RESEARCH

Open Access

CrossMark

On the translation-invariance of image distance metric

Bing Sun[1*], Jufu Feng[1] and Guoping Wang[2]

*Correspondence:
bsun@pku.edu.cn
[1] Key Laboratory of Machine Perception (Ministry of Education), Department of Machine Intelligence, School of Electronics Engineering and Computer Science, Peking University, Yiheyuan Road 5, 100871 Beijing, People's Republic of China
Full list of author information is available at the end of the article

Abstract

An appropriate choice of the distance metric is a fundamental problem in pattern recognition, machine learning and cluster analysis. Some methods that based on the distance of samples, e.g, the k-means clustering algorithm and the k-nearest neighbor classifier, are crucially relied on the performance of the distance metric. In this paper, the property of translation invariance for the distance metric of images is especially emphasized. The consideration is twofold. Firstly, some of the commonly used distance metrics, such as the Euclidean and Minkowski distance, are independent of the training set and/or the domain-specific knowledge. Secondly, the translation invariance is a necessary property for any intuitively reasonable image metric. The image Euclidean distance (IMED) and generalized Euclidean distance (GED) are image metrics that take the spatial relationship between pixels into consideration. Sun et al.(IEEE Conference on Computer Vision and Pattern Recognition, pp 1398–1405, 2009) showed that IMED is equivalent to a translation-invariant transform and proposed a metric learning algorithm based on the equivalency. In this paper, we provide a complete treatment on this topic and extend the equivalency to the discrete frequency domain. Based on the connection, we show that GED and IMED can be implemented as low-pass filters, which reduce the space and time complexities significantly. The transform domain metric learning proposed in (Sun et al. 2009) is also resembled as a translation-invariant counterpart of LDA. Experimental results demonstrate improvements in algorithm efficiency and performance boosts on the small sample size problems.

Keywords: Image Euclidean distance, Translation invariant, Linear discriminant analysis, Transform domain metric learning

Background

The distance measure of images plays a central role in computer vision and pattern recognition, which can be either learned from a training set, or specified according to a priori domain-specific knowledge. The problem of metric learning, has gained considerable interest in recent years (Hastie and Tibshirani 1996; Xing et al. 2003; Hertz and Pavel 2002; Bar-Hillel et al. 2003; Goldberger et al. 2005; Shalev-Shwartz et al. 2004; Chopra et al. 2005; Globerson et al. 2006; Weinberger et al. 2005; Lebanon 2006; Davis et al. 2007; Li et al. 2007). On the other hand, the fact that the standard Euclidean distance assumes that pixels are spatially independent yields counter-intuitive results, e.g, a perceptually large distortion can produce smaller distance (Jean 1990; Wang et al. 2005). By incorporating the spatial correlation of pixels, two classes of image metrics, namely

Sun *et al. Appl Inform* (2015) 2:11

Page 2 of 12

IMED (Wang et al. 2005) and GED (Jean 1990), were designed to deal with the spatial dependencies for image distances, which were demonstrated consistent performance improvements in many real world problems (Jean 1990; Wang et al. 2005; Chen et al. 2006; Wang et al. 2006; Zhu et al. 2007).

A key advantage of GED and IMED is that they can be embedded in any classification technique. The calculation of IMED is equivalent to performing a linear transform called the standardizing transform (ST) and then followed by the traditional Euclidean distance. Hence, feeding the ST-transformed images to a recognition algorithm automatically embeds IMED (Wang et al. 2005). The analogous transform for GED is referred as to the generalized Euclidean transform (GET) (Jean 1990).

IMED and GED are invariant to image translation, namely, if the same image translation is applied to two images, their IMED remains invariant. However, the associated transforms (ST and GET) are not translation invariant (TI). This left a problem whether IMED can be implemented by a TI transform. In (Sun et al. 2009), the authors gave a positive answer to the problem and provided a proof for simple cases, yet a few technical problems are left unresolved.

We should emphasize the importance of the translation invariances. Intuitively, as the relative distance between images should only depend on the relative position of them, translation invariance (TI) should be a fundamental requirement for any reasonable image metric. Yet few metric learning or linear subspace methods are aware of the TI property when dealing with images.

In this paper, we extend the theory in (Sun et al. 2009) to the discrete frequency domain to cover the practical cases. Based on the metric-transform connection, we show that both GED and IMED are essentially low-pass filters. The resulting filters lead to the fast implementations of GED and IMED, coinciding the algorithm proposed in (Sun et al. 2008), which reduces the space and time complexities significantly. The transform domain metric learning (TDML) proposed in (Sun et al. 2009) is also resembled as a translation-invariant counterpart of LDA. Experimental results demonstrate significant improvements of algorithm efficiency and performance boosts on the small sample size problems.

IMED and GED

Given an image X of size $n_1 \times n_2$, the *vectorization* of X is the vector $\mathbf{x} = \text{vec}(X)$, such that the $(n_2 i_1 + i_2)$th component of $\mathbf{x}$ is the intensity at the (i_1, i_2) pixel. This is a common technique to manipulate image data.

The assumption made in the standard Euclidean distance that the image pixels are spatially independent sometimes leads to counter-intuitive results (Jean 1990; Wang et al. 2005). To solve the problem, Wang et al. (2005) proposed the image Euclidean distance (IMED) defined as

$$d_G^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T G (\mathbf{x} - \mathbf{y}).$$

The entries g_{ij} of the metric matrix G are defined by the Gaussian function (Wang et al. 2005), i.e.,

Sun *et al. Appl Inform* (2015) 2:11

Page 3 of 12

$$g_{ij} = f\left(\|P_i - P_j\|\right)$$
$$= \frac{1}{2\pi\sigma^2}e^{-\frac{|P_i-P_j|^2}{2\sigma^2}}$$
$$= \frac{1}{2\pi\sigma^2}e^{-\frac{(i_1-j_1)^2+(i_2-j_2)^2}{2\sigma^2}}, \tag{1}$$

where $P_i = (i_1, i_2), P_j = (j_1, j_2)$. The $n_1 n_2 \times n_1 n_2$ metric matrix G solely defines the IMED, where the element g_{ij} represents how the component x_i affects the component x_j.

As suggested in (Wang et al. 2005), the calculation of IMED can be simplified by decomposing G to $A^T A$. The standardizing transform (ST) is the special case when $A^T = A$, written as $A = G^{\frac{1}{2}}$. By incorporating the standardizing transform matrix $G^{\frac{1}{2}}$, IMED can be easily embedded into almost any recognition algorithm. That is, feeding the ST-transformed image $G^{\frac{1}{2}}\mathbf{x}$ to a recognition algorithm automatically embeds IMED. Besides, Wang et al.showed that ST seems to have a smoothing effect (Wang et al. 2005) by illustrating a few eigen-vectors associated with the largest eigen-value of $G^{\frac{1}{2}}$, and then argued that since IMED is equivalent to a transform domain smoothing, it can tolerate small deformation and noises and hence improve recognition performances.

Another image metric, called the generalized Euclidean distance (GED) (Jean 1990), is essentially the same as IMED, except the distance measure coefficients between P_i and P_j. Specifically, the generating function for GED is the probability density function of the Laplace distribution

$$g_{ij} = e^{-\alpha\cdot(|i_1-j_1|+|i_2-j_2|)},$$

where α is a scale parameter.

As pointed out in (Wang et al. 2005), translation invariance (TI) is a necessary property for any intuitively reasonable image metric. Formally, for image X, Y, a distance measure $d(\cdot, \cdot)$ is translation invariant if and only if

$$d(X, Y) = d(X_\tau, Y_\tau),$$

where X_τ, Y_τ is an image translation of X, Y, respectively.

Both IMED and GED depend only on the relative position between pixels P_i and P_j, i.e., there exists a discrete function $g[\cdot, \cdot]$, such that

$$g_{ij} = g[i_1 - j_1, i_2 - j_2],$$

where

$$i = n_2 i_1 + i_2, \quad j = n_2 i_1 + i_2.$$

This makes g_{ij} invariant to image translation. However, the associated transform (ST and GET) are not translation invariant transforms. This left a problem whether IMED and GED can be decomposed to translation invariant transforms. That is, for any IMED or GED metric matrix G, does there exist a translation-invariant transform H such that $G = H^T H$?

Sun *et al. Appl Inform* (2015) 2:11

Page 4 of 12

The translation invariant transform of a translation invariant metric

In (Sun et al. 2009), the authors give a positive answer to the problem whether a translation invariant metric can be implemented by a translation invariant transform.

Theorem 1   *Given a translation invariant metric matrix G of n × n and thus a finitely sequence $g[i - j] = G(i, j)$ supported on $[-n, n]$, supposing that $\hat{g}(\omega) \geqslant 0$ (the discrete time Fourier transform of g[i]), there exists a translation invariant transform matrix H such that*

$$G = H^*H.$$

Specifically, define the filter h[i]

$$h[i] = \mathcal{F}^{-1}\left(\sqrt{\hat{g}(\omega)} e^{\sqrt{-1}\theta(\omega)} \right)$$

which satisfies that

$$G(i, j) = \langle h * \delta_j, h * \delta_i \rangle.$$

If h[i] is supported on $[-m, m]$, it can be equivalently written as

$$G = H^*H,$$

where H is the $(n + 2m) \times n$ LTI matrix of h[i] defined by

$$H(i, j) = \begin{cases} h[i - j - m], & \text{if } |i - j - m| \leqslant m, \\ 0, & \text{else.} \end{cases}$$

Each diagonal of H is constant, thus H is a Toeplitz matrix (Gray 2006) or diagonal-constant matrix.

A solid requirement of Theorem 1 is $\hat{g}(\omega) \geqslant 0$. The condition is satisfied when $G \geqslant 0$ is an infinite-sized matrix, as a consequence of the positive operator theorem (Rudin 1991) or the generalized Bochner's theorem on groups (Rudin 1990). In practice, G is a positive-definite matrix of finite size $n \times n$. Gray (2006) proved that as n approximates infinity, $\hat{g}(\omega)$ converges to a non-negative value.

Unlike the case of ST for IMED (Wang et al. 2005) and GET for GED (Jean 1990), the constructed translation-invariant transform matrix H is not a square matrix. Specifically, H is of size $(n + 2m) \times n$, where $[-m, m]$ is the support of the sequence $g[i]$.

Methods

Computational aspects

Unfortunately, Theorem 1 is presented in the continuous frequency domain only (Sun et al. 2009), which is not easy to be applied directly in practical problems because $\hat{g}(\omega)$ is a continuous function that has to be discretized. A naive extension of Theorem 1 can be constructed by using the circular convolution (Oppenheim et al. 1999) instead of the regular convolution.

Sun *et al. Appl Inform* (2015) 2:11

Page 5 of 12

Proposition 2 *If H_n is a circulant matrix, then the $n \times n$ metric matrix (which is also circulant) defined by $G_n = H_n^T H_n$ can be determined by*

$$G_n(i, j) = g[i - j],$$

where $g[i]$ is the auto-correlation function of $h[i]$, i.e.,

$$g[i] = h[i] \circledast_n h^*[i],$$

with $h^*[i] = \overline{h[-i]}$, where $\circledast_n$ denotes the n-point circular convolution, or equivalently in frequency domain,

$$\hat{g}[j] = |\hat{h}[j]|^2.$$

The above extension has problems. The first problem is that, for the same filter $h[i]$, the induced metric filters $g = h * h$ and $\tilde{g} = h \circledast_n h$ are different, i.e.,

$$H_n^* H_n \neq H_{m,n}^* H_{m,n},$$

because linear convolution and circular convolution don't equal generally.

The second problem is even worse: to derive a translation-invariant transform in discrete frequency domain, the matrix representation of the metric **G** must be a circulant matrix, which is not true for common cases, including both IMED and GED.

We adopt the following approach to overcome these problems: padding the finitely supported sequences to periodic sequences. Given $h[i]$ supported on $[-m, m)$ and $x[i]$ supported on $[0, n)$, define $\tilde{h}[i]$ and $\tilde{x}[i]$ of period-$(n + 2m)$ by

$$\tilde{h}[i] = \begin{cases} h[i], & i \in [-m, m) \\ 0, & i \in [m, m + n) \end{cases}$$

and

$$\tilde{x}[i] = \begin{cases} x[i], & i \in [0, n) \\ 0, & i \in [-m, 0) \cup [n, m + n). \end{cases}$$

By the circular convolution theorem (Oppenheim et al. 1999), the two types of convolution coincide:

$$h * x[i] = \begin{cases} \tilde{h} \circledast_{n+2m} \tilde{x}[i], & i \in [-m, m + n); \\ 0, & \text{else.} \end{cases}$$

In other words, the linear convolution of h and x on its support is a period of the circular convolution of their periodic expansion $\tilde{h}$ and $\tilde{x}$.

Now consider the two versions of metric filter: $g[i] = h * h^*[i]$ and $\tilde{g}[i] = \tilde{h} \circledast_{n+2m} \tilde{h}^*[i]$. Because

$$\forall i \in [0, \ n + 2m), \tilde{h}[i - m] = h[i - m],$$

hence $g[i] = \tilde{g}[i]$ if and only if

$$i \in [-2m, n) \bigcap (-n, +\infty).$$

Sun *et al. Appl Inform* (2015) 2:11

Page 6 of 12

On the other hand, by definition the metric filter is conjugate symmetric, i.e,,

$$g[i] = \overline{g[-i]}, \ \tilde{g}[i] = \overline{\tilde{g}[-i]},$$

so it can be asserted that $g[i] = \tilde{g}[i]$ when $i \in (-n, n)$.

The above statements assert that given a finitely supported translation-invariant transform $h[x]$, the induced metric $\tilde{g}[i]$ constructed by the padded period filter $\tilde{h}[i]$ is also translation invariant.

Hence, the analogous version of Theorem 1 can be given as follows.

Theorem 3 *Given the $[-m, m)$ supported metric filter $g[i]$, there exists a circular filter $\tilde{h}[i]$, such that $g[i]$ is equal to $\tilde{h} \circledast_{n+2m} \tilde{h}[i]$ on its support.*

Proof Define the period-$(n + 2m)$ sequence $\tilde{g}$ by

$$\tilde{g}[i] = \begin{cases} g[i], & i \in [-m, m) \\ 0, & i \in [m, m + n). \end{cases}$$

Let $\tilde{h}[i] = \mathcal{F}^{-1}\left(\sqrt{\widehat{\tilde{g}[i]}}\right)$ and the proof is complete. □

It is beneficial to derive the matrix representation of Theorem 3. Given the $n \times n$ metric matrix G_n, by Theorem 1, it determines a filter $h[i]$ supported on $[-m, m)$, and hence the $(n + 2m) \times n$ translate-invariant matrix $H_{m,n}$; by theorem 3, it determines a filter $\tilde{h}[i]$ of period $n + 2m$, and hence the $(n + 2m) \times (n + 2m)$ circular matrix $\tilde{H}_{m,n}$. Writing

$$G_n = H_{m,n}^* H_{m,n}$$
$$\tilde{G}_{n+2m} = \tilde{H}_{n+2m}^* \tilde{H}_{n+2m},$$

and it can be checked that G_n is the left-upper $n \times n$ block of $\tilde{G}_{n+2m}$.

The results in discrete frequency domain can be easily extended to multi-dimensional signal space the same as in continuous frequency domain (Sun et al. 2009). A convenient property of the extension is that the multi-dimensional data (e.g, 2d images) can be processed without vectorization.

The translation-invariant transforms of IMED and GED

To demonstrate that the proposed method can be applied to multi-dimensional cases directly, we write the metric matrices of IMED and GED in tenser form.

IMED The metric tensor g for IMED is defined in (Wang et al. 2005) by a Gaussian, i.e.,

$$g_{j_1 j_2}^{i_1 i_2} = \frac{1}{2\pi} e^{-\frac{d^2}{2}},$$

where

$$d = \sqrt{(i_1 - j_1)^2 + (i_2 - j_2)^2}.$$

The metric filter for IMED is separable, i.e.,

Sun *et al. Appl Inform* (2015) 2:11

Page 7 of 12

$$g[i_1, i_2] = \frac{1}{2\pi} e^{-\frac{i_1^2 + i_2^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{i_1^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{i_2^2}{2}} = g_0[i_1] g_0[i_2].$$

We choose the support length $m_1 = m_2 = 4$ ($g[4, 4] \approx 1.7911 \times 10^{-8}$), i.e., $g[i_1, i_2]$ is supported on $[-4, 4] \times [-4, 4]$. For 52×52 signals ($n_1 = n_2 = 52$), we build the period $n_1 + 2m_1 = 60$ sequence

$$\widetilde{g_0}[i] = \begin{cases} g_0[i] = \frac{1}{\sqrt{2\pi}} e^{-\frac{i^2}{2}}, & i \in [-4, 4] \\ 0, & i \in (4, 56). \end{cases}$$

It is easy to validate that $\widehat{\widetilde{g_0}}[j] \geqslant 0, \forall j$. Thus the separated period filter $\widetilde{h_0}[i]$ can be constructed by

$$\widetilde{h_0}[i] = \mathcal{F}^{-1}\left(\sqrt{\widehat{\widetilde{g}}[j]} \right),$$

and the overall filter is $\widetilde{h}[i_1, i_2] = \widetilde{h_0}[i_1] \widetilde{h_0}[i_2]$.

GED The metric tensor g for GED is defined in (Jean 1990) by a Laplacian, i.e.,

$$\mathrm{g}_{j_1 j_2}^{i_1 i_2} = r^d = e^{d \log r},$$

where $d = |i_1 - j_1| + |i_2 - j_2|$ is the l_1 distance of the two pixels and $r = 0.6$ is a decay constant. The metric filter for GED is separable, i.e.,

$$g[i_1, i_2] = r^{|i_1| + |i_2|} = r^{|i_1|} \cdot r^{|i_2|} = g_0[i_1] g_0[i_2].$$

We choose the support length $m_1 = m_2 = 15$ ($g[15, 15] \approx 2.2107 \times 10^{-7}$), i.e., $g[i_1, i_2]$ is supported on $[-15, 15] \times [-15, 15]$. For 30×30 signals ($n_1 = n_2 = 30$), we build the period $n_1 + 2m_1 = 60$ sequence

$$\widetilde{g_0}[i] = \begin{cases} g_0[i] = r^{|i|}, & i \in [-15, 15] \\ 0, & i \in (15, 45). \end{cases}$$

We can validate that $\widehat{\widetilde{g_0}}[j] \geqslant 0, \forall j$. Thus the separated period filter $\widetilde{h_0}[i]$ can be constructed by

$$\widetilde{h_0}[i] = \mathcal{F}^{-1}\left(\sqrt{\widehat{\widetilde{g}}[j]} \right),$$
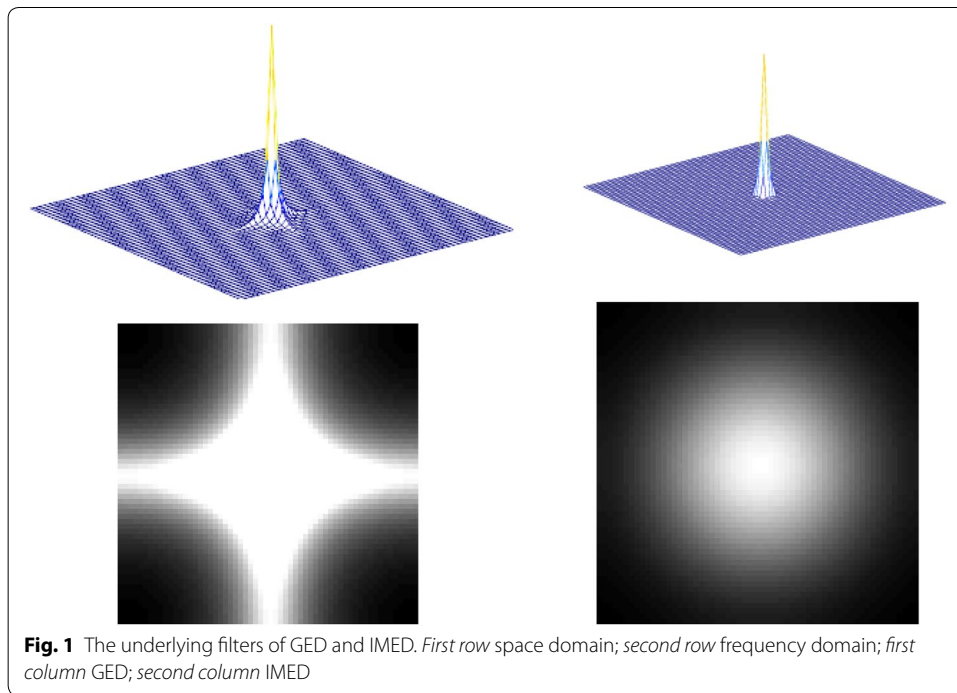
and the overall filter is $\widetilde{h}[i_1, i_2] = \widetilde{h_0}[i_1] \widetilde{h_0}[i_2]$.

The translation-invariant transforms of IMED and GED in space and frequency domain are drawn in Fig. 1. It clearly shows that applying the GED or IMED is equivalent to a low-pass filtering process, which is robust to small perturbation of images.

The fast implementation of IMED and GED

The advantages of the filtering decomposition over the GET or ST are not only the physical explanation but also the time and space complexity. Generally, the computational complexity associated with the filtering decomposition can be of $O(n \log n)$ due to the efficiency of FFT (Oppenheim et al. 1999).

In the case of IMED and GED, since the corresponding filters decay rapidly (Fig. 1), e.g., $g[4] = \frac{1}{\sqrt{2\pi}} e^{-\frac{4^2}{2}} \approx 1.34 \times 10^{-4}$ (IMED), the vector $\mathbf{g} = (g[0], \ldots, g[m], 0, \ldots, 0, g[-m], \ldots, g[-1])^T$ can be set of length n. Therefore

Sun *et al. Appl Inform* (2015) 2:11

Page 8 of 12

Fig. 1 The underlying filters of GED and IMED. *First row* space domain; *second row* frequency domain; *first column* GED; *second column* IMED

$G \approx \tilde{G}$ and the transform can be applied on the original X than the zero-padded image $\tilde{X}$. Finally, the period filter $\tilde{g}$ can be built using only several significant values. The templates of IMED ($\sigma = 1$) and GED ($\alpha = 2$) are

$$
\begin{pmatrix}
0 & 0.0012 & 0.0029 & 0.0012 & 0 \\
0.0012 & 0.0471 & 0.1198 & 0.0471 & 0.0012 \\
0.0029 & 0.1198 & 0.3046 & 0.1198 & 0.0029 \\
0.0012 & 0.0471 & 0.1198 & 0.0471 & 0.0012 \\
0 & 0.0012 & 0.0029 & 0.0012 & 0
\end{pmatrix},
$$

and

$$
\begin{pmatrix}
0.0003 & 0.0025 & 0.0183 & 0.0025 & 0.0003 \\
0.0025 & 0.0183 & 0.1353 & 0.0183 & 0.0025 \\
0.0183 & 0.1353 & 1.0000 & 0.1353 & 0.0183 \\
0.0025 & 0.0183 & 0.1353 & 0.0183 & 0.0025 \\
0.0003 & 0.0025 & 0.0183 & 0.0025 & 0.0003
\end{pmatrix},
$$

respectively.

Since the filter is of fixed size, the fast implementation can further reduces the space complexity from $O(n^2)$ to $O(1)$, and the time complexity from $O(n^2)$ to $O(n)$.

Transform domain metric learning

Generally, in order to learn a metric G, one can do optimization with respect to G. For images of size $n_1 \times n_2$, G has $n_1^2 \times n_2^2$ elements, making the optimization intractable. Another problem is G must satisfy the positive semi-definite constraint, i.e., $G \geqslant 0$, so it is not easy to find efficient algorithm to solve problem with such a constraint.

Theorem 1 can be equivalently written

$$
x^T G x = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{g}(\omega) |\hat{x}(\omega)|^2 d\omega. \tag{2}
$$

Sun *et al. Appl Inform* (2015) 2:11

Page 9 of 12

Equation (2) introduces great simplifications to the optimization problem of metric learning. With the translation-invariant assumption on G, things are much simpler. This is because the positive semi-definitive constraint $G \geqslant 0$ is reduced to a bound constraint $\hat{g}(\boldsymbol{\omega}) \geqslant 0$. Furthermore, the number of parameters is the sampling number on $\hat{g}$, which is usually chosen to be the same as the size of input data. An additional benefit of the translation-invariant approach is that it applies to any dimensionality without modifications, thus is unnecessary to stack the multi-dimensional data to vectors.

Suppose we have some data $\{x_i\}$, and are given the data label $\{y_i\}$. Let f_i be the Fourier transform of x_i, we compute the total "similar" and "dissimilar" power spectrum:

$$p_w(\boldsymbol{\omega}) = \sum_{i,j,y_i=y_j} |f_i(\boldsymbol{\omega}) - f_j(\boldsymbol{\omega})|^2, \qquad p_b(\boldsymbol{\omega}) = \sum_{i,j,y_i \neq y_j} |f_i(\boldsymbol{\omega}) - f_j(\boldsymbol{\omega})|^2.$$

The criterion here is that the filtered within-class distance is minimized, and the filtered between-class distance is maximized, simultaneously. This gives the objective functional

$$J_0(g) = \frac{\int_{T^d} \hat{g}(\boldsymbol{\omega}) p_w(\boldsymbol{\omega}) \mathrm{d}\boldsymbol{\omega}}{\int_{T^d} \hat{g}(\boldsymbol{\omega}) p_b(\boldsymbol{\omega}) \mathrm{d}\boldsymbol{\omega}}. \tag{3}$$

The objective (3) resembles the idea of LDA (Duda et al. 2000). In fact, TDML can be viewed as a translate-invariant solution to LDA.

Results

Experiments on the transform implementations of IMED

In this section, the standardizing transform (ST) and the translation invariant implementation of IMED are evaluated using the US postal service (USPS) and the FERET database. The USPS database consists of 16 by 16 pixel size normalized images of handwritten digits, divided into a training set of 7291 prototypes and a test set of 2007 pattern. The FERET database consists of 384 by 256 pixel size images of human faces, in which th 'fa' subset is chosen, including 1762 images.

The following algorithms are going to be compared, divided into 2 gourps:

1 The ST group

Algorithm 1 $U = G^{\frac{1}{2}} \mathrm{vec}(X)$, the original ST. It is memory expensive, and sometimes unfeasible, e.g, for the FERET database, the $G^{\frac{1}{2}}$ is of size 98304×98304, yielding a 36GiB usage of memory (4 bytes per element).

Algorithm 2 Since G is separable Wang et al. (2005), it can be shown $G_1^{\frac{1}{2}} X G_2^{\frac{1}{2}}$ is equivalent to Algorithm 1. This solves the memory problem. For the FERET database, only a 384×384 and a 256×256 matrices are needed.

2 The CST group (translation invariant transforms)

Algorithm 3 $(\mathbf{h}_1 \otimes \mathbf{h}_2^*) * X$, we need only a pre-computed 5×5 template.

Sun *et al. Appl Inform* (2015) 2:11

Page 10 of 12

Algorithm 4 Apply the template $\mathbf{h}_1$ to each column of *X*, then $\mathbf{h}_2$ to each row of *X*. This is the separated equivalent to Algorithm 3, in compared with Algorithm 2. Because $\mathbf{h}_1 = \mathbf{h}_2$, only one copy is in memory.

These algorithms were evaluated over the $7291 + 2007$ USPS images, and the 1762 FERET-fa images using MATLAB on a Dell PowerEdge 1950. The results (Table 1) demonstrate that the CST does improve the time efficiency significantly, especially in the case of large size images.

Also, we computed the Euclidean distance of CST-ed images, which has an error rate of $\sim 1\%$ comparing to the IMED of the original images, due to the approximate property of the convolution template.

Experiments on the transform domain metric learning

In this section, we conduct several sets of experiments. The experiments are performed on 3 face data sets (UMIST, Yale and ORL database). The images in UMIST, Yale and ORL data sets are resized to 28×23, 40×30 and 28×23, respectively.[1] We randomly select two images from each class as the training set, and use the remaining images for test. We repeat the process 20 times independently and the average results are calculated.

We first compare TDML with several other metrics, including the standard Euclidean distance (ED), IMED, GED, and a metric learning method XNZ (Xiang et al. 2008). The performances are evaluated in terms of recognition rate using a nearest neighbor classifier. The recognition results are shown in Table 2. TDML significantly outperforms all metrics.

Another set of experiments was to test whether embedding the learned TI metric in an image recognition technique, e.g., SVM (Vapnik 1998), can improve that algorithm's accuracy. Embedding a TI metric in an algorithm is simple: first, transform all images

Table 1 Time complexities

Algorithm	USPS (s)	FERET-fa (s)
ST(1)	0.3283	n/a
ST(2)	0.0970	130.13
CST(3)	0.7330	10.26
CST(4)	0.0584	15.34

Table 2 Comparison of image metrics on various databases (%)

	ED	IMED	GED	XNZ	TDML
UMIST	60.88	60.90	62.05	60.96	*73.92*
Yale	71.41	71.41	71.11	67.73	*75.26*
ORL	81.95	81.63	80.88	81.24	*84.06*

[1] The resization is necessary for traditional subspace and metric learning methods since they are vulnerable to the computational issue and small sample size problem from the curse of dimensionality. Our method doesn't suffer from it.

Sun *et al. Appl Inform* (2015) 2:11

Page 11 of 12

Table 3 SVM classification performances of the embedded metrics (%)

	ED	IMED	GED	TDML
UMIST	60.33	62.02	62.45	*69.53*
Yale	68.90	69.12	69.23	*72.30*
ORL	79.25	79.07	79.00	*80.38*

by the corresponding TI transform, and then run the algorithm with the transformed images as input data.

Table 3 gives the results of the metric when embedded to SVM. It can be found that TDML improves the performance of SVM better than IMED and GED.

Conclusion

In this paper, we extend the equivalency in (Sun et al. 2009) to the discrete frequency domain. We show that GED and IMED are low-pass filters, resulting in fast implementations which reduce the space and time complexities significantly. The transform domain metric learning (TDML) proposed in (Sun et al. 2009) is also resembled as a translation-invariant counterpart of LDA. Experimental results demonstrate significant improvement of algorithm efficiency and performance boosts on small sample size problems.

One possible future direction is the search for more effective metric learning algorithm. TDML is a simple and intuitive attempt and we expect novel methods that combine the concepts of margins, kernels, locality and non-linearity.

Author details
[1] Key Laboratory of Machine Perception (Ministry of Education), Department of Machine Intelligence, School of Electronics Engineering and Computer Science, Peking University, Yiheyuan Road 5, 100871 Beijing, People's Republic of China. [2] Graphics and Interactive Laboratory, School of Electronics Engineering and Computer Science, Peking University, Yiheyuan Road 5, 100871 Beijing, People's Republic of China.

References
Bar-Hillel A, Hertz T, Shental N, Weinshall D (2003) Learning distance functions using equivalence relations. Proc Int Conf Mach Learn 11–18
Chopra S, Hadsell R, LeCun Y (2005) Learning a similarity metric discriminatively, with application to face verification. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, vol. 1, pp 539–5461. doi:10.1109/CVPR.2005.202
Chen J, Wang R, Shan S, Chen X, Gao W (2006) Isomap based on the image euclidean distance. In: 18th International Conference on Pattern Recognition, 2006. ICPR 2006. vol. 2, pp 1110–1113. doi:10.1109/ICPR.2006.729
Davis JV, Kulis B, Jain P, Sra S, Dhillon IS (2007) Information-theoretic metric learning. In: Proceedings of the 24th International Conference on Machine Learning. ICML '07, ACM, New York, NY, USA, pp. 209–216. doi:10.1145/1273496.1273523. http://doi.acm.org/10.1145/1273496.1273523. Accessed 15 May 2013
Duda RO, Hart PE, Stork DG (2000) Pattern Classification, 2nd edn. Wiley-Interscience (2000)
Goldberger J, Roweis S, Hinton G, Salakhutdinov R (2005) Neighbourhood components analysis. In: Saul LK, Weiss Y, Bottou L (eds) Advances in neural information processing systems, vol 17. MIT Press, Cambridge, MA, pp 513–520

Sun *et al. Appl Inform* (2015) 2:11

Page 12 of 12

Globerson A, Roweis S (2006) Metric learning by collapsing classes. In: Weiss Y, Schölkopf B, Platt J (eds) Advances in neural information processing systems, vol 18. MIT Press, Cambridge, pp 451–458

Gray RM (2006) Toeplitz and circulant matrices: a review. Found Trends Commun Inform Theory 2(3):155–239

Hastie T, Tibshirani R (1996) Discriminant adaptive nearest neighbor classification. IEEE Trans Pat Anal Mach Intel 18(6):607–616. doi:10.1109/34.506411

Jean JSN (1990) A new distance measure for binary images. In: International Conference on Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., pp. 2061–2064. doi:10.1109/ICASSP.1990.115932

Lebanon G (2006) Metric learning for text documents. IEEE Trans Pat Anal Mach Intel 28(4):497–508. doi:10.1109/TPAMI.2006.77

Li F, Yang J, Wang J (2007) A transductive framework of distance metric learning by spectral dimensionality reduction. In: Proceedings of the 24th Annual International Conference on Machine Learning (ICML 2007), pp 513–520

Oppenheim AV, Schafer RW, Buck JR (1999) Discrete-Time Signal Processing, 2nd edn., Prentice Hall Signal Processing Series, Prentice Hall, Englewood Cliffs

Rudin W (1991) Functional Analysis, 2nd edn. McGraw-Hill Book Company, New York

Rudin W (1990) Fourier Analysis on Groups. Wiley, New York

Shalev-Shwartz S, Singer Y, Ng AY (2004) Online and batch learning of pseudo-metrics. In: Proceedings of the Twenty-first International Conference on Machine Learning. ICML '04, ACM, New York, p 94. doi:10.1145/1015330.1015376.http://doi.acm.org/10.1145/1015330.1015376. Accessed 11 03 2013

Shental N, Hertz T, Weinshall D, Pavel M (2002) Adjustment learning and relevant component analysis. In: ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part IV, Springer, London, pp. 776–792

Sun B, Feng J (2008) A fast algorithm for image euclidean distance. In: Chinese Conference on Pattern Recognition, 2008. CCPR '08, pp 1–5. doi:10.1109/CCPR.2008.32

Sun B, Feng J, Wang L (2009) Learning IMED via shift-invariant transformation. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009, pp 1398–1405. doi:10.1109/CVPR.2009.5206720

Vapnik VN (1998) Statistical Learning Theory. Wiley-Interscience

Weinberger KQ, Blitzer J, Saul LK (2005) Distance metric learning for large margin nearest neighbor classification. In: Advances in Neural Information Processing Systems, vol. 18, pp 1473–1480. http://books.nips.cc/papers/files/nips18/NIPS2005-0265.pdf

Wang L, Zhang Y, Feng J (2005) On the euclidean distance of images. IEEE Trans Pat Anal Mach Intel 27(8):1334–1339. doi:10.1109/TPAMI.2005.165

Wang R, Chen J, Shan S, Chen X, Gao W (2006) Enhancing training set for face detection. In: 18th International Conference on Pattern Recognition, 2006. ICPR 2006. vol. 3, pp 477–480. IEEE Computer Society, Washington, DC. doi:10.1109/ICPR.2006.493

Xing EP, Ng AY, Jordan MI, Russell S (2003) Distance metric learning, with application to clustering with side-information. In: Advances in Neural Information Processing Systems 15, vol. 15, pp 505–512. http://citeseerx.ist.psu.edu/viewdoc/summary?. doi:10.1.1.58.3667

Xiang S, Nie F, Zhang C (2008) Learning a mahalanobis distance metric for data clustering and classification. Pat Recogn 41(12):3600–3612. doi:10.1016/j.patcog.2008.05.018

Zhu S, Song Z, Feng J (2007) Face recognition using local binary patterns with image euclidean distance. In: SPIE, vol. 6790. doi:10.1117/12.750642