CrossMark

Principal component selection using interactive evolutionary computation

Yan Pei[1] (ORCID)

Abstract We propose a method to solve the selection problem of principal components in machine learning algorithms based on orthogonal transformation using interactive evolutionary computation. The orthogonal transformation presents a linear transformation method that preserves the inner product in the two coordinate systems, the one is before the transformation, and the other is after the transformation. One of the addressed subjects for machine learning algorithms based on orthogonal transformation is how to decide the number of principal components, and which of the principal components should be used to reconstruct the original data. In this work, we use the interactive differential evolution algorithm to study these subjects using real humans' subjective evaluation in optimization process. An image compression problem using principal component analysis is introduced to study the proposed method. We do not only solve the selection problem of principal components for machine learning algorithms based on orthogonal transformation using interactive evolutionary computation, but also can analyse the human aesthetical characteristics on visual perception and feature selection arising from the designed method and experimental evaluation. We also discuss and analyse potential research subjects and some open topics, which are invited to further investigate.

Keywords Interactive evolutionary computation · Orthogonal transformation · Aesthetical judgement · Human computer interaction · Visual perception · Principal component analysis · Principal component selection · Interactive differential evolution · Image compression

✉ Yan Pei
  peiyan@u-aizu.ac.jp
  http://web-ext.u-aizu.ac.jp/~peiyan/

[1] Computer Science Division, The University of Aizu, Tsuruga, Ikki-machi, Aizuwakamatsu, Fukushima 965-8580, Japan

1 Introduction

Interactive evolutionary computation (IEC) is an optimization approach that embeds human subjective evaluation into a process of system optimization [23]. It is used to optimize the problems where the explicit fitness functions are difficult or impossible to be established. Using a real human's subjective evaluation, the solution optimized by canonical evolutionary computation (EC) converges to the human's knowledge, experience, and preference. In the process, the selection (evaluation) of IEC algorithm uses the human's knowledge, experience, capability, and preference in its optimization so that the IEC is also referred to as the aesthetic selection [3]. From the optimization system viewpoint, the IEC-based optimization system can be implemented with three parts, i.e. a target system that is optimized, an EC algorithm that can use any type of EC algorithm or swarm optimization algorithm, and a real human as an algorithm evaluator. These three parts compose three research subjects in IEC field, i.e. IEC application research, IEC algorithm research, and human characteristics research of physiology and psychology from IEC evaluation, respectively [14].

Image processing is one of the research fields where IEC can be applied for solving some optimization problems. For example, IEC was used to color separation application in forensic image processing [2], IEC was applied to design image filter using human subjective evaluation [9], etc. On the contrary, image processing methods or algorithms also support a way to research the topics on human aesthetic judgements or selections in IEC. For example, the aesthetic judgement learning was studied using IEC in an evolutionary art system [6]. These two research directions are related to the first and third aspects of IEC research, i.e. IEC application research and human characteristic research, respectively. This paper concentrates on these two subjects as well.

Orthogonal transformation is a linear transformation that preserves lengths of vectors, and angles between vectors in inner product space. For example, $a \in V$ and $b \in V$ are two vectors, and V refers to an inner product space, if there is a linear transformation ($T : V \rightarrow V$) that can preserve the geometrical properties of vectors (a and b) after the transformation, i.e. $\langle a, b \rangle = \langle Ta, Tb \rangle$ ($\langle a, b \rangle$ means the inner product of vector a and vector b), the linear transformation ($T : V \rightarrow V$) is an orthogonal transformation. There are several algebraic and geometrical properties that are preserved before and after the transformation, many machine learning algorithms, therefore, were developed using orthogonal transformation technique.

Principal component analysis (PCA) is one of the machine learning algorithms that uses the orthogonal transformation technique [11]. The objective of PCA pursues to establish a linear transformation by which the total variance of the original data is projected to be maximum. After the transformation, the original data can be expressed by the linear combination of projected lower dimensional data. It is one of the dimensionality reduction methods for presenting the primary aspects of the data. One of the study subjects in PCA is the selection problem of principal components after the orthogonal transformation. It addresses that (1) how many principal components are enough to present the original data; and (2) which principal components are useful to reconstruct these original data. We attempt to discuss these subjects using the IEC. This presents the originality of this work.

This paper studies the selection problem of principal components for machine learning algorithms based on orthogonal transformation with the perspective of IEC. We design an image compression application using the PCA, and solve the selection problem of principal components using interactive differential evolution (IDE). The paired comparison mechanism of IDE is applied to reduce the user fatigue in the process of human subjective evaluation. We also discuss the topics that use the human's subjective evaluation for feature selection, and attempt to analyse the human characteristics by designed method and experiment. On the one hand, we design a method that solves the selection problem of principal components for machine learning algorithms based on orthogonal transformation. On the other hand, we can also use the designed method to analyse the human's characteristics on feature selection. There are two innovative contributions of this work.

Following this introductory section, we briefly make an overview of the differential evolution (DE), the interactive differential evolution (IDE) and its paired comparison mechanism, the orthogonal transformation, and the principal component analysis in Sect. 2. In Sect. 3, we present principal component selection problem, and use the IDE to solve an image compression problem. We explain how to select the principal components to restore an image, and use some quantitative metrics to evaluate the compressed images with different restore methods. The evaluation and discussion of the proposed method are presented in Sect. 4. We also report that how to analyse the characteristics of human using the proposed method, and address the primary discoveries from the human's subjective evaluation. Some human visual perception conclusions, and aesthetical judgements and characteristics are discovered by analysing the evaluation results. Finally, we conclude the whole work, and discuss the future works, open topics in Sect. 5.

2 An overview of related works and techniques

2.1 Differential evolution

Differential evolution (DE) is one of the population-based optimization algorithms [19, 22]. It searches the global optimum using a differential vector between two individuals, whose length is in proportion to the distribution size of individuals in general. Each parent individual generates its offspring. As a parent's individual is replaced with the generated one only when the fitness of a generated one exceeds that of the parent, we may say that DE operations have a similarity to an elite strategy or hill-climbing method. The most perspective feature of DE is powerful search capability with the quite simple algorithm implementation.

Suppose that an array on the left side of Fig. 1 means individuals, contour lines at the right side are a fitness landscape, and circles on the landscape are the individuals. DE algorithm for one search generation is described in the below and repeats until a satisfied solution(s) is(are) found or the search reaches the maximum generation.

(1) Select one individual as a target vector.
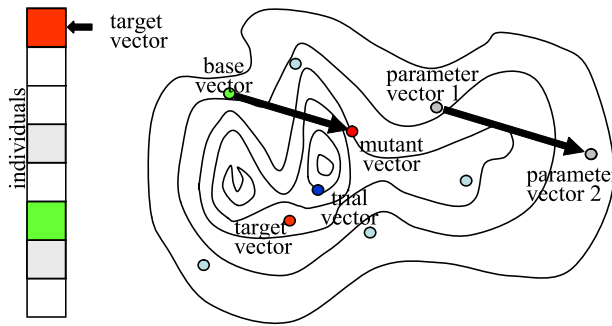(2) Select two other individuals as parameter vectors randomly and make a differential vector from them.

Fig. 1 Differential evolution algorithm (DE), when comparing target vector and trail vector, it is a paired comparison mechanism in DE algorithm nature, which benefits to human evaluator of interactive DE (IDE)

(3) Select the best individual from the remained individuals or one other individual randomly as a base vector.
(4) Conduct a mutant vector by adding a weighted differential vector to the base vector.
(5) Produce a trial vector by crossing the target vector and the mutant vector.
(6) Compare the target vector and the trial vector, and select a better one as offspring in the next generation.
(7) Go to the (1) and generate other offspring until all individuals are processed and compared, and then go to the next generation's processing.

The terms of vector and individual mean the same search points. The above steps (1)–(4) are summarized as in Eq. (1), which shows that the DE algorithm is easily implemented, where F is called as a scale factor. There are several DE variations in the number of differential vectors, selection methods of a base vector in the above step (3), crossover methods in the above step (5), and others.

$$\text{mutant vector} = \text{base vector} + F \times (\text{parameter vector } 2 - \text{parameter vector } 1). \quad (1)$$

Differential evolution is an algorithm that can control the balance of exploration and exploitation automatically thanks to a differential vector which average length is in proportion to the distribution size of individuals. We can say that DE searches around a base vector by narrowing its search area gradually because differential vectors have different lengths and different directions, and are added to the base vector to find search points [in the above step (4)]. Explanation in more detail, DE biases the search area to each target vector side [in the above steps (5) and (6)].

2.2 Paired comparison-based interactive differential evolution

The interface design of IEC algorithm is one of the study subjects in IEC algorithm research. Most IEC algorithms display all individuals to an IEC evaluator and request him/her to input a fitness evaluation to each of them as it is represented by interactive genetic algorithm (IGA). When individuals are images, for example, it is easy for an

IEC evaluator to compare them spatially and evaluate them. It is the reason why most IEC algorithms take this display-evaluation method. However, when individuals are sounds or movies, an IEC evaluator has to compare an individual with the others in memory and his/her mental load and fatigue become heavy. It was pointed out that human has a memory limitation and cannot process more than five to nine different information simultaneously [8]. Population sizes of many IEC algorithms frequently exceed this memory limitation, so displaying 10–20 sounds or movies to an IEC evaluator is not practical.

Paired comparison-based IEC solves this problem by replacing comparison of all individuals with paired comparisons and is expected to reduce IEC evaluator fatigue. The first approach to implement the paired comparison is a tournament IGA [7]. $N-1$ paired comparisons are iterated for N individuals in every generation, and fitness values are calculated using the number of winnings and fitness difference between each pair. The disadvantage of the tournament genetic algorithm (GA) is that the obtained fitness includes noise among individuals that are not comparable because the tournament is not a round robin competition against the canonical GA. The noise influences a GA selection operation and results in worse GA optimization performance.

Interactive DE (IDE) adopts the DE algorithm by replacing the fitness comparison in step (6) with real human subjective evaluation. The logic of step (6) is referred as a paired comparison mechanism. On the other hand, paired comparison-based IDE does not revise any parts of its algorithm because the algorithm includes paired comparison nature in the above algorithm step (6) [24]. Since it displays paired comparisons of individuals to an IDE evaluator without modifying the implementation of DE algorithm, the IDE is expected to be a promising IEC method. Some other comparison-based IDE algorithms were, therefore, developed by considering multiple comparison mechanism in canonical IDE process, such as triple and quadruple comparison-based IDE [17], triple comparison-based IDE based on memetic search [18], chaotic evolution [12,13], and paired comparison-based interactive chaotic evolution [15].

2.3 Orthogonal transformation

Orthogonal transformation is a linear transformation method, and it is a transformation in the inner product space. It can preserve geometrical and algebraic properties of data before and after the transformation. The geometrical size and shape of transferred data are as the same as that of the data before transformation. In the another word, the inner products of the transferred data and the original data are preserved, so the relation of data does not be destroyed by the transformation. Orthogonal transformation constructs a new coordinate system to present the data with some certain criteria.

The principal component analysis (PCA) [11] and linear discriminant analysis (LDA) [4] are two of the machine learning algorithms that use the orthogonal transformation to implement. The PCA pursues a reconstructed coordinate system where the total variance of original data achieves being maximum, and the LDA pursues that where not only the total distance sum of center points of all the group is maximum, but also the total variance sum of data in each group is minimum. We can design a variety of machine learning algorithms using orthogonal transformation with different

criteria. The principal component discriminant analysis is one of the machine learning algorithms that composes PCA and LDA into a uniform framework [16].

2.4 Principal component analysis

The PCA is one of the machine learning algorithms that uses the principle of orthogonal transformation. It is a statistical procedure, which transfers data into a coordinate system defined by linearly uncorrelated variables using orthogonal transformation technique [11]. In different fields, it has different name, such as the discrete Karhunen Loève transform in signal processing, and the Hotelling transform in multivariate quality control. The PCA yields a new coordinate system where the projected data have the first maximum variance in the first axis of the coordinate system, i.e. the first principal component, and have the second maximum variance in the second axis of the coordinate system, i.e. the second principal component, and so on. It is one of the image processing methods that can be applied to image compression and restore.

We firstly note some data symbols that are used in our derivations of the PCA. There is a set of data, which is presented as $\{x_1, x_2, \ldots, x_n\}$, and n is the number of the data. We also remark it as $X^T = [x_1, x_2, \ldots, x_n]$. The observed data have d-dimension. It can be expressed as in Eq. (2). We suppose that all the data are with zero mean. If the data are not with zero mean, we make a pre-processing that every datum minuses their mean value (Eqs. (3), (4)).

$$X^T = \begin{bmatrix} x_{11} & x_{21} & \ldots & x_{n1} \\ x_{12} & x_{22} & \ldots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1d} & x_{2d} & \ldots & x_{nd} \end{bmatrix}. \tag{2}$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{3}$$

$$x_i' = x_i - \bar{x}. \tag{4}$$

The total variance of the data that is projected to a direction v (new constructed coordinate system) can be expressed by Eqs. (5) and (6), where C is the co-variance matrix of the data. The objective of the data projection is to find a new coordinate system where the total variance has the maximum value. The aim of this optimization problem is to find a solution of Eq. (7). We can solve this optimization problem using Lagrangian multiplier method or matrix calculus method.

$$\sigma^2 = v^T C v. \tag{5}$$

$$C = \frac{1}{n} \sum_{i=1}^{n} x_i x_i{}^T = \frac{1}{n} X^T X. \tag{6}$$

$$v = \arg \max_{v} v^T C v. \tag{7}$$

3 Principal component selection using interactive evolutionary computation: a study of optimization of machine learning algorithm with human subjective evaluation

3.1 Principal component selection problems

The objective of PCA transfers the original data X into another presented data, e.g. Y. Eq. (8) shows this transformation, where $V^T = [v_1, v_2, \ldots, v_d]$ that can be obtained from the eigenvalue problem of Eq. (7). From Eq. (8), we can obtain $Y = V'X$ ($Y^T = [y_1, y_2, \ldots, y_n]$), i.e. $X = V'^{-1}Y$, so the final transferred data is in Eq. (10). If V' is an orthogonal matrix, the Eq. (10) can be rewritten as $X' = V'^T V'X$ (due to the property of orthogonal matrix, $V'^T = V'^{-1}$), and $V'^T = [v_1, v_2, \ldots, v_m]$ ($m \in Z^+, m <= d$), where $v_1, v_2, \ldots, v_m$ are d dimensional vectors, i.e. $v_i^T = [v_{i1}, v_{i2}, \ldots, v_{id}]$, ($i = 1, 2, \ldots, m$), such as in Eq. (9). Here, we use m principal components to reconstruct the original data.

$$Y = V'X = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_m^T \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1d} & x_{2d} & \cdots & x_{nd} \end{bmatrix}. \tag{8}$$

$$V' = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_m^T \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1d} \\ v_{12} & v_{22} & \cdots & v_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1d} & v_{2d} & \cdots & v_{md} \end{bmatrix}. \tag{9}$$

$$X' = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_m^T \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1d} & x_{2d} & \cdots & x_{nd} \end{bmatrix} = V'^T V'X. \tag{10}$$

The canonical selection method of principal component considers the eigenvalue to solve this problem by a threshold value. There are λ_i, ($i = 1, 2, \ldots, d$) from the eigenvalue solutions of Eq. (7). If the threshold of Eq. (11) achieves to a preset value, e.g. 80–90 %, the first m principal components are selected according to the ranking of their eigenvalues. This selection method considers the importance of principal component accordance to its eigenvalue. Whether is it also important to human visual perception, it is a research subject of this work.

$$\frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{d} \lambda_i} > \text{threshold}. \tag{11}$$

This work handles the selection problems of principal component that yield:

1. how many principal components are enough to reconstruct the original data from the human's visual perception?

2. which of the principal components are useful to reconstruct the original data?
3. which of the principal components are sensitive to the human's visual perception?

These are the selection problems of principal components of machine learning algorithms based on orthogonal transformation. We use the interactive evolutionary computation algorithm as a tool to discuss and analyse these subjects.

3.2 Image compression using principal component analysis

Digital images can be compressed using the PCA. In Eq. (10), before projecting the original X back to original coordinate system using $Y = V^T X$, we can truncate the matrix V^T to keep the certain number of principal components, i.e. $V'^T = [v_1, v_2, \ldots, v_m]$ ($m \in Z^+, m <= d$). The PCA is a dimensionality reduction method, by which we recover the matrix X, and transfer it into X' using the transformation as in Eq. (10). The two matrices (X and X') have the same dimensions, but the contents of them may not be totally the same since we truncate the matrix V^T. So the matrix X' can present the primary information of matrix X in the sense of the number of m and selection method of these principal components V'^T.

Let us study the left side of Eq. (10), it can be separated into two parts. One is the V'^T, and the other is the $V'X$. The size of these two matrices is $d \times m$ and $m \times n$, respectively, and the size of the original data, X, is $d \times n$. If the $(d \times m) + (m \times n) < (d \times n)$, i.e. in Eq. (12), it means that the matrix X can be compressed using matrix X' that can be presented by the multiplication of two matrices, i.e. V'^T and $V'X$ of Eq. (10).

$$m < \frac{d \times n}{d + n}.$$
(12)

If the X is an image, we can use the principle of Eq. (10) to compress it using the PCA. The image compression method using the PCA is a lossless compression since it reduces not only the redundancy information, but also the irrelevancy information. Here, we present three test images that are used in the evaluation parts in Fig. 2, they are all greyscale format images, and the size of all the images is 256×256.
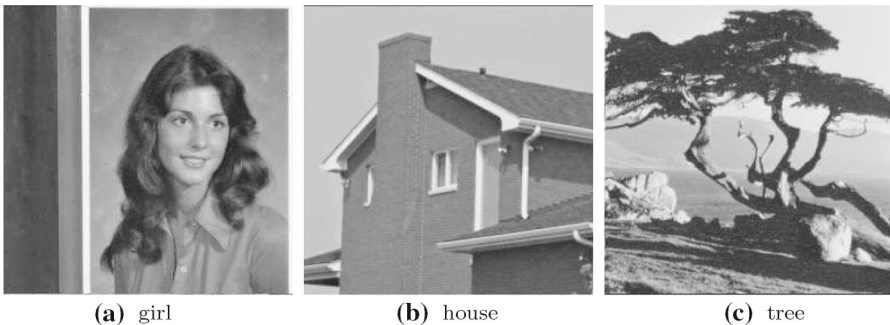

(a) girl   **(b)** house   **(c)** tree

Fig. 2 Three test images, **a** girl, **b** house, and **c** tree, they are all greyscale format images, and the size of all the images is 256×256

3.3 Selection of principal components using interactive differential evolution

We code the 256×256 image into a data set with 256 data by column and with 256 dimensions by rows. Using the PCA, we can establish a 256×256 co-variance matrix of these data. After we solve the eigenvalue problem of this co-variance matrix, we can obtain 256 eigenvalues and their corresponding eigenvectors. From the Eq. (12), the $d = 256$, $n = 256$ in the test images, so if the $m < 128$, the original test images can be compressed. However, we should make a pre-processing that makes a mean matrix with a size of 256×1, so the value of m should be less than or equal to 127. Because the principal components, whose corresponding eigenvalues are small, are less useful, and the maximum value of Matlab has limitation, we only investigate the first 64 principal components and their combinations for these three test image compression problems.

We use paired comparison-based IDE as an optimizer to find the best combination solution of principal components of three compressed test images with human subjective evaluation. The advantage of the paired comparison-based IDE is that it can display only two objects (two compressed image in this paper) for human's judgement rather than other IEC algorithms force human to give a rank for each individual simultaneously, such as IGA. As the feature of IEC algorithm setting, it needs less population size and less generation number to relieve evaluator's fatigue. We only use six individuals and five generations in IDE algorithms. The parameters of IDE are listed as in Table 1. We invited six subjects to involve in the evaluation, all of them are university students, male, and age range is about (20,25).

4 Evaluations and discussions

4.1 Discussion on image compression results

This work uses an image compression application to discuss the subject of principal component selection of machine learning algorithm based on orthogonal transformation. We need to firstly define some evaluation metrics to quantitatively discuss the image compression results. In image processing field, there are several evaluation metrics for quantitatively evaluating image, such as mean square error (MSE) [5], peak signal-to-noise ratio (PSNR), compression ratio (CR), structural similarity (SSIM) [25], and information entropy (IE) [21]. We use these evaluation metrics to

Table 1 IDE experiment parameters setting		
	Population size	6
	Search range of parameters	$[1, 2^{64}]$
	Scale factor F	1.0
	Crossover rate	1.0
	DE operation	DE/rand/1/bin
	Max. search generation, MAX_{NFC}	5

discuss and analyse our results. The definitions of MSE, PSNR, CR, SSIM, and IE are listed in Eqs. (13), (14), (15), (16), and (17), respectively. The x_i'/x' and x_i/x are compressed data and original data in Eqs. (13) and (15). The MAX_x is the maximum value in the data of x in Eq. (14). The μ and σ are mean and variance values in Eq. (16), where $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are two variables to stabilize the division with weak denominator; L is the dynamic range of the pixel-values (typically, it is $2^{\#\text{bits per pixel}} - 1$); and $k_1 = 0.01$ and $k_2 = 0.03$ by default. The $p(x)$ is the probability mass function of x in Eq. (17).

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (x_i' - x_i)^2. \tag{13}$$

$$\mathrm{PSNR} = 20 \log_{10} \frac{\mathrm{MAX}_x}{\mathrm{MSE}}. \tag{14}$$

$$\mathrm{CR} = \frac{x}{x'}. \tag{15}$$

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \tag{16}$$

$$\mathrm{IE} = -p(x) \log_2 p(x). \tag{17}$$

Table 2 shows MSE, PSNR, and SSIM values of each image with different principal component settings. Table 3 presents the information entropy of compressed data (image) by the PCA and the IDE with different principal components. The legend, PCA, in tables means that the images are compressed by the PCA with the same number of principal components used by IDE, but it uses the method of Eq. (11) to select the principal components. The legend, IDE, means that the images are compressed using the principal components obtained by IDE (both the number and combination of principal components). So the two methods have the same number of principal components, but different combination of that. In Tables 2 and 3, there are categories level = 0, level = 0.5, and level = 1, meaning that the images compressed obtained by which we use selected more than the 100, 50, and 1 percent of all the selected principal components, respectively. We analyse and discuss our proposal with these results.

From the MSE and PSNR evaluations in Table 2, it shows that the difference in these two metrics becomes larger when the applying selected principal component number is increasing. When we use all of the selected principal components from the human's selection, i.e. the condition of Level = 0, the MSE and PSNR values from the methods of IDE and PCA do not have many differences, because the PCA was proved as the best transform coding method for image compression when the data are under a normal distribution condition [10]. When we use the selection method of Eq. (11) to select the principal component to restore the images, it therefore presents the best effectiveness comparing with the original images. The PCA cannot automatically decide the number of principal components. However, our proposal can support a solution for finding a better selection method to decide the number and the combination of principal components from the human visual perception. This presents the originality of this work.

Table 2 The MSE, PSNR and SSIM results, IDE means the images are restored using the principal components selected by IDE, and PCA means images are restored using the Eq. (11) with the same principal component number as IDE, but different principal components' combination

	Image (a)		Image (b)		Image (c)	
	IDE	PCA	IDE	PCA	IDE	PCA
MSE						
Level = 0	1311.415	1145.929	4785.905	4531.623	5027.810	3826.290
Level = 0.5	1521.185	1200.512	5597.984	4794.802	5418.308	4115.758
Level = 1	1521.185	1200.512	6194.937	5459.347	5934.638	4943.077
PSNR						
Level = 0	16.953	17.539	11.331	11.568	11.117	12.303
Level = 0.5	16.309	17.337	10.650	11.323	10.792	11.986
Level = 1	16.309	17.337	10.210	10.759	10.397	11.191
SSMI						
Level = 0	0.629	0.654	0.644	0.682	0.665	0.705
Level = 0.5	0.576	0.635	0.562	0.640	0.617	0.676
Level = 1	0.576	0.635	0.524	0.551	0.551	0.604

Table 3 The information entropy (IE) results; it indicates that human can obtain more information from his or her visual perception

	Image1	IDE	PCA	Image2	IDE	PCA	Image3	IDE	PCA
Level = 0	3.243	3.436	3.434	2.901	3.193	3.148	3.141	3.273	3.199
Level = 0.5	3.243	3.477	3.432	2.901	3.286	3.143	3.141	3.338	3.228
Level = 1	3.243	3.477	3.432	2.901	3.227	3.226	3.141	3.445	3.327

The IDE and PCA have the same meaning as in Table 2

The SSIM is designed to improve conventional image evaluation methods, such as the MSE and the PSNR, which have been proven to be not inconsistent with human visual perception. From our evaluation result of SSIM in Table 2, it presents that SSIM has almost the same tendency with the PSNR. We cannot find any benefits from our results due to the test case by comparing the IDE and the PCA. However, the differences between these two methods are slight. We can conclude that the principal component selection methods of IDE and Eq. (11) under the condition with the same number of principal components are not much different. However, the method of Eq. (11) cannot decide a way to select the number of principal components; our proposal supports a method to do.

For the information entropy (IE) evaluation in Table 3, we can conclude that the information entropy obtained with the principal component selection by Eq. (11) is less than that obtained by IDE when the number of principal components is the same in both methods. On the one hand, human's subjective selection has noise, which is shown in the results of MSE and PSNR; it makes the information entropy become increasing. On the other hand, human's subjective selection can obtain more detail
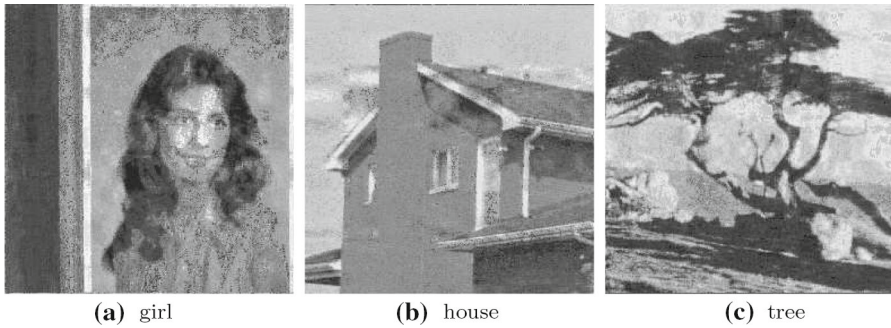
(a) girl   **(b)** house   **(c)** tree

Fig. 3 Three compressed test images, they are **a** girl, **b** house, and **c** tree, the compression ratios are 2.81, 2.53, and 2.69, respectively

information from the human visual perception. So the information entropy obtained by IDE method is more than that of PCA with selection method of Eq. (11). This is one of the advantages of our proposal, i.e. principal component selection by IDE can obtain more information of compressed images from the viewpoint of human visual perception.

Figure 3 shows a result of compressed image by considering all the subjects' selection. These compressed images do not have all the first m principal components, but they can also present the primary aspects of the original images by comparing with the original images in Fig. 2. There are 45, 50 and 47 principal components in compressed images (a), (b), and (c) of Fig. 3, respectively. Their corresponding compression ratios are 2.81, 2.53, and 2.69 for images (a), (b), and (c), respectively, using the Eqs. (15) and (18), where PCs means the number of selected principal components.

$$\text{compression} - \text{ratio} = \frac{256}{2 \times \text{PCs} + 1}. \tag{18}$$

### 4.2 Discussion on characteristics of selected principal components

The objective of PCA is to find a new coordinate system where the total variance of projected data can have the maximum value using orthogonal transformation technique. In image compression application, the first m principal components according to larger eigenvalues present the important aspects of the original data. As it is known, the coordinate (axis) with the maximal variance presents primary content of the data; on the contrary, that with the minimal variance presents the noise. The principal components with the larger eigenvalues, i.e. the first m principal components (with relative larger variance) can restore the image more clearly and more exactly. We, therefore, conclude that human's perception also follows this rule, i.e. human visual perception is sensitive to images restored using the first m principal components.

Figure 4 displays a result of the percentage of selected principal components in the final generation (the fifth generation) for each image, i.e. images (a), (b), and (c). In this figure, the sequence of the principal components is ranked with the their corresponding eigenvalue, i.e. the first principal component is with the largest eigenvalue, the second
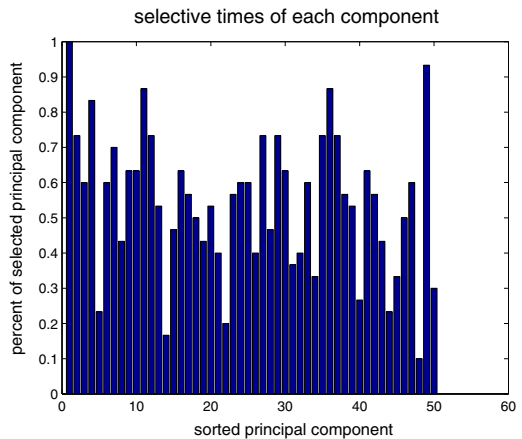
Fig. 4 The selected principal component percentage in the last generation (the fifth generation); the results indicate that the sensitiveness of human visual perception is not in accordance with the principal components ranked by their corresponding eigenvalues
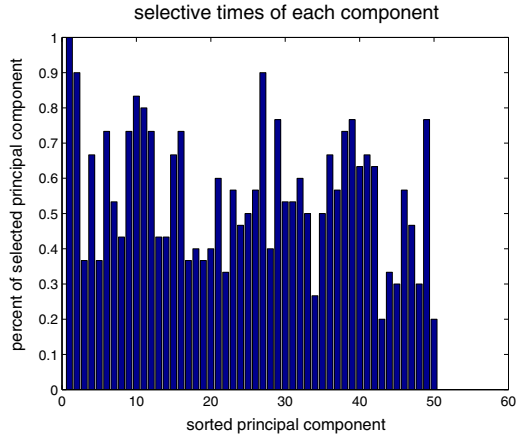
(1). image **(a)**
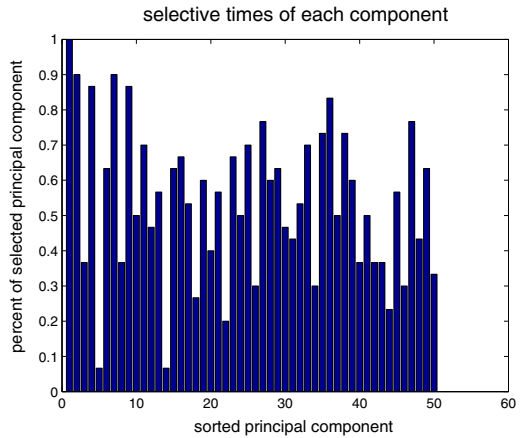
(2). image **(b)**

(3). image **(c)**

one with the second largest eigenvalue, etc., from the left to the right in X axis. We can observe that the percentage of the selected principal components is not decreased in accordance with the ranked eigenvalues for each principal component from left to right. It indicates that the visual sensitiveness of human's perception does not follow the derivative conclusion.

Human can percept more detail visual information that is presented by the principal components with the smaller eigenvalues. The principal component with the smaller eigenvalue can support more sensitive information in a equally important manner as can those with a larger eigenvalue for human visual perception. There is no difference among the principal components with smaller eigenvalues or larger eigenvalues from the human visual perception. This is the primary discovery of this work, and we can also use this conclusion to further develop our proposal for finding sensitive principal component from the data (images) for human in the viewpoint of human visual perception.

4.3 Discussion on subjects' aesthetical characteristics

One objective of this work is to investigate the relation between principal component selection and human's perception. Figure 5 shows the average number of selected principal components of subjects 1–6 in each generation. The average number of selected principal components does not increase with evolution optimization from one generation to the next. It indicates that it is not true that more principal component number means more sensitive information to human visual perception in the image compression application. However, for the human face perception, i.e. test image (a). girl, it may need more number of principal components to present a clear face because the results of subjects 1, 3 and 6 in Fig. 5 support this conclusion.

Another two results can be obtained from observing Fig. 5. The one is that the number of selected principal components depends on the different image compression tasks. The other is that different subjects conduct different selection decision makings that present the personal aesthetical selection, judgement, and characteristics. The human characteristic clustering analysis can be conducted in accordance with this discovery, i.e. subjects 1, 3, and 6 may have the same aesthetical thinking way on the test image (a), because they have the same tendency of selecting principal component.

We also use self-organized map (SOM) as a clustering tool to analyse subjects' aesthetical characteristics. The self-organized map is designed as 5×5 meshes. From the clustering results (Fig. 6), we can observe that subjects' aesthetical selections are different due to the application task, and we can obtain such knowledge from the selected principal components using our proposed method, i.e. selecting principal components using IEC. Especially, the clustering result of image (a), i.e. the subjects 1, 3, and 6 are clustered in the same class. It also verifies the conclusion from the observation of Fig. 5. This presents one of the originalities in this work.

What is the sensitive information from the human visual perception? We can obtain some qualitative results from our designed evaluation. We reconstruct the images using all the selected principal components 100 percent, 50 percent of principal components according to their selected rate, and 10 percent according to their selected rate by
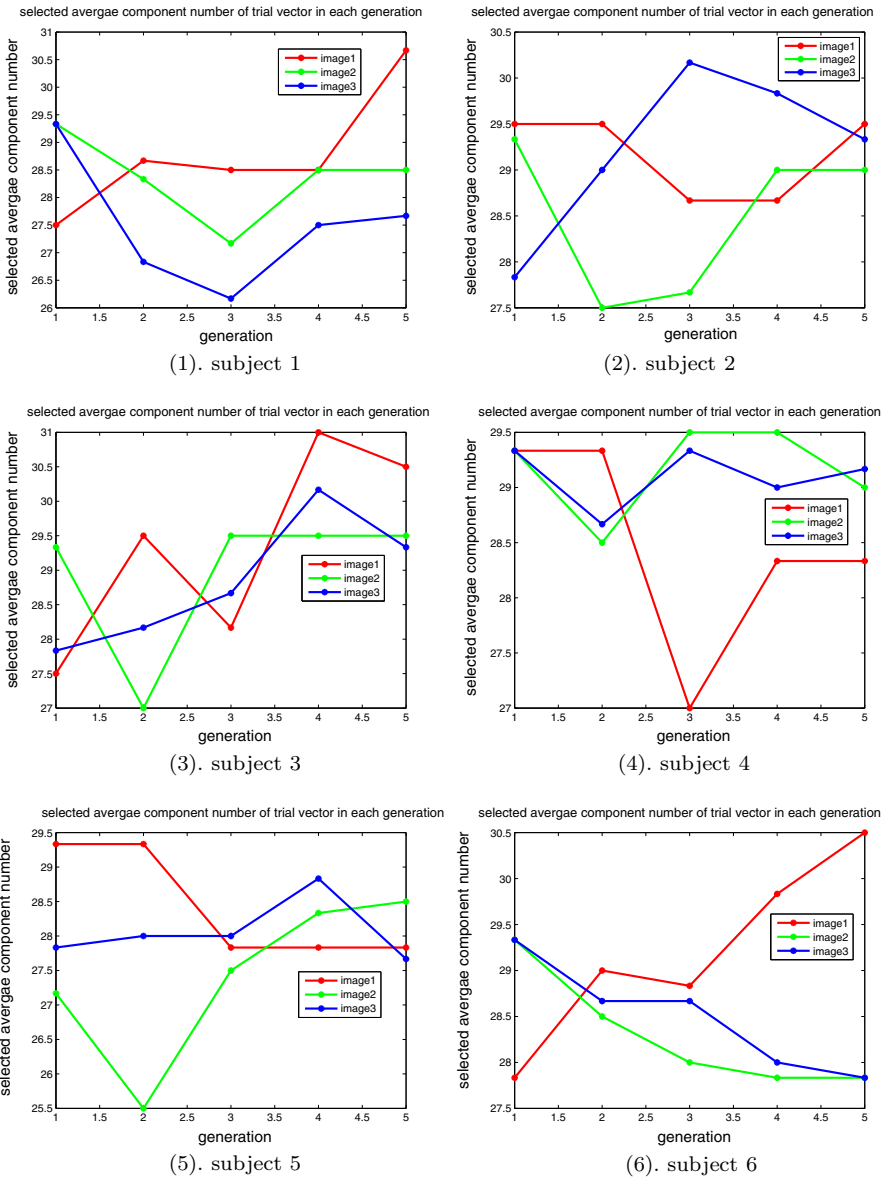
Fig. 5 Average number of selected principal component in each generation, by analysing the evaluation data in the optimization process; we can obtain the knowledge on subjects' aesthetical judgement. For example, subjects 1, 3, and 6 may have the same aesthetical thinking way on the test image (**a**)

the evaluators (see Fig. 7). Arising from observing and comparing these three group images, we can find that the shape and frame information is sensitive to the human visual perception, rather than the concrete information that can express the contents of the images in detail. For example, the image (a) presents a face of a girl, the
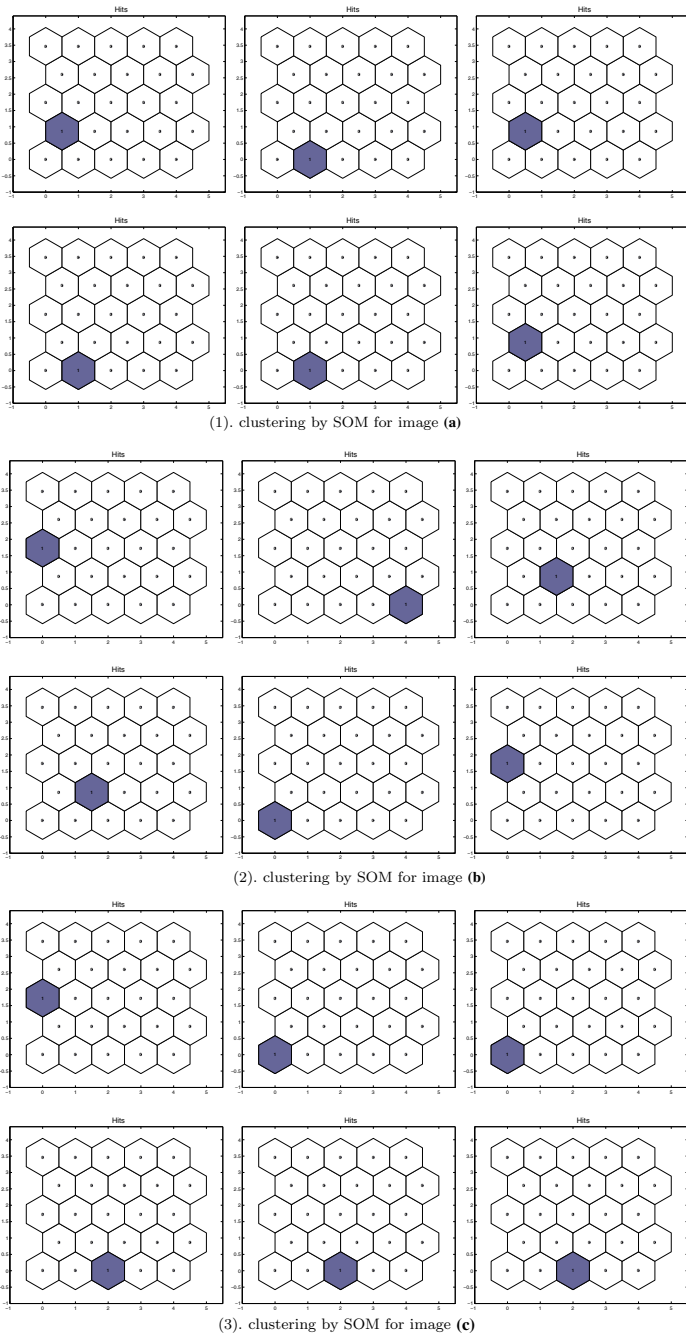
(1). clustering by SOM for image (**a**)

(2). clustering by SOM for image (**b**)

(3). clustering by SOM for image (**c**)

Fig. 6 Clustering analysis using self-organized map (SOM), for image (**a**), there are two classes: (subjects 1, 3, 6), and (subjects 2, 4, 5); for image (**b**), there are four classes: (subjects 1, 6), (subject 2), (subjects 3, 4), and (subject 5); for image (**c**), there are three classes: (subject 1), (subjects 2, 3), (subjects 4, 5,6). We can use this method to analyse the human aesthetical judgement and characteristic

(1). compressed images with 100% selected principal components
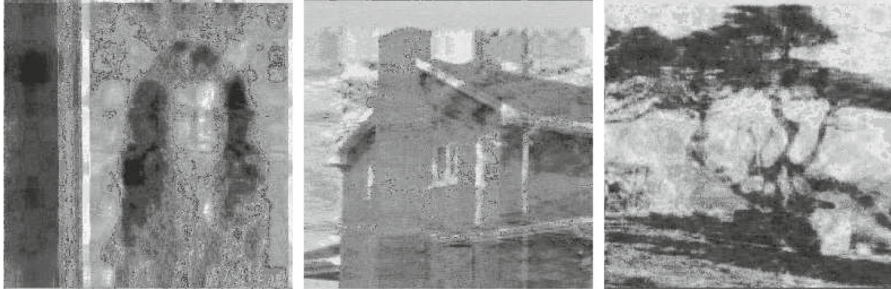
(2). compressed images with 50% selected principal components
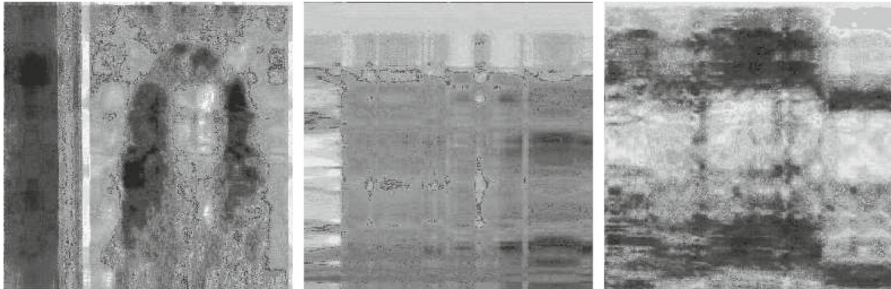
(3). compressed images with 10% selected principal components

Fig. 7 We use 100, 50, and 10 % selected principal components to reconstruct the images. We can obtain that human visual perception is sensitive to the shape and frame information from these images' comparisons

concrete five senses information is not presented in the image that is reconstructed by 10 percent of all selected principal components, but the face frame information is displayed evidently. This is one of the discoveries from our evaluations. The proposed method and design can be used to analyse the aesthetical characteristics and human perception by designing a variety of principal component selection applications.

5 Conclusion and future works

In this paper, we proposed a method to select principal components of machine learning algorithm based on orthogonal transformation using IEC. Conventional selection method cannot support an efficient way to decide the number of principal components

in related applications. Our proposal can efficiently solve this problem using human subjective evaluation. The proposal does not only support a way to decide the number of principal components, but also decide which principal components are sensitive to the human perception, because the optimization process involves human's aesthetical judgement. We used image compression applications to investigate our proposed method and discussed related research subjects. Another advantage of this method is that it also supports a way to investigate a certain human's aesthetical thoughts by analysing the evaluation data.

In the future, we will extend this work in other application fields to investigate the performance of proposed method, such as audio and video applications. The performance issues of optimization algorithm are also a primary subject for the future study. The paired comparison-based IDE can relieve the evaluators' fatigue theoretically; we need to use other IEC algorithms to investigate this issue, and study our proposed method using the other IEC algorithms for a comparison. In this work, we only use the PCA as an example to study our proposal, other machine learning algorithms and applications based on orthogonal transformation, such as linear discriminant analysis, kernel-based PCA [20], general discriminant analysis [1], and linear principal component discriminant analysis [16] also need to solve the selection problem of principal components. We will study principal component selection of other machine learning algorithms in the future. These research subjects are invited to further investigate in our future work.

References

1. Baudat G, Anouar F (2000) Generalized discriminant analysis using a kernel approach. Neural Comput 12(10):2385–2404
2. Berger CEH, Koeijer JA, Glas W, Madhuizen HT (2006) Color separation in forensic image processing. J Forensic Sci 51(1):100–102
3. Dawkins R (1986) Blind watchmaker. Alfred A Knopf, New York
4. Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugen 7(2):179–188
5. Lehmann EL, Casella G (1998) Theory of point estimation, vol 31. Springer Science & Business Media, Berlin
6. Li Y, Hu C, Minku LL, Zuo H (2013) Learning aesthetic judgements in evolutionary art systems. Genet Program Evolvable Mach 14(3):315–337
7. Lim IS, Thalmann D (2000) Tournament selection for browsing temporal signals. In: Proceedings of the 2000 ACM Symposium on Applied Computing, vol 2. ACM, pp 570–573
8. Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63(2):81–97
9. Ono S, Maeda H, Sakimoto K, Nakayama S (2014) User-system cooperative evolutionary computation for both quantitative and qualitative objective optimization in image processing filter design. Appl Soft Comput 15:203–218
10. Papoulis A, Pillai SU (2002) Probability, random variables, and stochastic processes. Tata McGraw-Hill Education, New York
11. Pearson K (1901) Liii. on lines and planes of closest fit to systems of points in space. Lond Edinb Dublin Philos Mag J Science 2(11):559–572
12. Pei Y (2013) A chaotic ergodicity based evolutionary computation algorithm. In: 2013 Ninth International Conference on Natural Computation (ICNC2013). 23–25 Jul 2013 IEEE, pp 454–459
13. Pei Y (2014) Chaotic evolution: fusion of chaotic ergodicity and evolutionary iteration for optimization. Nat Comput 13(1):79–96
14. Pei Y (2014) Study on efficient search in evolutionary computation. PhD thesis. Kyushu University

15. Pei Y (2015) From determinism and probability to chaos: chaotic evolution towards philosophy and methodology of chaotic optimization. Sci World J 2015:14. doi:10.1155/2015/704587
16. Pei Y (2015) Linear principal component discriminant analysis. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC2015), IEEE, Oct, 9–12 2015, pp 2108–2113
17. Pei Y, Takagi H (2013) Triple and quadruple comparison-based interactive differential evolution and differential evolution. In: Foundations of Genetic Algorithms Workshop XII (FOGA2013), ACM, Jan, 16–20 2013, pp 173–182
18. Pei Y, Takagi H, (2015) Local information of fitness landscape obtained by paired comparison-based memetic search for interactive differential evolution. In: 2015 IEEE Congress on Evolutionary Computation (IEEE CEC2015), IEEE, May, 25–28 2015, pp 2215-2221
19. Price K, Storn RM, Lampinen JA (2006) Differential evolution: a practical approach to global optimization. Springer Science & Business Media, Berlin
20. Schölkopf B, Smola A, Müller K-R (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput 10(5):1299–1319
21. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(3):379–423
22. Storn R, Price K (1997) Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. J Glob Optim 11(4):341–359
23. Takagi H (2001) Interactive evolutionary computation: fusion of the capabilities of ec optimization and human evaluation. Proc IEEE 89(9):1275–1296
24. Takagi H, Pallez D (2009) Paired comparison-based interactive differential evolution. In: World Congress on Nature & Biologically Inspired Computing 2009 (NaBIC 2009). 9–11 Dec 2009 IEEE, pp 475–480
25. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612