Corresponding Author: Mr. Songcan Chen, PhD

Corresponding Author's Institution: Nanjing University of Aeronautics & Astronautics

First Author: Qing Tian

Order of Authors: Qing Tian; Songcan Chen, PhD; Xiaoyang Tan, PhD

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.

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, P.R.China

{tianqing, s.chen, x.tan}@nuaa.edu.cn

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.

Keywords: ordinal regression; vector pattern; matrix pattern; tensor pattern; spatial structure; Euclidean distance; bilinear

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 conducted, 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 classify them into three categories; then for making a comparison, we choose one of currently popular topics in image classification, i.e., image-oriented ORs such as age estimation of face images, as a comparative platform, due to that *1) these ORs designed specially for 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 image classification such as face recognition.* And *next*, we develop three image-oriented OR variants by the compensation of spatial information with the aforementioned three strategies and then make a relatively comprehensive 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 using spatial structure information

In this subsection we analyze the existing separate schemes of the use of spatial structure information and summarize them into three main families as follows:

a) 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 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 [1] can be viewed as their representative. In [1], the authors developed the IMage Euclidean Distance (IMED) through 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 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 in [8] developed an adaptive IMED (AIMED) by further fusing gray level knowledge of image to IMED besides the spatial information to achieve more satisfactory identification performance for human face and handwritten digit. 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.

b) 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 (from optimizing the objective) is spatially smooth as much as possible [9-13]. The spatial smooth subspace learning (SSSL) proposed in [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 function to realize multi-scale image smoothing. Moreover, Chen et al. in [13] developed a regularized metric learning framework by imposing the Laplacian penalty and achieved competitive face recognition performance on several benchmark datasets. From these related researches it can easily be found that, structure-regularized modeling indeed can compensate the spatial information loss induced by tensor- or matrix-to-vector conversion. In the following comparative study, without loss of generality, we will take the SSSL as the basic regularization strategy. Actually, its extensions or variants can be employed in a similar way.

c) 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 related learning performance is thus

boosted, these strategies usually suffer from (1) high computational complexity; and (2) so-called "*small sample problem*", i.e., in which dimensionality of feature vector is higher than the sample set size, easily leading to over-fitting. Hence, a natural way to mitigate or handle these problems is operating directly image (or reshaped image) patterns. Along this line, many studies have been developed, for example, [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], directly based 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. implemented their dimensionality reduction or classification modeling directly manipulated on (high order) tensor patterns and applied them respectively for human gait recognition [20] and visual tracking [25]. Likewise, it can be found that direct operation on matrix or tensor data indeed can make more sufficient use of the inherent spatial structure information of data themselves, thus promoting the performance more sufficiently, compared with vector pattern counterparts.

1.3. Review of OR

Following the summary and categorization on separate spatial structure information utilization strategies, our next step is in position to taking the OR as a research platform, on which we will make an empirical comparison on three benchmark image datasets among our afore-summarized three families. As for OR, actually it is a special learning strategy used to design classifiers for ordinal classes for example, age estimation for human beings, thus its outputs are ordered discrete labels, leading to its duality of regression and classification. Due to its power, OR has so far been widely applied in such domains as recommender system [26], web page ranking [27], image retrieval [28], medical image diagnosis [29-30] and age estimation [31-32], and 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 is trivial, to the best of our knowledge, there has indeed no related study done yet.

Now for the sake of clarity, and without loss of generality, we will just take the linear version of KDLOR, a typical OR model proposed in [44], as a 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 called 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. In section 3, three re-modeled LDLOR counterparts derived from three spatial structure information utilizing 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 orderly 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 distinguish from the discriminant principles used in DA models such as LDA [45] due to the imposition of relative order constraint between all the data classes on LDLOR.

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

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

where S_w denotes the within-class scatter matrix as follows

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

where $m_k = \frac{1}{N_k} \sum_{x \in X_k} x$ is the mean vector of the k-th class and N_k is the corresponding

sample set size.

The formulation 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, when directly applied to image classification, LDLOR also suffers from the spatial information loss resulted from 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 first briefly review the implementations 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 form its new variants: IMED-LDLOR, SSSL-LDLOR and Bil-LDLOR.

3.1. IMED-LDLOR

It is known that conventional ED is one 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 its neglect for 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 is

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

Obviously, it can be found that in $d_E^2(x, y)$, the spatial relationships between pixels are not reflected due to 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 also 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 $d_{IMED}(x, y)$

based on ED which is formulated as follows:

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

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. As a result, the spatially smooth relationships between neighboring pixels are incorporated into the ED for reasonable image metric. However, in order to ensure IMED to be a valid metric, G must be positive semi-definite for which a Gaussian function is often used, leading to

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

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

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

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-converted 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 as (2); *and more importantly*, **2)** by the way of linear 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 which can be obtained by optimizing the following objective:

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

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

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

where $m_k^{IMED} = \frac{1}{N_k} \sum_{x \in X_k} G^{1/2}x, k = 1, 2, ..., K-1$ denotes the mean vector of the k-th ordinal

class, K is the class number, 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 we omit

its detail.

3.2. SSSL-LDLOR

In subspace learning such as LDA [45-46], LPP [46] and NPE [47], when operating on the

vectorized images, they also need a spatial information compensation for such loss as spatially

smooth information resulted from image vectorization process. In [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 image and be formulated as

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

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

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

. The 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 spatial smoothness can be reflected. Now, adding (5) to the objective function of basic LDLOR, we get the newly-modeled SSSL-LDLOR derivable from the following problem

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

By appropriately tuning the value of hyper-parameter λ, we can control spatial smoothness for good 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 for vectorized image, a more natural strategy is to establish a classifier directly on original images (or their reshaped matrix patterns). Based on such a starting point, there are many studies developed [14-18]. In [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 better recognition performance on face images. Inspired by the above studies, *we likewise introduce the idea of bilinear modeling to OR for image classification* to develop corresponding bilinear LDLOR, for short, Bil-LDLOR. Though such a idea is trivial, to our knowledge, there has indeed had no such an attempt.

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 is described as $u^T X_i v$, u and v respectively are the left and right weight vectors. Then we follow basic LDLOR to establish our Bil-LDLOR by

$$\min J(u, v, p) = \frac{1}{N} \sum_{k=1}^{K} \sum_{X_i^k \in X_k} u^T \left(X_i^k - M_k \right) v v^T \left(X_i^k - M_k \right)^T u - C \cdot \rho$$
$$s.t. \ u^T \cdot (M_{k+1} - M_k) \cdot v \geq \rho, k = 1, 2, ..., K - 1, \tag{7}$$

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, .., K, and the meaning 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) whose dimensionality is $n_1 n_2$, thus reducing the VC-dimension and more likely avoiding the *over-fitting* risk, especially when the size of training set is much less than the dimensionality; *more importantly*, **2)** Bil-LDLOR directly operates on matrix-pattern to avert the matrix-to-vector conversion, thus the spatial structure information involved in the data can be *more* desirably reflected. Intuitively, Bil-LDLOR should outperform the basic LDLOR in OR performance when operating on such structured data as images.

It can be observed that the objective (7) is not convex anymore as in (1) but bi-convex w.r.t. u and v, i.e., it is convex in u (v) for fixed v (u). Fortunately, we can adopt the off-the-shelf alternating iteration strategy to solve them. *More importantly*, it has theoretically been proven that such iteration process can converge to a local optimum [50], and experimentally shown that it just needs several rounds of alternative iterations to convergence.

The whole optimization procedure of Bil-LDLOR includes two main alternative optimization steps described as follows:

a) fixing v to optimize u

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

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

and

b) fixing u to optimize v

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

where $S_w^u = \dfrac{1}{N}\displaystyle\sum_{k=1}^{K}\sum_{x^u \in X_k}\left(x^u - m_k^u\right)\left(x^u - m_k^u\right)^T$, $m_k^u = \dfrac{1}{N_k}\displaystyle\sum_{X \in X_k} u^T \cdot X$, 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 give the complete solving procedure of Bil-LDLOR, illustrated in 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. $v_0 \xleftarrow{\;assign\;} random(n_1,1)$;

3. For i from 1 to *maximal Iteration*

4.   Fix v to optimize u using (8),

    $u_i \xleftarrow{\;update\;} u_{i-1}$;

5.   Fix u to optimize v using (9),

    $v_i \xleftarrow{\;update\;} v_{i-1}$;

6. End For

7. $u \xleftarrow{\;update\;} u_{\text{maxIter}}, v \xleftarrow{\;update\;} v_{\text{maxIter}}$;

8. Return u and v .

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 of the metric embedding, the structure regularization, and the bilinear modeling. Here we summarize these re-modeled LDLORs in Table 1.

Table 1. Overall summary for three re-modeled LDLORs

OR method	Main idea	Convexity	Input pattern	# of variables[*]
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	$n_1 + n_2$

*n_1 and n_2 are respectively the # of rows and columns of an image/matrix.

It can be noticed that from Table 1, both the objectives of IMED-LDLOR and SSSL-LDLOR can be solved by QP optimization, while Bil-LDLOR cannot due to the non-convexity in u and v of its objective function, but can still be solved alternatively with convergence guarantee.

4. Experiments

In this section, we conduct experiments and 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 used 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 defined as $\frac{1}{N}\sum_{i=1}^{N}|l_i^{predicted}-l_i^{groundtruth}|$, denotes the average deviation of the prediction from the ground-truth rank and the lower its value, the better the regression performance; while Acc, defined as $\frac{N_{right}}{N_{total}}$, denotes the classification accuracy of a classifier, and the higher the Acc value, the better the classification performance. *On the other hand*, considering the aforementioned duality of ORs, we further use a recently-proposed *Ordinal Classification Index* (OCI) evaluation criterion $\mathrm{OC}_{\beta}^{\gamma}$ [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*, $\mathrm{OC}_{\beta}^{\gamma}$ can well eliminate the influence of numerical scales used to label ordinal classes on MAE and Acc and thus more suitably measures the deviation of the predicted results from the actual ones. In the following experiment, we set parameter (γ, β) in $\mathrm{OC}_{\beta}^{\gamma}$ to (1, 0.75) as recommended in [51]. As for more details about $\mathrm{OC}_{\beta}^{\gamma}$, due to the complexity of its definition, we omit it here and please refer to the specific definition (7) in [51].

In our experiments, for 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). The tuning ranges of all hyper- or trade-off parameters involved in the experiments are {1e-5, 1e-3, 1e-1, 1e0, 1e1, 1e3, 1e5}. The results of Bil-LDLOR correspond to those after 10 round alternating iterations in each CV.

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 Figure 1.
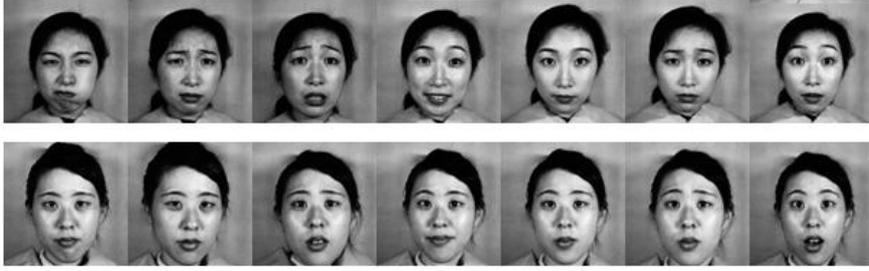
Figure 1. Examples from JAFFE dataset

The experimental results on JAFFE are tabulated in Tables 2a - 2c, 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).*

Table 2a. MAE comparisons among LDLORs on JAFFE (mean ± std-dev)

# NPer[*]	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

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

Table 2b. 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

Table 2c. $\mathrm{OC}_{\beta}^{\gamma}$ 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

From Tables 2a-2c 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-LDLORs 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 2a) 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 SSSL-LDLOR and Bil-LDLOR achieve more significant performance, respectively. For example, in the Acc index in Table 2b, 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 increase of IMED-LDLOR is especially slow and not so distinct, i.e., just about 2 percentages from 0.20 to 0.22, and is 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 sake of OR experiment, we select 6 consecutive ordinal interval angles from profile to frontal views for OR, and each angle 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 shown in Figure 2.
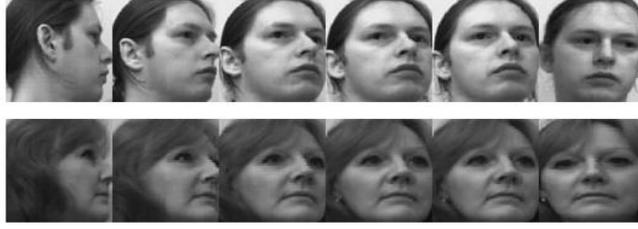
Figure 2. Examples from UMIST dataset

The experimental results on UMIST are respectively listed in Tables 3a - 3c as follows.

Table 3a. 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

Table 3b. 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

Table 3c. $\mathrm{OC}_{\beta}^{\gamma}$ 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

Observing the results from Tables 3a-3c 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 performance superiority is growing more obvious, e.g., in case of NPer = 42 and 43, its Acc performance in Table 3b 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 shows 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 fluctuates, which seems counterintuitive. Such an occurrence, besides the similar reason analyzed in section 4.1, may be further due 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 for our ordinal age group regression, some samples of them are shown in Figure 3.

Figure 3. Examples from FG-NET dataset

The age group regression results on FG-NET are respectively given in Tables 4a - 4c.

Table 4a. 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

Table 4b. 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***

Table 4c. OC_{β}^{γ} 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

From the results on FG-NET of age group regression, we can find 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 4b), the latter stays ahead with about 4 percentage points defeating the former one. Besides, 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 performances of neither the basic LDLOR nor IMED-based one has essentially been increased, e.g., for Acc index, their percentage points are improved respectively merely by 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 for avoiding redundancy.

4.4. Brief summary

Now jointly from all the above experimental results, we can find that for OR on image set, 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 some 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 prior knowledge. *It is worth noting that* though the objective of Bil-LDLOR is not jointly convex but biconvex, thus we can adopt the block coordinate gradient to iteratively optimize its solution, theoretically it has been proven that such iterations are convergent and experimentally it can be observed that about 10 round iterations can lead to a stable solution as illustrated in Figure 4.
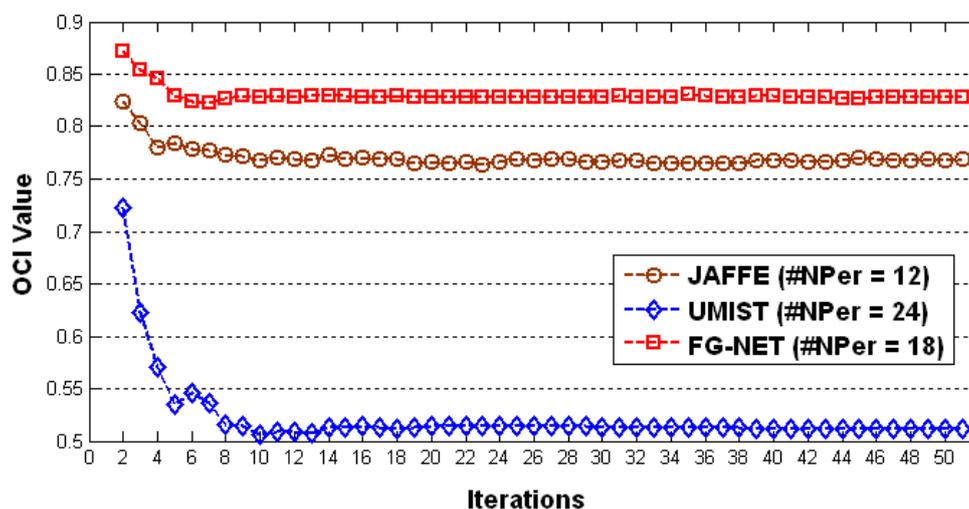
Figure 4. Relationship between convergence under OCI and alternate iterations of Bil-LDLOR

By contrast, though also embedded the spatial information, IMED just reflects the utilization of spatial information in metric rectification rather than the OR objective optimization, thus cannot definitely guarantee desirable results. In addition, *it is worth to note that* from the comparison of Tables 2a vs 2b and 4a vs 4b, 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^γ) as a reasonable evaluation index in ordinal classification or regression task.

5. Conclusions

In this paper, ***first*** through a systematic summary for existing 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 representative illustrations, 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, can as well achieve a distinct gain of 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, thus adopting the OR-specific OCI to more comprehensively evaluate OR is reasonable and recommended. Moreover, by *c*omparing the results among Tables 1c,

2c and 3c w.r.t. the OCI OC_β^γ, we can find that the OR difficulties of human facial expression on JAFFE and age 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.

In short, 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 and structure regularization, while the third category of structure-embedded, however, cannot generate performance benefits as intuitively expected, where the spatial information just is embedded for the metric modification not directly related to the OR objective. From all the above experimental results and corresponding analyses, 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

References

[1] L. Wang, Y. Zhang, J. Feng. On the Euclidean Distance of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(2): 1334-1339, 2005.

[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 in Computer Science, 4984: 973-982, 2008.

[5] Y. Liu, Y. Liu, K. Chan. Tensor Distance Based Multilinear Locality Preserved Maximum Information Embedding. IEEE Transactions on Neural Networks, 21(11): 1848-1854, 2010.

[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 Processing, 90: 2344-2352, 2010.

[8] J. Li, B. Lu. An adaptive image Euclidean distance. Pattern Recognition, 42: 349-357, 2009.

[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. Science China, 53(12): 2415-2428, 2010.

[11] Z. Lei, S. Li. Contextual constraints based linear discriminant analysis. Pattern Recognition Letters, 32: 626-632, 2011.

[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 Recognition, 40: 1533-1543, 2007.

[15] Z. Wang, S. Chen. New least squares support vector machines based on matrix patterns. Neural Processing Letters, 26: 41-56, 2007.

[16] Z. Wang, S. Chen. Matrix-pattern-oriented least squares support vector classifier with AdaBoost. Pattern Recognition Letters, 29(6): 745-753, 2008.

[17] Z. Wang, C. Zhu, D. Gao, et al. Three-fold structured classifier design based on matrix pattern. Pattern Recognition, 46: 1532-1555, 2013.

[18] Z. Wang, S. Chen, J. Liu, et al. Pattern Representation in feature extraction and classifier design: matrix versus vector. IEEE Transactions on Neural Networks, 19(5): 758-769, 2008.

[19] Z. Zhang, T. Chow. Maximum margin multisurface support tensor machines with application to image classification and segmentation. Expert Systems with Applications, 39: 849-860, 2012.

[20] D. Tao, X. Li, X. Wu, et al. General tensor discriminant analysis and gabor features for gait recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(10): 1700-1715, 2007.

[21] D. Tao, M. Song, X. Li, et al. Bayesian tensor approach for 3-D face modeling. IEEE

Transactions on Circuits and Systems for Video Technology, 18(10): 1397-1410, 2008.

[22] D. Tao, X. Li, X. Wu, et al. Tensor rank one discriminant analysis-A convergent method for discriminative multilinear subspace selection. Neurocomputing, 71: 1866-1882, 2008.

[23] J. Wen, X. Gao, Y. yuan, et al. Incremental tensor biased discriminant analysis: A new color-based visual tracking method. Neurocomputing, 73: 827-839, 2010.

[24] B. Wang, X. Gao, D. Tao, et al. A unified tensor level set for image segmentation. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 40(3): 857-867, 2010.

[25] L. Zhang, L. Zhang, D. Tao, et al. Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction. IEEE Transactions on Geoscience and Remote Sensing, 51(1): 242-256, 2013.

[26] K. Lakiotaki, N. Mastsatsinis, A. Tsoukias. Multicriteria user modeling in recommender systems. IEEE Intelligent Systems, 26(2): 64-76, 2011.

[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: 856-867, 2011.

[30] K. Gray, P. Aljabar, R. Heckemann, et al. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. NeuroImage, 65: 167-175, 2013.

[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 Computation, 19(3), 2007.

[36] R. Herbrich, T. Graepel, K. Obermayer. Large margin rank bound arises for ordinal regression. pages 115-132, MIT Press, Cambridge, MA, 2000.

[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. Fundamenta Informaticae, 47: 1-13, 2001.

[39] J. Cheng, Z. Wang, G. Pollastri. A neural network approach to ordinal regression. IEEE International Joint Conference on Neural Networks, 2008.

[40] S. Fouad, P. Tino. Adaptive metric learning vector quantization for ordinal classification. Neural computation, 24: 2825-2851, 2012.

[41] C. Seah, I. Tsang, Y. Ong. Transductive ordinal regression. IEEE Transactions on Neural Networks and Learning Systems, 23(7): 1074-1086, 2012.

[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 Transactions on Knowledge and Data Engineering, 22(6): 906-910, 2010.

[45] R. Duda, P. Hart, D. Stork. Pattern Classification. Wiley, 2000.

[46] X. He, S. Yan, Y. Yu, et al. Face recognition using Laplacian faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(3): 328-340, 2005.

[47] X. He, D. Cai, S. Yan, et al. Neighborhood preserving embedding. In ICCV, 2005.

[48] V. Vapnil. Statistical learning theory. New York: John Wiley & Sons. 1998.

[49] J. Suykens, Vandewalls J. Least squares support vector machine classifiers. Neural Process Letters. 9: 293-300, 1999.

[50] A. Rothman, E. Levina, J. Zhu. Sparse multivariate regression with covariance estimation. Journal of Computational and Graphical Statistics. 19(4): 947-962, 2010.

[51] J. Cardoso, R. Sousa. Measuring the performance of ordinal classification. International Journal of Pattern Recognition and Artificial Intelligence, 25(8): 1173-1195, 2011.

[52] J. Yang, D. Zhang, A. Frangi, et al. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(1), 2004.

Biography of the author(s)

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 Feb 2011 to Feb 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 Jiaotong 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 Sept. 2006 to Oct. 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.

Paper Title:

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

Authors:

Qing Tian, Songcan Chen[*], and Xiaoyang Tan

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics

Highlights:

1) Through summary, find three strategies of using image prior spatial information;

2) Apply these strategies to establish OR variants for classifying ordinal image data;

3) Conduct comprehensive comparisons among the developed novel OR variants;

4) Conclude the effectiveness of spatial structure depends on the embedding way.