Neurocomputing 136 (2014) 152–161

Contents lists available at ScienceDirect

Neurocomputing

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

Comparative study among three strategies of incorporating spatial structures to ordinal image regression

Qing Tian, Songcan Chen *, Xiaoyang Tan

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, PR China

ARTICLE INFO

Article history:
Received 26 July 2013
Received in revised form
19 November 2013
Accepted 3 January 2014
Communicated by Qingshan Liu
Available online 28 January 2014

Keywords:
Ordinal regression
Vector pattern
Matrix pattern
Spatial structure
Euclidean distance
Bilinear

ABSTRACT

Images usually have specific spatial structures, and related researches have shown that these structures can contribute to the establishment of more effective classification algorithms for images. So far though there have been many solutions of making use of such spatial structures separately proposed, little attention has been paid to their systematic summary, let their comparative study alone. On the other hand, we find that the existing image-oriented ordinal regression (OR) methods do not utilize such structure information, which motivates us to compensate a comparative study through embedding such spatial structure into ORs. Towards the end, in this paper, we (**1**) through a summary, find three typical strategies of using image prior spatial information, i.e., structure-embedded Euclidean distance strategy, structure-regularized modeling strategy for classifier learning, and direct manipulation strategy on images without vectorization for image; *more importantly*, (**2**) apply these strategies to establish corresponding ORs for classifying data with ordinal characteristic, conduct comprehensive comparisons and give analysis on them under three evaluation criteria. Experimental results on typical ordinal image datasets JAFFE, UMIST and FG-NET show that the latter two strategies can, on the whole, achieve distinct gain in OR performance and while the first one cannot necessarily as expected, which is due to whether the spatial information is directly embedded into the objective function involved or not.

1. Introduction

1.1. Background

Images have two-dimensional inherent spatial structures, in which explicit and implicit discriminative information beneficial to image classification is involved. For example, in human faces, the eyes, nose and mouth are distributed in different regions, and specific geometric relations exist between them. However, most current developed pattern recognition and machine learning algorithms are based on vector patterns, in which the process of matrix-to-vector conversion is performed, consequently, useful spatial structure information to classification is lost seriously, thus leaving the room of performance promotion.

Over past years, though strategies of taking advantage of spatial structure information have been separately developed for improving performance of image classification, a systematic summary and comparative study among them is still lacked. For this purpose, in this paper, we will ***first*** make a summary from those scattered related literature and group them into three categories; then for making a comparison among them, we choose one of currently popular topics in image classification, i.e., image-oriented OR, as the comparative platform. OR is a special machine learning paradigm and possesses the duality of classification and regression, thus often is applied in such scenarios in which the predicted labels are discrete but ordered [26,31], e.g., human facial age estimation and movie scoring and so on. Besides the duality, further reasons of choosing image-oriented OR as the comparative study paradigm are (***1***) *these specially-designed ORs for ordinal image classification have so far hardly exploited such spatial information, and* (***2***) *the multi-index-based synthetic evaluation originated from their duality of classification and regression can more be reflected from multi-facets for such information utilization than single-index evaluation for classification or regression.* And ***next***, we develop three image-oriented OR variants by the compensation of spatial information using the aforementioned three strategies and then make an extensive comparison from a joint view of regression and classification under three evaluation criteria of MAE, Acc and OCI.

1.2. Categorization of the strategies of utilizing spatial structure information

In this subsection we analyze the existing scattered schemes designed to utilize the spatial structure information and summarize them into three main families as follows:

1.2.1. Structure-embedded Euclidean distance strategy

It is known that Euclidean distance (ED) is one of the most often-used metric in pattern recognition. However, when it is used

* Corresponding author.
E-mail addresses: tianqing@nuaa.edu.cn (Q. Tian), s.chen@nuaa.edu.cn (S. Chen), x.tan@nuaa.edu.cn (X. Tan).

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2014.01.017

to similarity/distance measure between two images, the spatial structure information involved in them is not sufficiently reflected such that classification performance for the images is unfavorably affected. In order to compensate such loss, many attempts [1–8] have been done, among which Ref. [1] can be viewed as their representative. In Ref. [1], the authors developed an IMage Euclidean Distance (IMED) by means of embedding spatial structure of images to ED and applied it to handwritten digit and human face recognition with better performance than ED. Due to its insensitiveness to small distortion of images and generality able to be embedded into such classifiers as SVM, IMED can successively be extended. For example, Li et al. [4] extended the IMED to multi-view gender classification and achieved higher classification accuracy; Liu et al. [5] further proposed multi-linear locality-preserved maximum information embedding for face recognition with more stable performance. Moreover, Li and Lu [8] developed an adaptive IMED (AIMED) by further fusing gray level knowledge of image to IMED besides the prior spatial information to achieve more satisfactory identification performance for human face and handwritten digit. In summary, these methods originating from IMED are either modified to different applications or embedded into other learning tasks such as SVM for performance gain. Thus in the following comparative study, we just adopt IMED as basic embedding, but any of its effective variants can straightforwardly be utilized in a similar way.

1.2.2. Structure-regularized modeling strategy

In this family, the strategy of exploiting spatial structure usually adopts the regularization technique to penalize a related objective function such that the resulted solution (by optimizing the objective) is spatially smooth as much as possible [9–13]. The spatial smooth subspace learning (SSSL) proposed in Ref. [9] can be regarded as the representative, in which a Laplacian penalty is imposed to constrain the projection coefficients to be spatially smooth. Zuo et al. [12] went further by weighting the Laplacian penalty function with Gaussian weights to realize multi-scale image smoothing. Chen et al. in Ref. [13] developed a regularized metric learning framework by again imposing the Laplacian penalty and achieved competitive face recognition performance on several benchmark datasets. From these related researches it can be easily found that the structure-regularized modeling indeed can also compensate the spatial information loss induced by tensor- or matrix-to-vector conversion. Therefore, we also try to adopt such a spatially-regularized strategy for image-oriented OR. Considering that adapting those successive strategies from the spatial regularization [9] to our problem is trivial, thus without loss of generality, we here take the spatial smooth constraint in Ref. [9] as the basic regularization strategy to conduct the following comparative study.

1.2.3. Direct manipulation strategy on images

The strategies in former two families are all vector-pattern-oriented. Though the spatial structure information of images can get utilized and thus related learning performance is boosted, these strategies usually suffer from (1) high computational complexity; and (2) the so-called "*small sample problem*", i.e., the dimensionality of feature vector is higher than the training set size, leading to over-fitting. Hence, a natural way to mitigate or address these problems is operating directly image (or reshaped image) patterns. Along this line, many studies have been developed, for example Refs. [14–25], in which the works of Chen et al. [14–18] and Tao et al. [20–25] can be regarded as their representatives. More specifically, Chen et al. developed a series of classifiers, such as MatMHKS [14] and MatFE+MatCD [18], by bilinear projection on image (or reshaped image) patterns and achieved competitive performance in such classification tasks as human face and handwritten digit identification, against the vector-pattern-oriented counterparts; while Tao et al. developed their dimensionality reduction or classification modeling directly manipulated on (higher order) tensor patterns and applied them respectively for human gait recognition [20] and visual tracking [25]. It is the direct manipulation on matrix or tensor as operating unit such that the schemes like the bilinear projection on image (second-order tensor) can make more sufficient use of the inherent spatial structure information involved in the data than their vectorized counterparts. Out of the similar consideration, in the following comparative study, we take the bi-lateral manipulation on image as a direct learning scheme to make a comparison on ordinal learning performance with the other methods.

Finally, we tabulate a brief comparative summarization for the aforementioned three strategies in Table 1.

1.3. Review of OR

Following the categorization and summary for spatial structure information utilization strategies, our next step is in position to taking the OR as a research platform, on which we will make extensive empirical comparison on three image sets among we afore-summarized three groups of categories. Before that let us briefly give a review for OR, OR is actually a special learning strategy used to design classifiers for ordinal classes, e.g., human age estimation. Due to its duality of regression and classification and powerful ability, OR has so far been widely applied in domains such as the recommender system [26], web page ranking [27], image retrieval [28], medical image diagnosis [29–30] and age estimation [31–32]. In implementing them, various approaches have been put forward [33–44], including KDLOR [44], one of distinguished ORs. Though most of these ORs have achieved performance to different extents, however, when manipulated on images, almost all these methods neglect the compensation of spatial structure information for vectorized images, thus choosing the image-oriented OR as the research platform to give a comparison among the summarized three categories of using spatial structure is reasonable. *Though such a work of incorporating the spatial information to existing OR seems trivial, to the best of our knowledge, there has indeed no related research done yet.*

Now for the sake of clarity but without loss of generality, we will just take the linear version of KDLOR, a typical OR model proposed in Ref. [44], as the basic OR approach (herein denoted as LDLOR), and select IMED [1], SSSL [9] and bilinear modeling [14] as the comparative representatives of the three families of spatial structure information utilization to re-model LDLOR, thus yielding three modified LDLOR versions, respectively named as IMED–LDLOR, SSSL–LDLOR and Bil–LDLOR, and for which we conduct a series of experiments on several image benchmark datasets and report comparison results in terms of the OR-specific evaluation criteria.

The remainder of the paper is organized as follows. In Section 2, we briefly review a representative OR, i.e., LDLOR, which is taken as the base model (baseline). In Section 3, three re-modeled LDLOR counterparts derived from three spatial structure information utilizing

Table 1
A comparative summarization of the three strategies.

Strategy type	Input pattern type	Embedding fashion to the objective
Metric embedding	Vector	Indirect
Structure regularization	Vector	Direct
Direct manipulation on images	Matrix	Direct

strategies are detailed. Section 4 shows the experimental results and gives comparison analysis. The conclusions are drawn in Section 5.

2. Review of LDLOR

LDLOR, one of the distinguished ORs, aims to find the best projection direction along which the ordinal indices of the ordered classes can be preserved well after projection. Based on this principle, LDLOR has two main characteristics: maximizing the distance between each pair of mean vectors of neighboring ordinal classes, and simultaneously minimizing the within-class scatters, which makes it different from the discriminant principles used in DA models such as LDA [45] due to the imposition of relative order constraints between the data classes on LDLOR.

Now let $(x_i, y_i) \in R^l \times R$, $i = 1, 2, ..., N$ be the training set, where $x_i \in R^l$ denotes the i-th instance, $y_i \in \{1, 2, ..., K\}$ denotes its corresponding class label, N is the data set size, and K the total number of classes. Then LDLOR can be formulated as

$$\min J(w, \rho) = w^T \cdot S_w \cdot w - C \cdot \rho$$
$$\text{s.t. } w^T \cdot (m_{k+1} - m_k) \geq \rho, \quad k = 1, 2, ..., K-1, \tag{1}$$

where S_w denotes the within-class scatter matrix expressed as

$$S_w = \frac{1}{N} \sum_{k=1}^{K} \sum_{x \in X_k} (x - m_k)(x - m_k)^T,$$

where $m_k = (1/N_k) \sum_{x \in X_k} x$ denotes the mean vector of the k-th class and N_k is the set size of this class.

The expression in (1) is a typical quadratic programming (QP) problem and thus can be solved directly or via its dual-form using Lagrangian theorem.

From the formulation of its objective, it can be seen that when applied to image classification, LDLOR also suffers from the spatial information loss owing to the image-to-vector conversion.

3. Three re-modeled LDLORs fused spatial structure information

In order to utilize the spatial structure information involved in such data as images to LDLOR, in the following sub-sections, we will briefly review the theorems of IMED [1], SSSL [9] and bilinear modeling [14] (as the representatives of three spatial information utilization strategies), and then employ them to re-model the basic LDLOR to generate its new variants: IMED–LDLOR, SSSL–LDLOR and Bil–LDLOR.

3.1. IMED–LDLOR

It is known that conventional ED is an often-used metric applied to measure the similarity or distance between two vectors. However, when used to images, it usually yields unreasonable metric results, due to that it neglects the spatial relationships among image pixels. Specifically, now let x and y be two $M \times N$ images, their vectorized versions respectively be $x = (x^1, x^2, ..., x^{MN})^T$ and $y = (y^1, y^2, ..., y^{MN})^T$. Then the ED $d_E(x, y)$ between x and y can be written as

$$d_E^2(x, y) = \sum_{k=1}^{MN} (x^k - y^k)^2 = (x - y)^T (x - y).$$

Obviously, it can be found that in $d_E^2(x, y)$, the spatial relationships between neighboring pixels are not reflected due to the fact that all pixels are independently treated with the same weight. However, when these pixels in image lattice are closer to each other, their corresponding gray values intuitively should be more similar. In other words, images usually have locally spatial smoothness in gray levels. It is such a consideration that Wang et al. [1] invented so-called IMED $d_{IMED}(x, y)$ from ED formulated as

$$d_{IMED}^2(x, y) = \sum_{i,j=1}^{MN} g_{ij}(x^i - y^i)(x^j - y^j) = (x - y)^T G (x - y),$$

where $G = (g_{ij})_{MN \times MN}$ and g_{ij} is defined as the weight between the i-th and j-th pixels of the vectorized image according to their geometric or spatial ED in original two-dimensional lattice, and generally inversely proportional to the ED value. By this way, the spatially smooth relationships between neighboring pixels are skillfully incorporated into the ED for reasonable image metric. However, in order to let IMED to be a valid metric, G must be positive semi-definite for which a Gaussian function is often used, leading to

$$g_{ij} = f(|P_i - P_j|) = \frac{1}{2\pi\delta^2} \exp\{-|P_i - P_j|^2 / 2\delta^2\},$$

where $P_i, P_j (i, j = 1, 2, ..., MN)$ respectively denote the i-th and j-th pixels in the vectorized image. Further, since the positive semi-definition of G, $d_{IMED}^2(x, y)$ can be expanded as

$$\begin{aligned} d_{IMED}^2(x, y) &= (x - y)^T G (x - y) \\ &= (x - y)^T G^{1/2} G^{1/2} (x - y) \\ &= (u - v)^T (u - v) \\ &= d_E^2(u, v) \end{aligned} \tag{2}$$

where $u = G^{1/2}x$, $v = G^{1/2}y$. Therefore, by the way of (2), the IMED between the images of x and y is actually equivalent to the ED between the new-projected u and v via a linear transform.

From Eq. (2), it can be found that (**1**) IMED is easily to be embedded into other metric models in a similar way; *and more importantly*, (**2**) by the way of linear projection transform, the spatial relationships between neighboring pixels are embedded into the metric transformation. Next we embed IMED to LDLOR to generate a spatial structure information compensated LDLOR, i.e., IMED–LDLOR expressed as

$$\min J(w, \rho) = w^T \cdot {S_w}^{IMED} \cdot w - C \cdot \rho$$
$$\text{s.t. } w^T \cdot (m_{k+1}^{IMED} - m_k^{IMED}) \geq \rho, \quad k = 1, 2, ..., K-1, \tag{3}$$

where ${S_w}^{IMED}$ denotes the within-class scatter matrix

$$S_w^{IMED} = \frac{1}{N} \sum_{k=1}^{K} \sum_{x \in X_k} (G^{1/2}x - G^{1/2}m_k)(G^{1/2}x - G^{1/2}m_k)^T,$$

in which $m_k^{IMED} = \frac{1}{N_k} \sum_{x \in X_k} G^{1/2}x$, $k = 1, 2, ..., K-1$ denotes the mean vector of the k-th ordinal class, K is the total number of classes, and N_k the sample size of the k-th class.

Similar to LDLOR in (1), IMED–LDLOR in (3) can be solved in the same way, thus here we omit its detail.

3.2. SSSL–LDLOR

In subspace learning such as LDA [45–46], LPP [46] and NPE [47], when operating on vectorized images they also need a spatial information compensation for such loss as spatially smooth information resulted from the image vectorization process. For this purpose, in Ref. [9], Cai et al. established a spatially smooth learning framework, i.e. SSSL, in which they proposed to incorporate the spatially smooth information into model learning by the way of regularization. Specifically, let w be a projection vector with the same dimension as that of a vectorized n_1-by-n_2 image, D_1 (D_2) be a $n_1 \times n_1 (n_2 \times n_2)$ second-order gradient smoothing operator or matrix here along the rows (columns) of an image and

formulated as

$$D_j=\frac{1}{h_j^2}\begin{pmatrix} -1 & 1 & & & & & & 0\\ 1 & -2 & 1 & & & & & \\ & 1 & -2 & 1 & & & & \\ & & & \cdot & \cdot & \cdot & & \\ & & & & 1 & -2 & 1 & \\ & & & & & 1 & -2 & 1\\ 0 & & & & & & 1 & -1 \end{pmatrix}$$

where $h_j=1/n_j$, $j=1,2$. Next for describing the whole smooth on image space, we introduce the global second-order gradient convolution matrix Δ by

$$\Delta=D_1\otimes I_2+D_2\otimes I_1, \tag{4}$$

where I_j is an $n_j\times n_j$ identity matrix for $j=1,2$, and $\otimes$ denotes the Kronecker operator. Using the convolution matrix Δ, we can evaluate the whole spatial smoothness of w by

$$\Im=||\Delta\cdot w||^2=w^T\Delta^T\Delta w=w^T\Upsilon w. \tag{5}$$

An intuitive interpretation of (5) is that the closer to each other the entries of w, the less the value of $\Im$, and vice versa. Thus by this way, the spatial smoothness can be reflected. Now, adding (5) to the objective function of the (basic) LDLOR, we can get the newly-modeled SSSL–LDLOR derived from the following problem

$$\begin{aligned}&\min J(w,\rho)=w^T\cdot S_w\cdot w+\lambda w^T\cdot\Upsilon\cdot w-C\cdot\rho\\ &\text{s.t. } w^T\cdot(m_{k+1}-m_k)\geq\rho,\quad k=1,2,...,K-1,\end{aligned} \tag{6}$$

By appropriately tuning the value of hyper-parameter λ, we can control the spatial smoothness with desirable trade-off. Intuitively, SSSL–LDLOR should be superior to LDLOR.

Similar to LDLOR, the SSSL–LDLOR in (6) can be solved directly or via its dual form.

3.3. Bil–LDLOR

Different from previous two spatial structure information compensation strategies on vectorized image, a more natural way is to establish a classifier directly on images (or their reshaped matrix patterns). Based on such a starting point, many studies have been developed [14–18]. Specifically, in Refs. [14–18], the authors designed a series of matrix-oriented classifiers by using the bilinear discriminant functions to replace the linear ones in existing vector-oriented classifiers such as Support Vector Machines (SVMs) [48] and Least Squares Support Vector Machines (LS-SVMs) [49], consequently, obtaining competitive performance in face recognition. Motivated by the above studies, *we likewise introduce the idea of bilinear modeling to OR for image classification* to develop a corresponding bilinear LDLOR, i.e., Bil–LDLOR. Though such an idea seems trivial, to our knowledge, there indeed has had no such an attempt yet.

To establish Bil–LDLOR, let us define $X_i\in R^{n_1\times n_2}$ as an image and a corresponding bilinear (discriminant) function operating on it as u^TX_iv, u and v respectively are the left and right weight vectors. Then we follow the basic LDLOR to establish our Bil–LDLOR as

$$\begin{aligned}&\min J(u,v,\rho)=\frac{1}{N}\sum_{k=1}^{K}\sum_{X_i^k\in X_k}u^T(X_i^k-M_k)vv^T(X_i^k-M_k)^Tu-C\cdot\rho\\ &\text{s.t. } u^T\cdot(M_{k+1}-M_k)\cdot v\geq\rho,\quad k=1,2,...,K-1,\end{aligned}, \tag{7}$$

where M_k denotes the mean matrix of the k-th class, $X_i^k\in R^{n_1\times n_2}$ is a sample from the k-th class set X_k, $k=1,2,\ldots,K$, and the meanings of all the other notations are the same as those in (1).

Compared with the derivation of basic LDLOR in (1), the objective of Bil–LDLOR in (7) brings several key advantages: (**1**) the left and right weight vectors $u\in R^{n_1}$ and $v\in R^{n_2}$ in (7) can respectively be determined by n_1 and n_2 free variables, totally being n_1+n_2, much fewer than $w\in R^{n_1\times n_2}$ in (1), thus reducing the VC-dimension and prone to avoiding the *over-fitting*, especially when the training set size is much fewer than the dimensionality; *more importantly*, (**2**) Bil–LDLOR can directly operate on matrix-pattern to avert the matrix-to-vector conversion, thus the spatial structure information involved in the data can be *more* desirably reflected. Consequently, Bil–LDLOR should beat the basic LDLOR when operating on such structured data as images.

Next let us describe the optimization process for (7). From the formulation of objective (7), we can find it is not jointly convex anymore w.r.t. (u, v) but just bi-convex w.r.t. either of them [50], i.e., it is convex w.r.t. u (v) for fixed v (u). Therefore, we can only adopt an alternating optimization iteration strategy to solve (7). The whole optimization procedure of Bil–LDLOR consists of two alternative optimization steps and is described as follows:

(a) Fixing v to optimize u

$$\begin{aligned}&\min J(u,\rho)=u^T\cdot S_w^v\cdot u-C\cdot\rho,\\ &\text{s.t. } u^T\cdot(m_{k+1}^v-m_k^v)\geq\rho,\quad k=1,2,...,K-1,\end{aligned} \tag{8}$$

where $S_w^v=\frac{1}{N}\sum_{k=1}^{K}\sum_{x^v\in X_k}(x^v-m_k^v)(x^v-m_k^v)^T$, $m_k^v=\frac{1}{N_k}\sum_{X\in X_k}X\cdot v$, and $x^v=X\cdot v$;

and

(b) Fixing u to optimize v

$$\begin{aligned}&\min J(v,\rho)=v^T\cdot S_w^u\cdot v-C\cdot\rho,\\ &s.t.\ (m_{k+1}^u-m_k^u)\cdot v\geq\rho, k=1,2,...,K-1,\end{aligned} \tag{9}$$

where $S_w^u=\frac{1}{N}\sum_{k=1}^{K}\sum_{x^u\in X_k}(x^u-m_k^u)(x^u-m_k^u)^T$, $m_k^u=\frac{1}{N_k}\sum_{X\in X_k}u^T\cdot X$, and $x^u=u^T\cdot X$. It can be easily found that both the sub-objectives of (8) and (9) are formally the same as that of (1), hence the implementation for (1) can be directly copied here.

Now we list the complete solving procedure of Bil–LDLOR in the following *Algorithm Bil-LDLOR*.

Algorithm Bil-LDLOR

Input: $X_1,X_2,...,X_N,n_1,n_2$
Output: u, v

1. Compute the mean matrix M_i of the i-th class
2. $v_0\overset{initialize}{\leftarrow}random(n_1,1)$;
3. For i from *1* to *maximal Iteration*
4. Fix v to optimize u using (8), $u_i\overset{update}{\leftarrow}u_{i-1}$;
5. Fix u to optimize v using (9), $v_i\overset{update}{\leftarrow}v_{i-1}$;
6. End For
7. $u\overset{update}{\leftarrow}u_{\text{maxIter}}$, $v\overset{update}{\leftarrow}v_{\text{maxIter}}$;
8. Return u and v.

With the biconvex optimization theory [50], we theoretically prove the *Algorithm Bil-LDLOR* able to converge a local minimum (for more details, please refer to Appendix) and experimentally also witness that it just takes several alternative iterations to convergence.

3.4. Overall comparisons between three re-modeled LDLORs

In previous sub-sections, we introduced three variants of LDLOR fused spatial structure information by three representative strategies, i.e., the metric embedding, the structure regularization,

Table 2
Overall summary for three re-modeled LDLORs.

OR method	Main idea	Convexity	Input pattern type	No. of variables[a]
IMED–LDLOR	Metric embedding	Yes	Vector	$n_1 \times n_2$
SSSL–LDLOR	Structure regularization	Yes	Vector	$n_1 \times n_2$
Bil–LDLOR	Matrix bilateral projection	No	Matrix	n_1+n_2

[a] n_1 and n_2 are respectively the no. of rows and columns of an image/matrix.

and the bilinear modeling. Here we briefly summarize a comparison between we re-modeled LDLORs in Table 2.

It can be noticed that from Table 2, both the objectives of IMED–LDLOR and SSSL–LDLOR can be solved by QP optimization, while Bil–LDLOR cannot due to its non-convexity in u and v, but still can be solved alternatively with convergence guarantee.

4. Experiments

In this section, we conduct experiments to make empirical comparisons among LDLOR, IMED–LDLOR, SSSL–LDLOR and Bil–LDLOR on three benchmark image datasets, i.e., JAFFE (for human facial expression intensity regression), UMIST (for human head pose regression), and FG-NET (for human age group regression), *their classes all are ordinal*. To eliminate the influence of image size to experiments, all images are cropped and resized to 16×16, and the raw (pixel) gray levels are directly extracted as features to represent images.

Considering OR's characteristics of both classification and regression, we use the most often-used criteria of mean absolute error (MAE) and classification accuracy (Acc) to respectively evaluate its regression deviation and classification performance. Specifically, $MAE = (1/N)\sum_{i=1}^{N}|l_i^{predicted} - l_i^{groundtruth}|$ denotes the average deviation of the prediction from the ground-truth and the lower its value, the better the regression performance; while $Acc = (N_{right}/N_{total})$ denotes the classification accuracy of a classifier, and the higher the Acc value, the better the classification performance. *On the other hand*, considering the duality of OR, we further use a recently-proposed *Ordinal Classification Index* (OCI) evaluation criterion OC_β^γ [51] that is specially designed for ORs, and the lower the OCI value, the better the OR performance. *It should be noted that different from MAE and Acc*, OC_β^γ can well eliminate the influence of numerical scales used to label ordinal classes on MAE and Acc and thus can more suitably measure the deviation of the predicted results from groundtruth. In the following experiments, we set parameter (γ, β) in OC_β^γ to (1, 0.75) as recommended in Ref. [51]. For more details about OC_β^γ, due to the complexity of its definition, we omit it here but please refer to the specific definition (7) in Ref. [51].

In our experiments, on each dataset, we adopt a nearest class-mean classifier to perform final ordinal classification and report the averaged results over 20 random splits by cross-validation (CV). It should be noted that due to the bi-convexity of Bil–LDLOR expressed in (7) and the random initialization of v_0 in *Algorithm Bil-LDLOR*, the resulting solution (convergent) is not so satisfactorily stable with varying initial value of v_0 as shown in Fig. 1. However, it can be further found that though such single solution can lead to a fluctuating or unstable performance on testing set, an averaged solution over many corresponding solutions to different initializations (such as 50 in Fig. 1, actually 10 is enough) of v_0 (or u_0) is quite stable. Thus in our experiments, for Bil–LDLOR, all the tabulated results are averaged over 10 repetitions on once training data split. *Here it is worthwhile pointing out* that except the Bil–LDLOR, the objectives of the remaining three ORs, i.e., LDLOR, IMED–LDLOR, and SSSL–LDLOR, are all convex, thus their individual globally optimal solutions can respectively be obtained and each can yield a stable performance unaffected by any random initialization, implying that their *averaging over the repetitions* as done for the Bil–LDLOR is always unchanged.

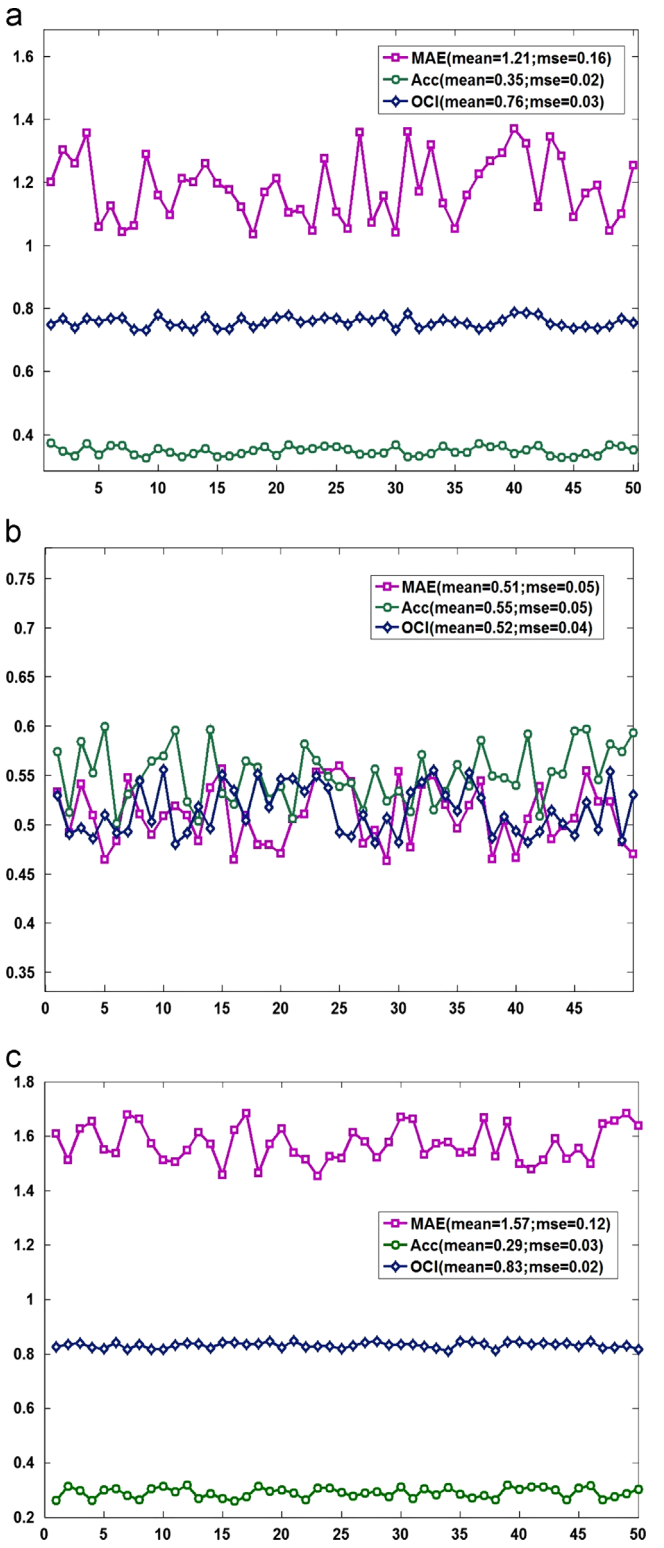

Fig. 1. The resulting solution of Bil–LDLOR with varying initialization of v_0 (the horizontal axis represents the number of repetitions in once data split, a total of 50 repetitions). (a) on JAFFE (# each class=12), (b) on UMIST (# each class=24) and (c) on FG-NET (# each class=18).

Besides above experiment setting, the tuning range of all hyper- or trade-off parameters involved in the experiments is {1e−5, 1e−3, 1e−1, 1e0, 1e1, 1e3, 1e5}.

4.1. JAFFE dataset

The original JAFFE dataset contains 213 images of 7 facial expressions (6 basic facial expressions+1 neutral) posed by 10 Japanese female models and each image has been rated on 6 emotion adjectives by 60 Japanese subjects. In the experiment, we select 29 samples each class, these selected samples cover all 7 (ordinal) facial expressions from disgust to surprise, and some examples of them are shown in Fig. 2.

The experimental results on JAFFE are tabulated in Tables 3a, 3b and 3c, respectively according to MAE, Acc and OC_β^γ. *Note that the underlined bold results* (*including the ones on UMIST and FG-NET*) *are statistically best compared with the other methods in the same row after* ***t-test*** (*significance value* $p=0.05$).

From Tables 3a, 3b and 3c respectively for the evaluation indices of MAE, Acc and OC, it can be observed that for facial expression regression on JAFFE, compared with the baseline LDLOR, SSSL- and Bil–LDLOR(s) both perform better in the three evaluation indices, especially SSSL–LDLOR, which partially indicates that ORs using *either direct spatially regularized objectives or direct operation on images* can outperform the corresponding vectorized versions. However, though embedded spatial information, IMED–LDLOR mostly behaves the worst (even worse than LDLOR in case of the number of training samples ranging from 12 to 24), slightly better just in MAE (Table 3a) when the training set is small, e.g., NPer=4. *On the other hand*, as the number of training samples grows from 4 to 24 (with an increment of 4), on the whole, the performances of all the approaches are getting better and better to different extents, especially both the SSSL–LDLOR and Bil–LDLOR can achieve more significant performance, respectively. For example, in the Acc index in Table 3b, the OR classification accuracy gets an improvement of about 20 percentages from 0.23 to 0.42 for SSSL based regularization and of 13 percentages from 0.27 to 0.40 for direct OR modeling. However, the Acc performance of IMED–LDLOR is increasing especially slow and not so distinct, i.e., just about 2 percentages from 0.20 to 0.22, and is merely a tenth of that of SSSL–LDLOR. Such a result may be due to that the embedding of the spatial structure information into ED is just for metric but not for final OR criterion to be optimized.

4.2. UMIST dataset

The original UMIST dataset consists of 564 images of 20 individuals. For the sake of OR experiment, we select 6 consecutive ordinal interval angles from profile to frontal views and each angle is associated with 56 samples. I.e., 6 ordinal head pose classes and each class with 56 samples are selected for head pose regression. Some samples are exhibited in Fig. 3.

The experimental results on UMIST are respectively listed in Tables 4a, 4b, and 4c as follows.

Investigating the results from Tables 4a–4c on UMIST for human head-pose regression, we can discover that (**1**) SSSL–LDLOR in all evaluation indices occupies the first position with absolute performance superiority. More importantly, with the increasing size of training set, its superiority is growing more obvious, e.g., in case of NPer=42 and 43, its Acc performance in

Table 3a
MAE comparisons among LDLORs on JAFFE (mean ± std-dev).

# NPer[a]	LDLOR	IMED–LDLOR	SSSL–LDLOR	Bil–LDLOR
4	1.92 ± 0.21	***1.60 ± 0.09***	1.78 ± 0.18	1.65 ± 0.14
8	1.55 ± 0.16	1.48 ± 0.07	1.43 ± 0.17	***1.29 ± 0.24***
12	1.28 ± 0.12	1.44 ± 0.05	***1.17 ± 0.15***	1.22 ± 0.24
16	1.22 ± 0.36	1.46 ± 0.24	***1.02 ± 0.10***	1.15 ± 0.23
20	1.06 ± 0.24	1.40 ± 0.06	***0.92 ± 0.13***	1.03 ± 0.27
24	0.95 ± 0.16	1.36 ± 0.06	***0.79 ± 0.14***	0.86 ± 0.13

The number of training samples of each ordinal class is written in italics; the experimental result is written in bold-italics when it is relatively the best but not statistically significant compared to the others in the same row; the experimental result is written in bold-italics-underline when it is the best one with statistical significance compared to the others in the same row.

[a] "**# NPer**" represents the number of training samples of each ordinal class (similarly hereinafter).

Table 3b
Acc comparisons among LDLORs on JAFFE (mean ± std-dev).

# NPer	LDLOR	IMED–LDLOR	SSSL–LDLOR	Bil–LDLOR
4	0.22 ± 0.03	0.20 ± 0.04	0.23 ± 0.04	***0.27 ± 0.04***
8	0.28 ± 0.04	0.22 ± 0.04	0.29 ± 0.04	***0.33 ± 0.04***
12	0.30 ± 0.03	0.21 ± 0.03	0.33 ± 0.04	***0.35 ± 0.05***
16	0.32 ± 0.07	0.20 ± 0.04	***0.38 ± 0.05***	0.37 ± 0.05
20	0.35 ± 0.06	0.23 ± 0.04	***0.40 ± 0.06***	0.39 ± 0.07
24	0.37 ± 0.10	0.22 ± 0.06	*0.42 ± 0.09*	0.40 ± 0.04

For the significance of italics, bold-italics, and bold-italics-underline, see footnote to Table 3a

Table 3c
OC_β^γ comparisons among LDLORs on JAFFE (mean ± std-dev).

# NPer	LDLOR	IMED–LDLOR	SSSL–LDLOR	Bil–LDLOR
4	0.88 ± 0.02	***0.84 ± 0.00***	0.87 ± 0.02	0.84 ± 0.03
8	0.84 ± 0.03	0.82 ± 0.02	0.81 ± 0.03	***0.79 ± 0.05***
12	0.79 ± 0.03	0.81 ± 0.02	***0.76 ± 0.03***	0.77 ± 0.06
16	0.76 ± 0.06	0.80 ± 0.04	***0.73 ± 0.04***	0.75 ± 0.06
20	0.73 ± 0.05	0.79 ± 0.02	***0.69 ± 0.04***	0.70 ± 0.07
24	0.69 ± 0.07	0.77 ± 0.02	***0.62 ± 0.07***	0.64 ± 0.04

For the significance of italics, bold-italics, and bold-italics-underline, see footnote to Table 3a.

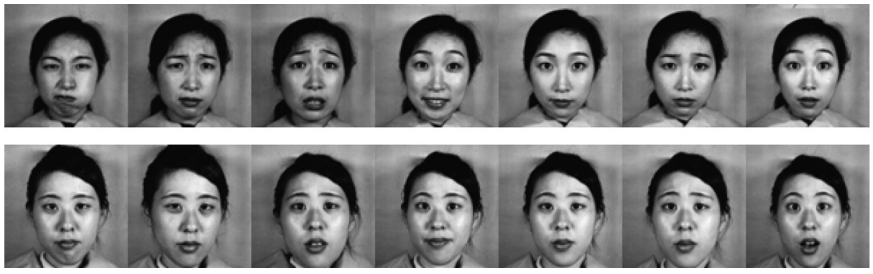

Fig. 2. Examples from JAFFE dataset.

Fig. 3. Examples from UMIST dataset.

Table 4a
MAE comparisons among LDLORs on UMIST (mean ± std-dev).

# NPer	LDLOR	IMED–LDLOR	SSSL–LDLOR	Bil–LDLOR
6	0.74 ± 0.11	0.85 ± 0.15	***0.72 ± 0.10***	0.81 ± 0.14
12	0.61 ± 0.08	1.00 ± 0.15	***0.57 ± 0.05***	0.59 ± 0.09
18	0.58 ± 0.07	1.12 ± 0.09	***0.48 ± 0.05***	0.54 ± 0.06
24	0.58 ± 0.08	1.09 ± 0.13	***0.43 ± 0.03***	0.50 ± 0.06
30	0.60 ± 0.05	1.05 ± 0.12	***0.38 ± 0.04***	0.49 ± 0.06
36	0.67 ± 0.11	1.14 ± 0.04	***0.37 ± 0.05***	0.46 ± 0.05
42	0.85 ± 0.15	1.15 ± 0.07	***0.33 ± 0.05***	0.45 ± 0.07
48	1.26 ± 0.29	1.21 ± 0.14	***0.30 ± 0.07***	0.42 ± 0.07

For the significance of italics, bold-italics, and bold-italics-underline, see footnote to Table 3a.

Table 4b
Acc comparisons among LDLORs on UMIST (mean ± std-dev).

# NPer	LDLOR	IMED–LDLOR	SSSL–LDLOR	Bil–LDLOR
6	0.43 ± 0.05	0.41 ± 0.05	***0.46 ± 0.05***	0.44 ± 0.05
12	0.49 ± 0.05	0.38 ± 0.05	***0.52 ± 0.04***	0.52 ± 0.05
18	0.51 ± 0.04	0.35 ± 0.05	***0.58 ± 0.04***	0.54 ± 0.06
24	0.51 ± 0.05	0.31 ± 0.03	***0.60 ± 0.03***	0.56 ± 0.04
30	0.49 ± 0.03	0.27 ± 0.03	***0.64 ± 0.03***	0.57 ± 0.04
36	0.45 ± 0.04	0.24 ± 0.03	***0.65 ± 0.05***	0.59 ± 0.04
42	0.39 ± 0.07	0.25 ± 0.02	***0.68 ± 0.05***	0.60 ± 0.06
48	0.32 ± 0.08	0.24 ± 0.03	***0.70 ± 0.06***	0.62 ± 0.06

For the significance of italics, bold-italics, and bold-italics-underline, see footnote to Table 3a.

Table 4c
OC_β^γ comparisons among LDLORs on UMIST (mean ± std-dev).

# NPer	LDLOR	IMED–LDLOR	SSSL–LDLOR	Bil–LDLOR
6	0.64 ± 0.05	0.67 ± 0.06	***0.62 ± 0.05***	0.67 ± 0.05
12	0.57 ± 0.05	0.72 ± 0.05	***0.55 ± 0.03***	0.56 ± 0.05
18	0.56 ± 0.04	0.72 ± 0.01	***0.49 ± 0.03***	0.53 ± 0.06
24	0.56 ± 0.05	0.71 ± 0.05	***0.46 ± 0.03***	0.51 ± 0.04
30	0.57 ± 0.03	0.71 ± 0.05	***0.41 ± 0.04***	0.50 ± 0.04
36	0.60 ± 0.05	0.73 ± 0.02	***0.40 ± 0.05***	0.48 ± 0.04
42	0.67 ± 0.06	0.74 ± 0.03	***0.36 ± 0.05***	0.46 ± 0.06
48	0.77 ± 0.08	0.77 ± 0.04	***0.33 ± 0.07***	0.43 ± 0.06

For the significance of italics, bold-italics, and bold-italics-underline, see footnote to Table 3a.

Table 4b is about two times that of the basic LDLOR and even three times of IMED–LDLOR, which shows that making use of the spatial information by the regularization is significantly effective; (**2**) the performance of Bil–LDLOR directly-modeled on images is better than both LDLOR and the IMED–LDLOR but inferior to SSSL–LDLOR, which witnesses that compared to the vectorized ORs without compensation of spatial information, direct manipulation on images can likewise make use of the spatial information and thus promote its OR performance; and (**3**) IMED–LDLOR mostly yields the worst performance, e.g., just an average Acc of 0.31, even inferior to 0.45 of the LDLOR. More surprisingly, with the increasing training samples, its performances on UMIST and JAFFE do not monotonically increase as expected but significantly fluctuate, which seems counterintuitive. Such an occurrence, besides the similar reason analyzed in Section 4.1, may be further owing to the unaligned head poses in images of this dataset.

4.3. FG-NET dataset

The FG-NET dataset contains a number of individuals aging from 0 to 69. In our experiment, we divide all the samples into 8 ordinal categories, i.e., 0–1 years old, 2–4 years old, 5–8 years old, 9–12 years old, 13–16 years old, 17–29 years old, 30–43 years old, and 44–69 years old. 43 typical samples for each category are selected and some examples of them are shown in Fig. 4.

The age group regression results on FG-NET are respectively listed in Tables 5a, 5b and 5c.

From the results on FG-NET of age group regression, we can discover some hints: *On the one hand*, almost all the best results in performance are led by SSSL–LDLOR or Bil–LDLOR, and in particular for Acc (Table 5b), the latter LDLOR variant stays ahead with about 4 percentage points defeating the former one. Besides that, either for MAE, Acc or OCI, the performances of both SSSL–LDLOR and Bil-LDLOR have been improved with distinct significance, e.g., with the training samples increasing from 6 to 36, their Acc performances are increased by about 9 percentages (respectively from 0.19 to 0.28 and from 0.23 to 0.32). By contrast, the performance of neither the basic LDLOR nor IMED-based one has essentially been increased, e.g., for Acc index, their percentage points are improved respectively by merely about 3. *On the other hand*, with the increasing size of training set, all the indices of both SSSL–LDLOR and Bil–LDLOR are monotonically improved with significant extent, while those of both the basic and the IMED-based ones, however, do not emerge a similar monotonic trend. The reasons behind can similarly be analyzed as in Sections 4.1 and 4.2, thus are omitted here.

4.4. Brief summary

Now jointly from all the above experimental results and analyses, we can find that for OR on image dataset, both SSSL–LDLOR and Bil–LDLOR can make good use of the spatial information involved in the data and consequently improve their OR performance with significance. By analyzing their essences, we can witness that both SSSL–LDLOR and Bil–LDLOR impose the spatial smooth constraints to the OR objectives, thus improve their OR performance through purposely respecting the prior spatial knowledge. *It is worth noting that* though the objective of Bil–LDLOR is not jointly convex but biconvex, using *Algorithm Bil-LDLOR*, we can get a convergent solution within about 10 iterations as illustrated in Fig. 5.

Fig. 4. Examples from FG-NET dataset.

Table 5a
MAE comparisons among LDLORs on FG-NET (mean ± std-dev).

# NPer	LDLOR	IMED–LDLOR	SSSL–LDLOR	Bil–LDLOR
6	1.94 ± 0.20	***1.77 ± 0.07***	2.07 ± 0.25	1.84 ± 0.16
12	1.65 ± 0.17	1.74 ± 0.22	1.71 ± 0.16	***1.60 ± 0.17***
18	1.71 ± 0.18	1.69 ± 0.06	***1.51 ± 0.12***	1.57 ± 0.14
24	1.76 ± 0.12	1.70 ± 0.07	***1.41 ± 0.08***	1.51 ± 0.17
30	2.18 ± 0.26	1.79 ± 0.13	***1.30 ± 0.10***	1.50 ± 0.15
36	2.26 ± 0.18	1.87 ± 0.11	***1.27 ± 0.12***	1.49 ± 0.20

For the significance of italics, bold-italics, and bold-italics-underline, see footnote to Table 3a.

Table 5b
Acc comparisons among LDLORs on FG-NET (mean ± std-dev).

# NPer	LDLOR	IMED–LDLOR	SSSL–LDLOR	Bil–LDLOR
6	0.19 ± 0.03	0.18 ± 0.02	0.19 ± 0.03	***0.23 ± 0.03***
12	0.22 ± 0.03	0.21 ± 0.02	0.21 ± 0.04	***0.27 ± 0.02***
18	0.22 ± 0.03	0.19 ± 0.02	0.24 ± 0.04	***0.29 ± 0.02***
24	0.22 ± 0.03	0.17 ± 0.02	0.27 ± 0.03	***0.31 ± 0.03***
30	0.19 ± 0.04	0.16 ± 0.02	0.27 ± 0.02	***0.31 ± 0.04***
36	0.20 ± 0.07	0.15 ± 0.03	0.28 ± 0.03	***0.32 ± 0.06***

For the significance of italics, bold-italics, and bold-italics-underline, see footnote to Table 3a.

Table 5c
OC'_β comparisons among LDLORs on FG-NET (mean ± std-dev).

# NPer	LDLOR	IMED–LDLOR	SSSL–LDLOR	Bil–LDLOR
6	0.89 ± 0.02	***0.86 ± 0.01***	0.90 ± 0.02	0.87 ± 0.01
12	0.86 ± 0.02	0.85 ± 0.01	0.87 ± 0.02	***0.84 ± 0.02***
18	0.87 ± 0.02	0.85 ± 0.01	0.84 ± 0.02	***0.83 ± 0.01***
24	0.87 ± 0.02	0.85 ± 0.01	0.82 ± 0.02	***0.81 ± 0.02***
30	0.90 ± 0.03	0.87 ± 0.00	***0.80 ± 0.02***	0.81 ± 0.03
36	0.90 ± 0.02	0.87 ± 0.00	***0.79 ± 0.03***	0.80 ± 0.04

For the significance of italics, bold-italics, and bold-italics-underline, see footnote to Table 3a.

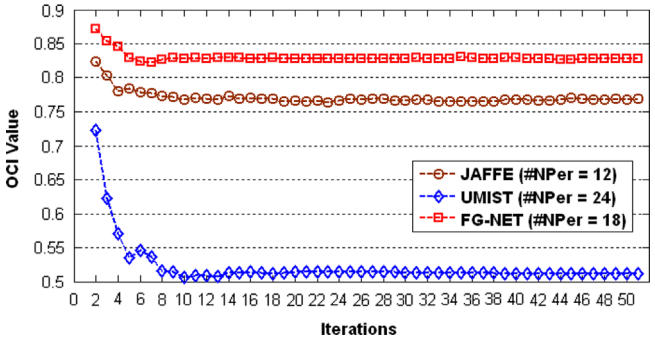

Fig. 5. Relationship between convergence under OCI and alternate iterations of Bil–LDLOR

By contrast, though embedded the spatial information, IMED just reflects the utilization of spatial information in metric rectification rather than in the OR objective optimization, thus cannot definitely guarantee desirable results. In addition, *it is worth to note that* from the comparison between Table 3a vs. 3b and 5a vs. 5b, some inconsistencies also exist among the evaluation indices. For example, in case of NPer equal to 18, 24, 30 and 36 on FG-NET respectively for human head pose regression, compared to SSSL–LDLOR, all the Acc results of Bil–LDLOR are all significantly better but neither of its MAE results is dominant, which indicates that *good classification performance does not necessarily mean good regression performance, and* vice versa. Therefore, neither MAE nor Acc is comprehensive enough to afford the evaluation for ORs. Relative to MAE and Acc, OCI is a more preferable OR-specific measure index in that its definition is more prone to OR nature than MAE or Acc through eliminating the influence of numerical scales used to label ordinal classes, thus we recommend biasedly OCI (OC'_β) as a reasonable evaluation index in ordinal classification or regression task.

5. Conclusions

In this paper, ***first*** through a systematic summary for separately-proposed spatial structure information utilization schemes, we classified them into three main categories of the structure-embedded Euclidean distance preserving, the structure-regularized modeling and the direct manipulation on images; ***second***, to further make a comparison among them in conditions that the spatial structure information is rarely reflected in existing image-oriented ORs, we respectively took IMED, SSSL and Bilinear modeling as their representatives, and applied them to re-model the LDLOR (as the baseline/basic approach of OR) and develop corresponding variants: IMED–LDLOR, SSSL–LDLOR, and Bil–LDLOR, and then conducted sufficient experiments on JAFFE, UMIST and FG-Net respectively for human facial expression, head pose and age group regressions, with conclusions that

(a) Direct OR modeling methods on images, such as Bil–LDLOR, can effectively preserve and utilize the spatial information involved in the images to some extent by a similar way as 2DPCA [52] and thus achieve a significant OR performance improvement.
(b) The structure-regularized based ORs, such as SSSL–LDLOR, as well can achieve distinct benefit in OR performance by imposing an regularization in terms of spatial information into their objectives.
(c) The structure-embedded ORs, such as IMED–LDLOR, though embedded the spatial information, usually cannot yield a significance in improving the OR performance. The reason lies in that the spatial information is not taken into account for objective optimization and that it is further affected by some other potential factors to be discovered.
(d) In OR experiments, the results from the indices MAE (used to evaluate the regression deviation) and Acc (measuring the classification accuracy) are not always consistent, which is due

to that they are not bound together in optimization. And in view of the duality of OR, adopting the OR-specific OCI to more comprehensively evaluate ORs is reasonable and recommended. Moreover, by comparing the results among Tables 3c, 4c and 5c w.r.t. the OCI (OC_β^γ), we can find that the difficulties of human facial expression OR on JAFFE and age OR on FG-NET are almost at the same level, both harder than that of human head pose regression on UMIST. That is, human head pose regression is relatively easy than the other two, this is consistent with human intuition.

From the comparative study in this paper, we can see that both the two strategies of structure-regularized and direct manipulation on images can well obtain a distinct improvement in OR performance by directly imposing the spatial information in their objectives respectively through direct manipulation or structure regularization, while the third category of structure-embedded, however, cannot generate performance benefits as intuitively expected, where the spatial information is just embedded for the metric modification not directly related to the OR objective. Therefore, we can infer that whether the spatial information can boost the performance of a classifier (or regressor) depends on the embedding way of the spatial structure information.

Acknowledgments

This work is partially supported by NSFC (61170151 and 61073112), Jiangsu SFC(BK2012793), Research Fund for the Doctoral Program (RFDP) (20123218110033), Funding of Jiangsu Innovation Program for Graduate Education (CXLX13_159), the Fundamental Research Funds for the Central Universities (NZ2013306) and Jiangsu Qinglan project.

Appendix

According to Ref. [50], the sequence of solving a biconvex objective can converge if it holds that (1) its objective value is lower-bounded and (2) its value sequence can decrease monotonically during the optimization process. Accordingly, we give the detailed proof for the convergence of the bi-convex Bil–LDLOR as follows.

The objective of Bil–LDLOR is lower-bounded

The function value of objective (7) is lower-bounded, because the objective of (7) equals

$$
\begin{aligned}
J(u,v,\rho) &= \frac{1}{N}\sum_{k=1}^{K}\sum_{X_i^k\in X_k} u^T(X_i^k-M_k)vv^T(X_i^k-M_k)^Tu - C\cdot\rho \\
&= \frac{1}{N}\sum_{k=1}^{K}\sum_{X_i^k\in X_k} (u^T\cdot X_i^k\cdot v - u^T\cdot M_k\cdot v)(u^T\cdot X_i^k\cdot v - u^T\cdot M_k\cdot v)^T - C\cdot\rho \\
&= \frac{1}{N}\sum_{k=1}^{K}\sum_{X_i^k\in X_k} ||u^T\cdot X_i^k\cdot v - u^T\cdot M_k\cdot v||^2 - C\cdot\rho \\
&\geq \frac{1}{N}\sum_{k=1}^{K}\sum_{X_i^k\in X_k} ||u^T\cdot X_i^k\cdot v - u^T\cdot M_k\cdot v||^2 - C \\
&\quad\cdot \min_{k=1,2,\ldots,K-1}\{u^T\cdot(M_{k+1}-M_k)\cdot v\} \\
&\quad (\textit{Note}: \text{ The second term is derived from the constraints in (7)}) \\
&\geq \frac{1}{N}\sum_{k=1}^{K}\sum_{X_i^k\in X_k} ||u^T\cdot X_i^k\cdot v - u^T\cdot M_k\cdot v||^2 - C \\
&\quad\cdot \min_{k=1,2,\ldots,K-1}\{||u^T||\cdot||(M_{k+1}-M_k)||\cdot||v||\}
\end{aligned}
$$

in which the first term and the trade-off parameter C are non-negative, and for given training set, the $||(M_{k+1}-M_k)||$ is a constant and $\min_{k=1,2,\ldots,K-1}\{||u^T||\cdot||(M_{k+1}-M_k)||\cdot||v||\}$ is bounded. Thus $J(u,v,\rho)$ is lower-bounded.

The objective value of Bil–LDLOR decreases monotonically during the optimization process using Algorithm Bil–LDLOR

Here, let $J(u^i,v^i,\rho^i)$ denote the function value of objective (7) at the i-th iteration using *Algorithm Bil-LDLOR*. For fixed v, the formula (7) is transformed to (8). Obviously, problem (8) is a convex quadratic programming w.r.t. u, thus we can adopt any off-the-shelf optimization approach (such as SMO [53]) to obtain a unique globally optimal solution u^{i+1}, its corresponding function value is $J(u^{i+1},v^i,\rho^i)$, as a result, $J(u^i,v^i,\rho^i)\geq J(u^{i+1},v^i,\rho^i)$; next for fixed u^{i+1}, the formula (7) is transformed to (9), the latter is as well a convex quadratic programming w.r.t. (v,ρ), as done in the previous iteration, we also get a globally optimal solution (v^{i+1},ρ^{i+1}), whose functional value is $J(u^{i+1},v^{i+1},\rho^{i+1})$, consequently, $J(u^{i+1},v^i,\rho^i)\geq J(u^{i+1},v^{i+1},\rho^{i+1})$, finally $J(u^i,v^i,\rho^i)\geq J(u^{i+1},v^i,\rho^i)\geq J(u^{i+1},v^{i+1},\rho^{i+1})$, implying the iteration sequence $\{J(u^i,v^i,\rho^i)\}$ decreases monotonically. Incorporating that $J(u,v,\rho)$ is lower-bounded, consequently, the sequence $\{J(u^i,v^i,\rho^i)\}$ of solving objective (7) can converge.

References

[1] L. Wang, Y. Zhang, J. Feng, On the Euclidean distance of images, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2) (2005) 1334–1339.
[2] T. Tangkuampien, D. Suter, 3D Object Pose Inference via Kernel Principal Component Analysis with Image Euclidean Distance, in: BMVC, 2006.
[3] J. Chen, R. Wang, S. Shan, et al. Isomap Based on the Image Euclidean Distance, in: ICPR, 2006.
[4] J. Li, B. Lu., A framework for multi-view gender classification, Lecture Notes Comput. Sci. 4984 (2008) 973–982.
[5] Y. Liu, Y. Liu, K. Chan., Tensor distance based multilinear locality preserved maximum information embedding, IEEE Trans. Neural Netw. 21 (11) (2010) 1848–1854.
[6] B. Sun, J. Feng, L. Wang, Learning IMED via Shift-Invariant Transformation, in: CVPR, 2009.
[7] W. Zuo, H. Zhang, D. zhang, et al., Post-processed LDA for face and palmprint recognition: what is the rationale, Signal Process. 90 (2010) 2344–2352.
[8] J. Li, B. Lu, An adaptive image Euclidean distance, Pattern Recognit. 42 (2009) 349–357.
[9] D. Cai, X. He, Y. Hu, et al., Learning a Spatial Smooth Subspace for Face Recognition, in: CVPR, 2007.
[10] S. Gu, Y. Tan, X. He., Laplacian smoothing transform for face recognition, Sci. China 53 (12) (2010) 2415–2428.
[11] Z. Lei, S. Li, Contextual constraints based linear discriminant analysis, Pattern Recognit. Lett. 32 (2011) 626–632.
[12] W. Zuo, L. Liu, K. Wang, et al., Spatially Smooth Subspace Face Recognition Using LOG and DOG Penalties, in: ISNN, 2009.
[13] X. Chen, Z. Tong, H. Liu, et al., Metric Learning with Two-Dimensional Smoothness for Visual Analysis, in: CVPR, 2012.
[14] S. Chen, Z. Wang, Y. Tian., Matrix-pattern-oriented Ho-Kashyap classifier with regularization learning, Pattern Recognit. 40 (2007) 1533–1543.
[15] Z. Wang, S. Chen, New least squares support vector machines based on matrix patterns, Neural Process. Lett. 26 (2007) 41–56.
[16] Z. Wang, S. Chen, Matrix-pattern-oriented least squares support vector classifier with AdaBoost, Pattern Recognit. Lett. 29 (6) (2008) 745–753.
[17] Z. Wang, C. Zhu, D. Gao, et al., Three-fold structured classifier design based on matrix pattern, Pattern Recognit. 46 (2013) 1532–1555.
[18] Z. Wang, S. Chen, J. Liu, et al., Pattern representation in feature extraction and classifier design: matrix versus vector, IEEE Trans. Neural Netw. 19 (5) (2008) 758–769.
[19] Z. Zhang, T. Chow, Maximum margin multisurface support tensor machines with application to image classification and segmentation, Expert Syst. Appl. 39 (2012) 849–860.
[20] D. Tao, X. Li, X. Wu, et al., General tensor discriminant analysis and gabor features for gait recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (10) (2007) 1700–1715.
[21] D. Tao, M. Song, X. Li, et al., Bayesian tensor approach for 3-D face modeling, IEEE Trans. Circuits Syst. Video Technol. 18 (10) (2008) 1397–1410.
[22] D. Tao, X. Li, X. Wu, et al., Tensor rank one discriminant analysis – a convergent method for discriminative multilinear subspace selection, Neurocomputing 71 (2008) 1866–1882.

[23] J. Wen, X. Gao, Y. yuan, et al., Incremental tensor biased discriminant analysis: a new color-based visual tracking method, Neurocomputing 73 (2010) 827–839.
[24] B. Wang, X. Gao, D. Tao, et al., A unified tensor level set for image segmentation, IEEE Trans. Syst. Man Cybern. -Part B: Cybern. 40 (3) (2010) 857–867.
[25] L. Zhang, L. Zhang, D. Tao, et al., Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction, IEEE Trans. Geosci. Remote Sens. 51 (1) (2013) 242–256.
[26] K. Lakiotaki, N. Mastsatsinis, A. Tsoukias, Multicriteria user modeling in recommender systems, IEEE Intell. Syst. 26 (2) (2011) 64–76.
[27] T. Joachims, Optimizing Search Engineer Using Click-Through Data, in: ACM SIGKDD, 2012.
[28] H. Wu, H.Q. Lu, S.D. Ma, A Practical SVM-Based Algorithm for Ordinal Regression in Image Retrieval, in: ACM Multimedia, 2003.
[29] D. Zhang, Y. Wang, L. Zhou, et al., Multimodal classification of Alzheimer's disease and mild cognitive impairment, NeuroImage 55 (2011) 856–867.
[30] K. Gray, P. Aljabar, R. Heckemann, et al., Random forest-based similarity measures for multi-modal classification of Alzheimer's disease, NeuroImage 65 (2013) 167–175.
[31] C. Li, Q. Liu, J. Liu, et al., Learning Ordinal Discriminative Features for Age Estimation, in: CVPR, 2012.
[32] K. Chang, C. Chen, Y. Huang, Ordinal Hyperplanes Ranker With Cost Sensitivies for Age Estimation, in: CVPR, 2011.
[33] R. Herbrich, T. Graepel, K. Obermayer, Support Vector Learning for Ordinal Regression, in: ICANN, 1999.
[34] W. Chu, S. Keerthi, New Approaches to Support Vector Ordinal Regression, in: ICML, 2005.
[35] W. Chu, S. Keerthi, Support vector ordinal regression, Neural Comput. 19 (3) (2007) 792–815.
[36] R. Herbrich, T. Graepel, K. Obermayer., Large Margin Rank Bound Arises for Ordinal Regression, MIT Press, Cambridge, MA (2000) 115–132.
[37] A. Shashua, A. Levin, Ranking with Large Margin Principle: Two Approaches, in: NIPS, 2003.
[38] S. Kramer, G. Widmer, B. Pfahringer, et al., Prediction of ordinal classes using regression trees, Fundam. Inform. 47 (2001) 1–13.
[39] J. Cheng, Z. Wang, G. Pollastri, A neural network approach to ordinal regression, in: Proceedings of the IEEE International Joint Conference on Neural Networks, 2008.
[40] S. Fouad, P. Tino, Adaptive metric learning vector quantization for ordinal classification, Neural Comput. 24 (2012) 2825–2851.
[41] C. Seah, I. Tsang, Y. Ong, Transductive ordinal regression, IEEE Trans. Neural Netw. Learn. Syst. 23 (7) (2012) 1074–1086.
[42] Y. Liu, Y. Liu, K. Chan, Ordinal Regression Via Manifold Learning, in: AAAI, 2011.
[43] Y. Liu, Y. Liu, S. Zhong, et al., Semi-Supervised Manifold Ordinal Regression for Image Ranking, in: ACM Multimedia, 2011.
[44] B. Sun, J. Li, D. Wu, et al., Kernel discriminant learning for ordinal regression, IEEE Trans. Knowl. Data Eng. 22 (6) (2010) 906–910.
[45] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley-Interscience, 2000.
[46] X. He, S. Yan, Y. Yu, et al., Face recognition using Laplacian faces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.
[47] X. He, D. Cai, S. Yan, et al., Neighborhood Preserving Embedding, in: ICCV, 2005.
[48] V. Vapnil, Statistical Learning Theory, John Wiley & Sons, New York, 1998.
[49] J. Suykens, J. Vandewalls, Least squares support vector machine classifiers, Neural Process Lett. 9 (1999) 293–300.
[50] J. Gorski, F. Pfeuffer, K. Klamroth, Biconvex sets and optimization with biconvex functions: a survey and extensions, Math. Methods Oper. Res. 66 (3) (2007) 373–407.
[51] J. Cardoso, R. Sousa, Measuring the performance of ordinal classification, Int. J. Pattern Recognit. Artif. Intell. 25 (8) (2011) 1173–1195.
[52] J. Yang, D. Zhang, A. Frangi, et al., Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) (2004).
[53] B.E. Boser, I.M. Guyon, and V.N. Vapnik, A training algorithm for optimal margin classifier, in: Proceedings of the 5th ACM Workshop on Computational Learning Theory, 1992.

Qing Tian received the B.S. degree in computer science from Southwest University for Nationalities, China, and the M.S. degree in computer science from Zhejiang University of Technology, China, respectively with the honors of *Sichuan provincial level outstanding graduate* and *Zhejiang provincial level outstanding graduate* in 2008 and 2011. From February 2011 to February 2012, as a researcher in the field of gender/age recognition, he worked in Arcsoft, U.S. Now he is a Ph.D. candidate in computer science at Nanjing University of Aeronautics and Astronautics, and his current research interests include machine learning and pattern recognition.

Songcan Chen received the B.S. degree from Hangzhou University (now merged into Zhejiang University), the M.S. degree from Shanghai Jiao Tong University and the Ph.D. degree from Nanjing University of Aeronautics and Astronautics (NUAA) in 1983, 1985, and 1997, respectively. He joined in NUAA in 1986, and since 1998, he has been a full-time Professor with the Department of Computer Science and Engineering. He has authored/co-authored over 170 scientific peer-reviewed papers and ever obtained Honorable Mentions of 2006, 2007 and 2010 Best Paper Awards of Pattern Recognition Journal respectively. His current research interests include pattern recognition, machine learning, and neural computing.

Xiaoyang Tan received his B.S. and M.S. degrees in computer applications from Nanjing University of Aeronautics and Astronautics (NUAA) in 1993 and 1996, respectively. Then he worked at NUAA in June 1996 as an assistant lecturer. He received a Ph.D. degree from Department of Computer Science and Technology of Nanjing University, China, in 2005. From September 2006 to October 2007, he worked as a postdoctoral researcher in the LEAR (Learning and Recognition in Vision) team at INRIA Rhone-Alpes in Grenoble, France. His research interests are in face recognition, machine learning, pattern recognition, and computer vision.