Optics & Laser Technology 56 (2014) 290–298

Contents lists available at ScienceDirect

Optics & Laser Technology

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

Kernel maximum likelihood scaled locally linear embedding for night vision images

Jing Han, Jiang Yue, Yi Zhang, Lian-fa Bai *

Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, China

ARTICLE INFO

Article history:
Received 7 July 2013
Received in revised form
5 August 2013
Accepted 30 August 2013
Available online 23 September 2013

Keywords:
Manifold learning
Kernel
Night vision image

ABSTRACT

This paper proposes a robust method to analyze night vision data. A new kernel manifold algorithm is designed to match an ideal distribution with a complex one in natural data. First, an outlier-probability based on similarity metric is derived by solving the maximum likelihood in kernel space, which is corresponding with classification property for considering the statistical information on manifold. Then a robust nonlinear mapping is completed by scaling the embedding process of kernel LLE with the outlier-probability. In the simulations of artificial manifolds, real low-light-level (LLL) and infrared image sets, the proposed method show remarkable performances in dimension reduction and classification.

1. Introduction

With the development of night vision technology (LLL and infrared imaging), it is essential to represent high-dimensional night vision data in low-dimensional space and preserve intrinsic properties to facilitate subsequent recognition and visualization. Many manifold dimension reduction algorithms with appealing performances, including LLE, ISOMAP, and LPP [1–6], can be introduced to discover intrinsic structures in night vision data. However, heavy noises in LLL image and low contrast in infrared image may produce outliers, which will lead to inaccurate embedding.

Aiming at reducing the interference of outlier, two types of solutions are mainly developed: preprocessing procedures of detecting outliers and filtering noisy data [4,7]; improving the local distance metric to fit the classification property [8,9]. Nevertheless, the intrinsic manifold structures of night vision image usually have a high degree of complexity and randomness, which causes the obtained classification results to deviate from actual situation when the classificatory distribution is discrete or selection of samples are not representative [10,11]. So automatic unsupervised manifold method should be studied in night vision data, it will have a better similarity estimation, which is corresponding with classification property. This similarity estimation discovered from statistical regularities of large datasets will provide an effective measure for selection of neighbors and inhibition of outliers.

This paper proposes a robust manifold dimension reduction and classification algorithm against deformed distributed data, kernel space maximum likelihood (KML) scaled LLE (KLLE) method (KML-KLLE), to solve the problems mentioned above. A KML similarity metric is presented to detect outliers and select neighborhood to produce an accurate mapping of high-dimensional night vision data. Compared with works in [4,7,8], the KML similarity emphasizes local statistical distribution of the manifold to evaluate the outlier-probability instead of computing the distance metric among data. Besides, an outlier-probability scaled KLLE method is proposed to suppress outliers in LLL and infrared images.

Section 2 introduces the KML similarity metric. Section 3 details the outlier-probability scaled KLLE method, which is the whole KML-KLLE algorithm. Section 4 presents several experiments on artificial manifolds and real LLL and infrared image sets. Section 5 makes some discussion about the proposed algorithm. Finally, Section 6 serves as the conclusion.

2. KML similarity metric

Maximum likelihood (ML) method based on probability discriminant function and Bayesian criterion analyzes the statistical information of noise image comprehensively. It is especially suitable for estimation of the abnormal distribution and detection of outliers in images with fine structures. In addition, the ML method has simple prior knowledge integration and algorithm structure to be implemented [12–14]. Inspired by the concepts and advantages, we introduce ML to evaluate the outlier-probability of each sample point in feature space of LLL and infrared images.

* Corresponding author. Tel.: +86 258 431 5330.
E-mail addresses: mrblf@163.com, njusthanjing@163.com (L.-f. Bai).

0030-3992/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.optlastec.2013.08.028

The dataset in high-dimensional input space is $\{\mathbf{x}_i\}_{i=1,\ldots,M}$. Each sample point $\mathbf{x}_i=(x_{i1},x_{i2},\ldots,x_{iD})^T$ has ML value with their N-neighbors $\mathbf{X}=\{\mathbf{x}_j\}_{j=1,\ldots,N}\in w$:

$$ML(\mathbf{x}_i)=P(\mathbf{x}_i/w)/P(w)=\frac{P(w)}{(2\pi)^{q/2}|\mathbf{\Sigma}_w|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu}_w)^T\mathbf{\Sigma}_w^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_w)\right) \tag{1}$$

here $\boldsymbol{\mu}_w$ and $\mathbf{\Sigma}_w$ is the mean value and covariance in neighborhood w. Assuming that classification of each neighborhood has the same prior probability $P(w)$, the ML similarity estimation function $s(\mathbf{x}_i)$ can be expressed as follow by taking the logarithm of $ML(\mathbf{x}_i)$.

$$s(\mathbf{x}_i)=\ln|\mathbf{\Sigma}_w|+(\mathbf{x}_i-\boldsymbol{\mu}_w)^T\mathbf{\Sigma}_w^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_w)$$
$$\mathbf{\Sigma}_w=\frac{1}{N}\sum_{j=1}^{N}(\mathbf{x}_j-\boldsymbol{\mu}_w)(\mathbf{x}_j-\boldsymbol{\mu}_w)^T,\quad \boldsymbol{\mu}_w=\frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_j \tag{2}$$

here larger $s(\mathbf{x}_i)$ indicates higher outlier-probability of $\mathbf{x}_i$.

ML estimation is under the premise of normal distributed features in images, whereas the characteristics of night vision image do not follow a normal distribution. To eliminate the negative effects of Gaussian distribution assumption, we use kernel ML (KML), which nonlinearly maps the input signal to kernel space through kernel function $\phi(\cdot)$ and detects outliers by ML in kernel space. On the one hand, the nonlinear information of night vision data can be mined. On the other hand, Gaussian distribution assumption of data in kernel space means the complicated input spatial distribution, which is much closer to real situation.

The only difference between the input space and the kernel space is that we nonlinearly map each input sample $\mathbf{x}_i$ in input space to kernel space through kernel function $\phi(\cdot)$, so the sample in kernel space can be expressed as $\phi(\mathbf{x}_i)$. Then the mean value and covariance of neighborhood w in kernel space are modified as $\boldsymbol{\mu}_{w\phi}=(1/N)\sum_{j=1}^{N}\phi(\mathbf{x}_j)$ and -

$\mathbf{\Sigma}_{w\phi}=(1/N)\sum_{j=1}^{N}(\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi})(\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi})^T$. Thus the KML similarity metric function is transformed as follow.

$$s(\phi(\mathbf{x}_i))=\ln|\mathbf{\Sigma}_{w\phi}|+(\phi(\mathbf{x}_i)-\boldsymbol{\mu}_{w\phi})^T\mathbf{\Sigma}_{w\phi}^{-1}(\phi(\mathbf{x}_i)-\boldsymbol{\mu}_{w\phi})$$
$$\mathbf{\Sigma}_{w\phi}=\frac{1}{N}\sum_{j=1}^{N}(\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi})(\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi})^T,\quad \boldsymbol{\mu}_{w\phi}=\frac{1}{N}\sum_{j=1}^{N}\phi(\mathbf{x}_j) \tag{3}$$

The eigenvalue of $\mathbf{\Sigma}_{w\phi}$ is $\mathbf{\Lambda}_{w\phi}=\{\Lambda_{w\phi}^l\}_{l=1,\ldots,N}$ and eigenvector is $\mathbf{V}_{w\phi}=\{\mathbf{V}_{w\phi}^l\}_{l=1,\ldots,N}$. Set $\phi_w(\mathbf{x}_j)=\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi}$ and $\phi_w(\mathbf{X})=\{\phi_w(\mathbf{x}_j)\}_{j=1,\ldots,N}$, there are $\mathbf{\Sigma}_{w\phi}=(1/N)\phi_w(\mathbf{X})\phi_w^T(\mathbf{X})$ and $\mathbf{V}_{w\phi}^l=\sum_{j=1}^{N}\alpha_j^l\phi_w(\mathbf{x}_j)=\phi_w(\mathbf{X})\boldsymbol{\alpha}^l$, $\boldsymbol{\alpha}^l=[\alpha_1^l,\alpha_2^l,\ldots,\alpha_N^l]^T$. Insert $\mathbf{\Sigma}_{w\phi}=(1/N)\phi_w(\mathbf{X})\phi_w^T(\mathbf{X})$ and $\mathbf{V}_{w\phi}^l=\phi_w(\mathbf{X})\boldsymbol{\alpha}^l$ into $\Lambda_{w\phi}^l\mathbf{V}_{w\phi}^l=\mathbf{\Sigma}_{w\phi}\mathbf{V}_{w\phi}^l$ can acquire Eq. (4).

$$N\Lambda_{w\phi}^l\phi_w(\mathbf{X})\boldsymbol{\alpha}^l=\phi_w(\mathbf{X})\phi_w^T(\mathbf{X})\phi_w(\mathbf{X})\boldsymbol{\alpha}^l \tag{4}$$

Multiply Eq. (4) by $\phi_w^T(\mathbf{X})$ on both sides of the left and set $\mathbf{K}_w=\langle\phi_w(\mathbf{X}),\phi_w(\mathbf{X})\rangle$ to obtain $N\Lambda_{w\phi}^l\boldsymbol{\alpha}^l=\mathbf{K}_w\boldsymbol{\alpha}^l$. $\mathbf{K}_w=\mathbf{K}-\mathbf{IK}-\mathbf{KI}+\mathbf{IKI}$ is the kernel matrix, here $\mathbf{K}=\{k(\mathbf{x}_{ki},\mathbf{x}_{kj})\}_{ki,kj=1..N}$, $k(\mathbf{x}_{ki},\mathbf{x}_{kj})=\langle\phi(\mathbf{x}_{ki}),\phi(\mathbf{x}_{kj})\rangle$ and $\mathbf{I}$ is the unit matrix. Set $\mathbf{\Lambda}_w$ as the eigenvalue of $\mathbf{K}_w$ and $\mathbf{V}_w$ as the eigenvector. From $N\Lambda_{w\phi}^l\boldsymbol{\alpha}^l=\mathbf{K}_w\boldsymbol{\alpha}^l$ we can get $\mathbf{\Lambda}_w=N\mathbf{\Lambda}_{w\phi}$, $\mathbf{V}_w=\boldsymbol{\alpha}=\{\boldsymbol{\alpha}^l\}_{l=1,\ldots,N}$ and $\mathbf{V}_{w\phi}=\phi_w(\mathbf{X})\mathbf{V}_w$. So the pseudo-inverse of $\mathbf{\Sigma}_{w\phi}$ is:

$$\mathbf{\Sigma}_{w\phi}^{-1}=\mathbf{V}_{w\phi}\mathbf{\Lambda}_{w\phi}^{-1}\mathbf{V}_{w\phi}^T=\phi_w(\mathbf{X})\mathbf{V}_w\left(\frac{1}{N}\mathbf{\Lambda}_w\right)^{-1}\mathbf{V}_w^T\phi_w^T(\mathbf{X})=N\phi_w(\mathbf{X})\mathbf{K}_w^{-1}\phi_w^T(\mathbf{X}) \tag{5}$$

Then the KML similarity metric function $s(\phi(\mathbf{x}_i))$ is finally converted as:

$$\begin{aligned}s(\phi(\mathbf{x}_i))&=\ln|\mathbf{\Sigma}_{w\phi}|+(\phi(\mathbf{x}_i)-\boldsymbol{\mu}_{w\phi})^T\mathbf{\Sigma}_{w\phi}^{-1}(\phi(\mathbf{x}_i)-\boldsymbol{\mu}_{w\phi})\\&=\ln|\mathbf{\Sigma}_{w\phi}|+N(\phi(\mathbf{x}_i)-\boldsymbol{\mu}_{w\phi})^T\phi_w(\mathbf{X})\mathbf{K}_w^{-1}\phi_w^T(\mathbf{X})(\phi(\mathbf{x}_i)-\boldsymbol{\mu}_{w\phi})\end{aligned} \tag{6}$$

here

$$\begin{aligned}\phi(\mathbf{x}_i)^T\phi_w(\mathbf{X})&=\phi(\mathbf{x}_i)^T[(\phi(\mathbf{x}_1),\phi(\mathbf{x}_2),\ldots,\phi(\mathbf{x}_N))-\boldsymbol{\mu}_{w\phi}]=\mathbf{K}_{\mathbf{x}_i}\\&=(k(\mathbf{x}_i,\mathbf{x}_1),k(\mathbf{x}_i,\mathbf{x}_2),\ldots,k(\mathbf{x}_i,\mathbf{x}_N))-\sum_{j=1}^{N}k(\mathbf{x}_i,\mathbf{x}_j)/N\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\mu}_{w\phi}^T\phi_w(\mathbf{X})&=\boldsymbol{\mu}_{w\phi}^T[(\phi(\mathbf{x}_1),\phi(\mathbf{x}_2),\ldots,\phi(\mathbf{x}_N))-\boldsymbol{\mu}_{w\phi}]\\&=\mathbf{K}_{\boldsymbol{\mu}}=\textstyle\sum_{j=1}^{N}(k(\mathbf{x}_j,\mathbf{x}_1),k(\mathbf{x}_j,\mathbf{x}_2),\ldots,k(\mathbf{x}_j,\mathbf{x}_N))/N\\&\quad-\textstyle\sum_{ki=1}^{N}\sum_{kj=1}^{N}k(\mathbf{x}_{ki},\mathbf{x}_{kj})/N^2\end{aligned}$$

Simultaneously, based on the definition of determinant, $\det(\mathbf{\Sigma}_{w\phi})$ is the directed area or directed volume of super parallel polyhedron constituted by row or column vectors in $\mathbf{\Sigma}_{w\phi}$. Therefore $|\mathbf{\Sigma}_{w\phi}|$ can be approximately calculated as the summation of distances between centralized sample vectors and eigenvector in $\mathbf{\Sigma}_{w\phi}$:

$$\begin{aligned}|\mathbf{\Sigma}_{w\phi}|&\approx\sum_{j=1}^{N}(\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi})^T\mathbf{V}_{w\phi}\mathbf{V}_{w\phi}^T(\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi})\\&=\sum_{j=1}^{N}(\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi})^T\phi_w(\mathbf{X})\mathbf{V}_w\mathbf{V}_w^T\phi_w^T(\mathbf{X})(\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi})\\&=\sum_{j=1}^{N}\mathbf{K}_p\mathbf{V}_w\mathbf{V}_w^T\mathbf{K}_p^T\end{aligned} \tag{7}$$

here

$$\begin{aligned}\mathbf{K}_p&=\phi(\mathbf{x}_j)^T\phi_w(\mathbf{X})-\boldsymbol{\mu}_{w\phi}^T\phi_w(\mathbf{X})\\&=(k(\mathbf{x}_j,\mathbf{x}_1),k(\mathbf{x}_j,\mathbf{x}_2),\ldots,k(\mathbf{x}_j,\mathbf{x}_N))-\sum_{ki=1}^{N}k(\mathbf{x}_j,\mathbf{x}_{ki})/N-\mathbf{K}_{\boldsymbol{\mu}}.\end{aligned}$$

As shown in Eq. (6), $(\phi(\mathbf{x}_i)-\boldsymbol{\mu}_{w\phi})^T\mathbf{\Sigma}_{w\phi}^{-1}(\phi(\mathbf{x}_i)-\boldsymbol{\mu}_{w\phi})$ is kernel mahalanobis distance (KMD) actually, which means the covariance distance between sample point and overall neighbors in kernel space. And in Eq. (7), $\sum(\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi})^T\mathbf{V}_{w\phi}\mathbf{V}_{w\phi}^T(\phi(\mathbf{x}_j)-\boldsymbol{\mu}_{w\phi})$ is kernel PCA (KPCA) essentially, which represents the projection distance summation for overall neighbors on principal component of neighborhood in kernel space. So KML similarity metric is the combination of KPCA and KMD. Only when each sample point has the most compact neighborhood with minimum covariance distance, can it select the best neighbors and have the highest similarity with neighborhood to gain minimum outlier-probability and highest reliability.

3. KML outlier-probability scaled KLLE

Locally linear embedding (LLE) method assumes that original sample set is uniformly and continuously distributed on manifold and that linear method can be used to analyze the intrinsic structure of data. However, it is difficult to satisfy in practice. It is known that performing LLE in an appropriate kernel space means a complex distribution of data in the input space [15,16]. We implement kernel LLE method, and further employ the KML outlier-probability in scaling KLLE to decrease the interference of outliers. It is the KML-KLLE algorithm, which essentially integrates the KML similarity and KLLE method.

First, $\mathbf{F}_c$ denotes the clean dataset, which is identified by outlier-probability s_i of each sample point through $s_i\le\tau$, $\tau>0$ is the threshold of outlier. Use KNN method to choose N-neighbors of $\mathbf{x}_i$ and require all neighbors to meet $\{\mathbf{x}_j\}_{j=1,\ldots,N}\in\mathbf{F}_c$.

Second, calculate the reconstruction weights $\boldsymbol{\eta}=\{\eta_{ij}\}_{i=1,\ldots,M,j=1,\ldots,N}$ of the neighbors to minimize the reconstruction error $J(\boldsymbol{\eta})$ in kernel space.

$$J(\boldsymbol{\eta})=\sum_{i=1}^{M}\|\phi(\mathbf{x}_i)-\sum_{j=1}^{N}\eta_{ij}\phi(\mathbf{x}_j)\|^2 \tag{8}$$

Set $J(\boldsymbol{\eta}_i)=\|\phi(\mathbf{x}_i)-\sum_{j=1}^{N}\eta_{ij}\phi(\mathbf{x}_j)^2\|$, $\mathbf{G}_i=[\phi(\mathbf{x}_i)-\phi(\mathbf{x}_{i1}),\ldots,\phi(\mathbf{x}_i)-\phi(\mathbf{x}_{iN})]$, $\boldsymbol{\eta}_i=[\eta_{i1},\eta_{i2},\ldots,\eta_{iN}]^T$ and the constraint $\mathbf{e}^T\boldsymbol{\eta}_i=1$, there is