{Z}^n$. Here we briefly review the t-cost norm.

The t-cost norm (Das et al., 1992) defines two points in the rectangular grid as neighbors when their respective hypercubes (or hypervoxels) share a hyperplane of any dimension. The cost associated with these points can be at most t, $1 \leqslant t \leqslant n$, such that if two consecutive points on a shortest path share a hyperplane of dimension r, the distance between them is taken as $\min(t, n - r)$. There are n distinct t-cost norms defined by

$$D_t(\mathbf{x}) = \sum_{i=1}^{t} x_{(i)}, \quad 1 \leqslant t \leqslant n,$$

where $x_{(i)}$ is the i-th absolute largest component of $\mathbf{x}$, i.e., $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is a permutation of $(|x_1|, |x_2|, \ldots, |x_n|)$ such that $x_{(1)} \geqslant x_{(2)} \geqslant \cdots \geqslant x_{(n)}$. The MRE of this norm is given by Das et al. (1992)

$$\varepsilon_{max}^{D_t} = \max \left(\sqrt{t} - 1, 1 - \frac{t}{\sqrt{n}} \right).$$

Mukherjee generalized the t-cost norm as follows (Mukherjee, 2011):

$$D_M(\mathbf{x}) = \max_{1 \leqslant t \leqslant n} \{ w_t D_t(\mathbf{x}) \},$$

where w_t's are non-negative real constants. Based on this weighted norm, he then derived an approximation for D_2 using the following weight assignment: $w_t = 1\sqrt{t}$ for $1 \leqslant t \leqslant n$. Note that D_M consistently underestimates D_2 and the corresponding MRE is given by Mukherjee (2011)

$$\varepsilon_{max}^{D_M} = 1 - \frac{1}{\sqrt{\sum_{i=1}^{n} (\sqrt{i} - \sqrt{i-1})^2}}. \tag{5}$$

In a recent study (Celebi et al., 2011), we examined various Euclidean norm approximations in detail and compared their average and maximum errors using numerical simulations. Here we show that two of those approximations, namely Barni et al.'s norm (Barni et al., 1995, 2000) and Seol and Cheun's norm (Seol and Cheun, 2008), are viable alternatives to D_M.

Barni et al. (1995, 2000) formulated a generic approximation for D_2 as

$$D_B(\mathbf{x}) = \delta \sum_{i=1}^{n} \alpha_i x_{(i)},$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and $\delta > 0$ are approximation parameters. Note that a non-increasing ordering and strict positivity of the component weights, i.e., $\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_n > 0$ is a necessary and sufficient condition for D_B to define a norm (Barni et al., 2000).

Barni et al. showed that the minimization of (4) is equivalent to determining the weight vector $\boldsymbol{\alpha}$ and the scale factor δ that solve the following minimax problem:

$$\min_{\boldsymbol{\alpha}, \delta} \max_{\mathbf{x} \in V} |D_B(\mathbf{x}) - 1|,$$

where $V = \{ \mathbf{x} \in \mathbb{R}^n : x_1 \geqslant x_2 \geqslant \cdots \geqslant x_n \geqslant 0, D_2(\mathbf{x}) = 1 \}$. The optimal solution and its MRE are given by

$$\alpha_i^* = \sqrt{i} - \sqrt{i-1}, \quad \delta^* = \frac{2}{1 + \sqrt{\sum_{i=1}^{n} \alpha_i^{*2}}}, \quad \varepsilon_{max}^{D_B} = 1 - \delta^*. \tag{6}$$

Note the striking similarity between (5) and (6). Interestingly, a similar but less rigorous approach had been published earlier by Ohashi (1994). It should also be noted that several authors approached the problem from a Euclidean distance transform perspective and derived similar approximations for the 2- and 3-dimensional cases, see for example (Borgefors, 1986; Verwer, 1991). Furthermore, computation of weighted (Chamfer) distances in arbitrary dimensions on general point lattices is discussed in (Fouard et al., 2007).

More recently, Seol and Cheun (2008) proposed an approximation of the form

$$D_{a,b}(\mathbf{x}) = a D_\infty(\mathbf{x}) + b D_1(\mathbf{x}), \tag{7}$$

where a and b are strictly positive parameters to be determined by solving the following 2×2 linear system

$$a \mathrm{E}\left(D_\infty^2\right) + b \mathrm{E}(D_\infty D_1) = \mathrm{E}(D_2 D_\infty),$$
$$a \mathrm{E}(D_\infty D_1) + b \mathrm{E}\left(D_1^2\right) = \mathrm{E}(D_2 D_1),$$

where $\mathrm{E}(\cdot)$ is the expectation operator.

Seol and Cheun estimated the optimal values of a and b using 100,000 n-dimensional vectors whose components are independent and identically distributed, standard Gaussian random variables. In (Celebi et al., 2011), we demonstrated that a fixed number of samples from the unit hypersphere gives biased estimates for the MRE. The basic reason behind this is the fact that a fixed number of samples fail to suffice as the dimension of the space increases.

It is easy to see that D_B and $D_{a,b}$ fit into the general form

$$\widetilde{D}(\mathbf{x}) = \sum_{i=1}^{n} w_i x_{(i)},$$

which is a weighted D_1 norm. For D_B the weights are $w_1 = \delta^*$ and $w_{i \neq 1} = \delta^* \alpha_i^*$, whereas for $D_{a,b}$ they are $w_1 = a + b$ and $w_{i \neq 1} = b$. Clearly, D_B has a more elaborate design in which each component is assigned a weight proportional to its ranking (absolute value). However, this weighting scheme also presents a drawback in that a full ordering of the component absolute values is required.

D_B and $D_{a,b}$ can also be written as linear combinations of the D_1 and D_∞ norms, as in (7). D_1 overestimates the D_2 norm, whereas D_∞ underestimates it (Chaudhuri et al., 1992). Therefore, it is natural to expect a suitable linear combination of D_1 and D_∞ to give an approximation to D_2 better than either of them (Rhodes, 1995). Note that Rosenfeld and Pfaltz (1968) obtained a 2-dimensional
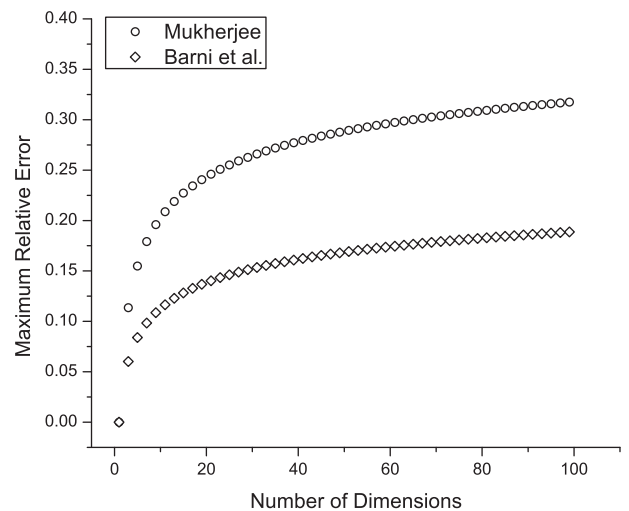

Fig. 1. Maximum relative errors for D_M and D_B.

approximation by combining D_1 and D_∞ nonlinearly as follows: $\widetilde{D}(\mathbf{x}) = \max(\lfloor 2(D_1(\mathbf{x}) + 1)/3 \rfloor, D_\infty(\mathbf{x}))$.

2. Comparison of the Euclidean norm approximations

Due to their formulations, the MREs for D_M and D_B can be calculated analytically using (5) and (6), respectively. In Fig. 1 we plot the theoretical errors for these norms for $n \leqslant 100$. It can be seen that D_B is not only more accurate than D_M, but also it scales significantly better.

The operation counts for each norm are given in Table 1 (**ABS**: absolute value, **COMP**: comparison, **ADD**: addition, **MULT**: multiplication, **SQRT**: square root). The following conclusions can be drawn:

▷ D_B and D_M have the highest computational cost due to the fact that they require sorting of the absolute values of the vector components.
▷ $D_{a,b}$ has the lowest computational cost among the approximate norms. A significant advantage of this norm is that it requires only two multiplications regardless of the value of n.
▷ $D_{a,b}$ can be used to approximate D_2^2 (squared Euclidean norm) using an extra multiplication. On the other hand, the computational cost of D_B (D_M) is higher than that of D_2^2 due to the extra absolute value and sorting operations involved.

In Table 2 we display the percentage average and maximum errors for $D_{a,b}, D_B$, and D_M for $n \leqslant 8$. Average relative error (ARE) is defined as

$$\varepsilon_{\text{avg}}^{\widetilde{D}} = \frac{1}{|S|} \sum_{\mathbf{x} \in S} \left| \widetilde{D}(\mathbf{x}) - 1 \right|,$$

where S is a finite subset of the unit hypersphere S_2^{n-1}, and $|S|$ denotes the number of elements in S. An efficient way to pick a random point on S_2^{n-1} is to generate n independent Gaussian random variables $x_1, x_2, \ldots, x_n$ with zero mean and unit variance. The distribution of the unit vectors

$$\left\{ \mathbf{y} = (y_1, y_2, \ldots, y_n) : y_i = x_i \bigg/ \left(\sum_{j=1}^{n} x_j^2 \right)^{1/2}, \quad i = 1, 2, \ldots, n \right\}$$

will then be uniform over the surface of the hypersphere (Muller, 1959). For each approximate norm, the ARE and MRE values were calculated over an increasing number of points, $2^{20}, 2^{21}, \ldots$ (that are uniformly distributed on the hypersphere) until the error values converge, i.e., the error values do not differ by more than $\epsilon = 10^{-5}$ in two consecutive iterations.

In Table 2, the error values under the column "$D_M(\mathbb{R}^n)$" were obtained using the aforementioned iterative scheme, whereas those under the column "$D_M(\mathbb{Z}^n)$" are taken from Mukherjee (2011). Motivated by the fact that D_M consistently underestimates D_2, we also experimented with a normalized form of this approximate norm given by $D_{\widehat{M}}(\mathbf{x}) = D_M(\mathbf{x})/\delta^*$. Note that $\delta^* < 1$ for $n \geqslant 2$ (6).

Note that for D_M and D_B, two types of maximum error were considered: empirical maximum error (MRE$_e$), which is calculated numerically over S and the theoretical maximum error (MRE$_t$), which is calculated analytically using (5) and (6), respectively.

By examining Table 2, the following observations can be made regarding the maximum error:

▷ The most accurate approximation is D_B. This is because this norm is designed to minimize the maximum error.
▷ The proposed normalization is quite effective since the resulting norm, $D_{\widehat{M}}$, is, on the average, only 8.6% less accurate than D_B, whereas both $D_M(\mathbb{R}^n)$ and $D_M(\mathbb{Z}^n)$ are, on the average, about 85% less accurate than D_B.
▷ The least accurate approximations are D_M and $D_{a,b}$ for $n \leqslant 4$ and $n > 4$, respectively.
▷ As n is increased, the error increases in all approximations. However, as can also be seen in Fig. 1, the error grows faster in some approximations than others.
▷ For D_B, the empirical and theoretical errors agree almost perfectly in all cases, which demonstrates the validity of the presented iterative error calculation scheme. As for D_M, the agreement in each case is close, but not as close as that observed in D_B. We have confirmed that using a smaller convergence threshold (ϵ) alleviates this problem at the expense of increased computational cost.

On the other hand, with respect to average error we can see that:

▷ $D_{a,b}$ is the most accurate approximation. This is because this norm is designed to minimize the average error.
▷ $D_M(\mathbb{R}^n)$ and $D_M(\mathbb{Z}^n)$ are the least accurate approximations. Furthermore, the errors given by Mukherjee are lower than those that we obtained (over $\mathbb{R}^n$), and the discrepancy between the outcomes of the two error calculation schemes increases as n is increased. The optimistic average error values given by

Table 1
Operation counts for the norms.

Norm	ABS	COMP	ADD	MULT	SQRT
D_∞	n	$n-1$	0	0	0
D_1	n	0	$n-1$	0	0
D_2	0	0	$n-1$	n	1
D_B	n	$\mathcal{O}(n\log n)$	$n-1$	n	0
$D_{a,b}$	n	$n-1$	n	2	0
D_M	n	$\mathcal{O}(n\log n)$	$n-1$	n	0

Table 2
Percentage average and maximum errors for the approximate Euclidean norms.

n	$D_{a,b}$		D_B			$D_{\widehat{M}}$		D_M				
								$\mathbb{R}^n$		$\mathbb{Z}^n$		
	ARE	MRE$_e$	ARE	MRE$_e$	MRE$_t$	ARE	MRE$_e$	ARE	MRE$_e$	ARE	MRE$_e$	MRE$_t$
2	2.00	5.25	2.41	3.96	3.96	2.48	4.12	2.55	7.61	2.40	7.61	7.61
3	2.39	9.98	3.00	6.02	6.02	2.97	6.40	4.14	11.35	3.63	11.35	11.35
4	2.57	13.64	3.44	7.39	7.39	3.28	7.97	5.21	13.75	4.29	13.75	13.75
5	2.68	16.59	3.77	8.39	8.39	3.53	9.16	5.98	15.47	4.65	15.46	15.49
6	2.73	18.88	4.01	9.19	9.19	3.73	10.12	6.55	16.80	4.85	16.79	16.83
7	2.76	20.67	4.18	9.84	9.84	3.92	10.91	7.00	17.90	5.00	17.86	17.92
8	2.77	21.92	4.31	10.39	10.39	4.10	11.59	7.35	18.78	5.04	18.75	18.82

Table 3
Percentage average and maximum errors for $D_{\widehat{M}}$.

n	$D_M^{\delta^*}$			$D_{\widehat{M}}^{\hat{\delta}}$		
	ARE	MRE_e	δ^*	ARE	MRE_e	$\hat{\delta}$
2	2.48	4.12	0.960434	2.41	3.96	0.961971
3	2.97	6.40	0.939809	2.79	6.02	0.943192
4	3.28	7.97	0.926150	2.99	7.39	0.931336
5	3.53	9.16	0.916059	3.13	8.40	0.922654
6	3.73	10.12	0.908117	3.23	9.18	0.915927
7	3.92	10.91	0.901603	3.31	9.84	0.910619
8	4.10	11.59	0.896101	3.40	10.39	0.905850

Mukherjee are due to the fact that his approximation was primarily intended for use in digital geometry and hence the calculations were performed in $\mathbb{Z}^n$ (rather than $\mathbb{R}^n$) using a very small number of points ranging from 32 to 512 (Mukherjee, 2011). In fact, Mukherjee used progressively fewer points with increasing n to calculate the error values. In (Celebi et al., 2011), we demonstrated that more points are required in higher dimensions to obtain unbiased error estimates.

In the calculation of $D_{\widehat{M}}$, we assumed that the optimal scaling factor for D_M is the same as that of D_B, i.e., δ^*. In order to check this assumption, we performed a one-dimensional grid search over $[\delta^*,1]$ for each n value. The results are shown in Table 3. It can be seen that:

▷ $D_{\widehat{M}}^{\hat{\delta}}$ is significantly more accurate than $D_M^{\delta^*}$ with respect to both ARE and MRE.

▷ $D_{\widehat{M}}^{\hat{\delta}}$ and D_B have almost identical MREs. Since D_B is analytically optimized for the maximum error it can be concluded that $D_{\widehat{M}}^{\hat{\delta}}$ can reach the same optimality by means of a suitable scaling factor.

▷ Interestingly, $D_{\widehat{M}}^{\hat{\delta}}$ is more accurate than D_B with respect to ARE. This could be due to the fact that the two approximations take different paths towards minimizing the MRE.

3. Conclusions

In this paper, we examined the weighted t-cost norm recently proposed by Mukherjee (2011) with respect to its ability to approximate the Euclidean norm in $\mathbb{R}^n$. We evaluated the average and maximum errors of this norm using numerical simulations and compared the results to those of two other well-known Euclid-ean norm approximations. The results demonstrated that, because it was designed for digital geometry applications in $\mathbb{Z}^n$, the original weighted t-cost norm is not particularly suited to approximate the Euclidean norm in $\mathbb{R}^n$. It is also shown, however, that when normalized with an appropriate scaling factor, Mukherjee's norm becomes competitive with an analytically optimized approximation with respect to both average and maximum relative errors.

References

Barni, M., Bartolini, F., Buti, F., Cappellini, V., 1995. Optimum linear approximation of the Euclidean norm to speed up vector median filtering, In: Proc. of the 2nd IEEE Internat. Conf. on Image Processing, pp. 362–365.
Barni, M., Buti, F., Bartolini, F., Cappellini, V., 2000. A quasi-Euclidean norm to speed up vector median filtering. IEEE Trans. Image Process. 9 (10), 1704–1709.
Borgefors, G., 1986. Distance transformations in digital images. Comput. Vision Graphics Image Process. 34 (3), 344–371.
Celebi, M.E., Celiker, F., Kingravi, H., 2011. On Euclidean Norm Approximations. Pattern Recognition 44 (2), 278–283.
Chaudhuri, D., Murthy, C.A., Chaudhuri, B.B., 1992. A modified metric to compute distance. Pattern Recognition 25 (7), 667–677.
Das, P.P., Chakrabarti, P.P., Chatterji, B., 1987. Generalized distances in digital geometry. Inform. Sci. 42 (1), 51–67.
Das, P.P., Chakrabarti, P.P., Chatterji, B., 1987. Distance functions in digital geometry. Inform. Sci. 42 (2), 113–136.
Das, P.P., Mukherjee, J., Chatterji, B.N., 1992. The t-cost distance in digital geometry. Inform. Sci. 59 (1), 1–20.
Fouard, C., Strand, R., Borgefors, G., 2007. Weighted distance transforms generalized to modules and their computation on point lattices. Pattern Recognition 40 (9), 2453–2474.
Mukherjee, J., 2011. On approximating Euclidean metrics by weighted t-cost distances in arbitrary dimension. Pattern Recognition Lett. 32 (6), 824–831.
Muller, M.E., 1959. A note on a method for generating points uniformly on N-dimensional spheres. Comm. ACM 2 (4), 19–20.
Ohashi, Y., 1994. Fast Linear Approximations of Euclidean Distance in Higher Dimensions. In: Heckbert, P. (Ed.), Graphics Gems, vol. IV. Academic Press.
Rhodes, F., 1995. On the metrics of Chaudhuri Murthy and Chaudhuri. Pattern Recognition 28 (5), 745–752.
Rosenfeld, A., Pfaltz, J.L., 1968. Distance functions on digital pictures. Pattern Recognition 1 (1), 33–61.
Seol, C., Cheun, K., 2008. A low complexity Euclidean norm approximation. IEEE Trans. Signal Process. 56 (4), 1721–1726.
Verwer, B.J.H., 1991. Local distances for distance transformations in two and three dimensions. Pattern Recognition Lett. 12 (11), 671–682.