

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
28 July 2005 (28.07.2005)

PCT

(10) International Publication Number
WO 2005/067572 A2

(51) International Patent Classification: **Not classified**

(21) International Application Number:
PCT/US2005/001671

(22) International Filing Date: 13 January 2005 (13.01.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/536,210 13 January 2004 (13.01.2004) US

(71) Applicant (for all designated States except US): **NEW YORK UNIVERSITY** [US/US]; 70 Washington Square South, New York, NY 10012-1091 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **VASILESCU, Manuela, Alex, O.** [US/US]; 3916 Wellwood Road, Seaford, NY 11783 (US). **TERZOPOULOS, Demetri** [US/US]; 719 Broadway, 12th Floor, New York, NY (US).

(74) Agent: **ABELEV, Gary**; Dorsey & Whitney LLP, 250 Park Avenue, New York, NY 10177 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

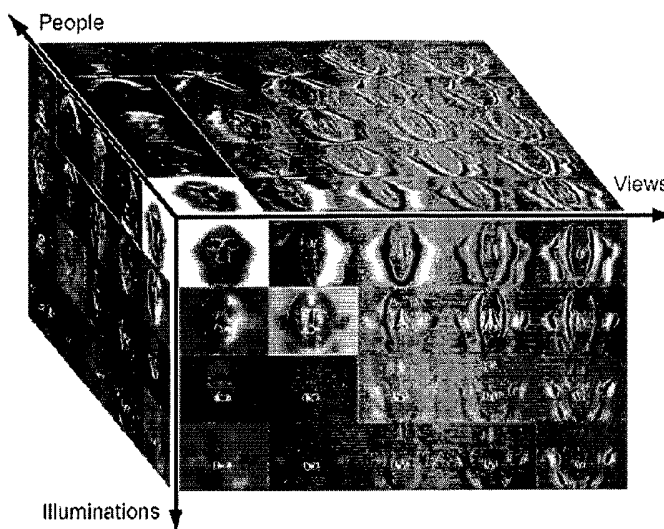
[Continued on next page]

(54) Title: METHOD, SYSTEM, STORAGE MEDIUM, AND DATA STRUCTURE FOR IMAGE RECOGNITION USING MULTILINEAR INDEPENDENT COMPONENT ANALYSIS



300

A



310

B

(57) Abstract: A method, system, computer-readable medium and data structure are provided for processing image data in connection with image recognition. A response of an image to a basis tensor can be determined after the image is applied thereto. The image response can be flattened. A coefficient vector may be extracted from the image response. The extracted coefficient vector may be compared to a plurality of different parameters stored in coefficient rows of a matrix.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD, SYSTEM, STORAGE MEDIUM, AND DATA STRUCTURE FOR
IMAGE RECOGNITION USING MULTILINEAR INDEPENDENT COMPONENT
ANALYSIS

5 CROSS REFERENCE TO RELATED APPLICATION(S)

The present application claims priority from U.S. Patent Application Serial No. 60/536,210, filed January 13, 2004, entitled "Face Recognition Using Multilinear Independent Component Analysis," the entire disclosure of which is incorporated herein by reference.

10 FIELD OF THE INVENTION

The present invention relates generally to image recognition. In particular, the present invention relates to a logic arrangement, data structure, system and method for acquiring data, and more particularly to a logic arrangement, data structure, system and method for acquiring data describing at least one characteristic
15 of an object, synthesizing new data, recognizing acquired data and reducing the amount of data describing one or more characteristics of the object (e.g., a human being).

BACKGROUND INFORMATION

An important problem in data analysis for pattern recognition and
20 signal processing is finding a suitable representation. For historical and computational simplicity reasons, linear models that optimally encode particular statistical properties of the data have been desirable. In particular, the linear, appearance-based face recognition method known as "Eigenfaces" is based on the principal component analysis ("PCA") technique of facial image ensembles. See L.
25 Sirovich et al., "Low dimensional procedure for the characterization of human faces," Journal of the Optical Society of America A, 4:519-524, 1987, and M.A. Turk and A.P. Pentland, "Face recognition using eigenfaces," Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586-590, Hawaii, 1991, both of which are hereby incorporated by this reference. The PCA
30 technique encodes pairwise relationships between pixels, the second-order statistics,

correlational structure of the training image ensemble, but it ignores all higher-order pixel relationships, the higher-order statistical dependencies. In contrast, a generalization of the PCA technique known as independent component analysis (“ICA”) technique learns a set of statistically independent components by analyzing the higher-order dependencies in the training data in addition to the correlations. See A. Hyvarinen et al., Independent Component Analysis, Wiley, New York, 2001, which is hereby incorporated by this reference. However, the ICA technique does not distinguish between higher-order statistics that rise from different factors inherent to an image formation—factors pertaining to scene structure, illumination and imaging.

The ICA technique has been employed in face recognition and, like the PCA technique, it works best when person identity is the only factor that is permitted to vary. See M.S. Bartlett, “Face Image Analysis by Unsupervised Learning,” Kluwer Academic, Boston, 2001, and M.S. Bartlett et al., “Face recognition by independent component analysis,” IEEE Transactions on Neural Networks, 13(6):1450-1464, 2002, both of which are hereby incorporated by this reference. If additional factors, such as illumination, viewpoint, and expression can modify facial images, recognition rates may decrease dramatically. The problem is addressed by multilinear analysis, but the specific recognition algorithm proposed in M.A.O. Vasilescu et al. “Multilinear analysis for facial image recognition,” In Proc. Int. Conf on Pattern Recognition, Quebec City, August 2002 was based on linear algebra, and such algorithm does not fully exploit the multilinear approach.

OBJECTS AND SUMMARY OF THE INVENTION

One of the objects of exemplary embodiments of the present invention is to overcome the above-described deficiencies. Another object of the present invention is to provide a method, system, storage medium, and data structure for generating an object descriptor.

According to an exemplary embodiment of the present invention such method can include steps of computing a response of an image to a basis tensor, flattening the image response, extracting a coefficient vector from the image response,

and comparing the extracted coefficient to a plurality of different parameters stored in rows of a coefficient matrix.

In another exemplary embodiment of the present invention, a computer system can be provided which includes a storage arrangement (e.g., a memory), and a processor which is capable of receiving data associated with an image, and provided in communication with the storage arrangement. The storage arrangement can store computer-executable instructions for performing a method of processing data. For example, a response of an image to a basis tensor may be determined, the image response can be flattened, a coefficient vector can be extracted from the image response, and the extracted coefficient may be compared to a plurality of different parameters stored in rows of a coefficient matrix.

In yet another exemplary embodiment of the present invention, a computer-readable medium is provided having stored thereon computer-executable instructions for performing a method. The method includes steps of computing a response of an image to a basis tensor, flattening the image response, extracting a coefficient vector from the image response, and comparing the extracted coefficient to a plurality of different parameters stored in rows of a coefficient matrix.

In yet another exemplary embodiment of the present invention, a method of processing data is provided. The method includes steps of applying a multilinear independent component analysis to image data to create a factorial code; and generating a representation of the data having a plurality of sets of coefficients that encode people, viewpoints, and illuminations, wherein each set is statistically independent.

In yet another exemplary embodiment of the present invention, a data structure is provided having set of coefficient vectors for a target image. In particular, the vectors include an identifier of the target, a viewpoint of the target, and an illumination direction of the target.

BRIEF DESCRIPTION OF THE DRAWINGS

Further objects, features and advantages of the invention will become apparent from the following detailed description taken in conjunction with the accompanying figures showing illustrative embodiments of the invention, in which:

Figure 1A illustrates exemplary Principal Components Analysis (“PCA”) eigenvectors (e.g., eigenfaces), which are the principal axes of variation across all images;

Figure 1B shows a partial visualization of an exemplary “TensorFaces” representation;

Figure 2A illustrates independent components for image data which can be used with the exemplary embodiment of the present invention;

Figure 2B shows a partial visualization of an exemplary multilinear independent component analysis (“MICA”) representation which can be used with the exemplary embodiment of the present invention;

Figure 3A shows an image representation of the set of coefficient vectors used by the either by “TensorFaces” or MICA representation to represent a particular image which can be used with the exemplary embodiment of the present invention;

Figure 3B shows an image response of a new image to a basis tensor;

Figure 4A shows the full set of images for one of the subjects with viewpoints arrayed horizontally and illuminations arrayed vertically which can be used with the exemplary embodiment of the present invention;

Figure 4B shows a scan of subjects from which the images in Figure 4A were rendered;

Figure 5 shows a block diagram of one exemplary embodiment of a computer system that performs the MICA procedure according to the present invention;

Figure 6 shows a block diagram of one exemplary embodiment of a method of processing image data using a multilinear independent component analysis according to the present invention; and .

Figure 7A-7F are block diagrams of sample tensors and equivalent mode-1, mode-2 and mode-3 flattened tensors according to an exemplary embodiment of the present invention.

Throughout the figures, the same reference numerals and characters, unless otherwise stated, are used to denote like features, elements, components or portions of the illustrated embodiments. Moreover, while the present invention will now be described in detail with reference to the figures, it is done so in connection with the illustrative embodiments. It is intended that changes and modifications can be made to the described embodiments without departing from the true scope and spirit of the subject invention as defined by the appended claims.

DETAILED DESCRIPTION

Exemplary embodiments of the present invention relate to and may utilize a nonlinear, multifactor model of image ensembles that generalizes conventional ICA technique. Whereas the ICA technique employs linear (matrix) algebra, the exemplary embodiment of a Multilinear ICA ("MICA") procedure according to the present invention uses multilinear (tensor) algebra. Unlike its conventional, linear counterpart, the MICA procedure is able to learn the interactions of multiple factors inherent to image formation and separately encode the higher order statistics of each of these factors. Unlike the multilinear generalization of "Eigenfaces", referred to as "TensorFaces" which encodes only second order statistics associated with the different factors inherent to image formation, the MICA procedure can also encode higher order dependencies associated with the different factors.

The multilinearICA procedure of the exemplary embodiment of the present invention can be understood in the context of the mathematics of PCA, multilinearPCA, and ICA techniques.

For example, the principal component analysis of an ensemble of I_2 images can be determined by performing a singularvalue decomposition (“SVD”) on a $I_1 \times I_2$ data matrix $\mathbf{D}$ whose columns are the “vectored” I_1 -pixel “centered” images. Figure 1A illustrates the exemplary PCA eigenvectors 300 (eigenfaces), which are the principal axes of variation across all images. Figure 1B shows a partial visualization of the $75 \times 6 \times 6 \times 8560$ TensorFaces representation 310 of $\mathcal{D}$, obtained as $T = Z \times_s$ $\mathbf{U}_{\text{pixels}}$. Each vectored-centered image can be obtained by subtracting the mean image of the ensemble from each input image, and likely identically arranging the resulting pixels into a column vector. The matrix $\mathbf{D} \in \mathbb{R}^{I_1 \times I_2}$ is a two-mode mathematical object that has two associated vector spaces, a row space and a column space.

In a factor analysis of $\mathbf{D}$, the SVD technique orthogonalizes these two spaces and decomposes the matrix as

$$\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (1)$$

with the product of an orthogonal column-space represented by the left matrix $\mathbf{U} \in \mathbb{R}^{I_1 \times J_1}$, a diagonal singular value matrix $\mathbf{\Sigma} \in \mathbb{R}^{I_1 \times J_2}$ with diagonal entries $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ called the singular values of $\mathbf{D}$, and an orthogonal row space represented by the right matrix $\mathbf{V} \in \mathbb{R}^{J_2 \times J_2}$. The eigenvectors $\mathbf{U}$ are also called the “principal component” (or Karhunen-Loeve) directions of $\mathbf{D}$. Exemplary optimal dimensionality reduction in matrix principal component analysis can be obtained by a truncation of the singular value decomposition (i.e., deleting eigenvectors associated with the smallest singular values).

The analysis of an ensemble of images resulting from the confluence of multiple factors related to scene structure, illumination, and viewpoint is a problem of multilinear algebra. See M.A.O. Vasilescu et al., “Multilinear analysis of image ensembles: Tensorfaces,” In Proc. European Conf. on Computer Vision (ECCV 2002), pp. 447-460, Copenhagen, Denmark, May 2002. Within this mathematical

framework, the image ensemble can be represented as a higher-order tensor. This image data tensor $\mathcal{D}$ should be decomposed in order to separate and parsimoniously represent the constituent factors. To this end, an *N-mode SVD* procedure may be used as a multilinear extension of the above-mentioned conventional matrix SVD technique.

A tensor, also known as an n-way array or multidimensional matrix or n-mode matrix, is a higher order generalization of a vector (first order tensor) and a matrix (second order tensor). The description and use of tensors is described in greater detail in International patent application publication no. WO 03/055119A3, filed December 6, 2002, entitled "Logic Arrangement, Data Structure, System and Method for Multilinear Representation of Multimodal Data Ensembles for Synthesis, Recognition and Compression," which is hereby incorporated by this reference as though set forth fully herein. For example, a tensor can be defined as a multi-linear mapping over a set of vector spaces. The tensor can be represented in the following manner: $\mathcal{A} \in IR^{I_1 \times I_2 \times \dots \times I_N}$, where $\mathcal{A}$ is a tensor. The order of the tensor $\mathcal{A}$ is N . A tensor can be formed by a group of primitives. Each primitive is a set of mode vectors, such that a first primitive is a set of mode-1 vectors, a second vector is a set of mode-2 vectors, an n^{th} primitive is a set of mode-n vectors, etc. In an alternate embodiment, the primitives can be row vectors of a matrix, column vectors of a matrix, index of a vector, etc. An element of tensor $\mathcal{A}$ can be denoted as $\mathcal{A}_{i_1 \dots i_n \dots i_N}$ or $a_{i_1 \dots i_n \dots i_N}$ or where $1 \leq i_n \leq I_n$. Scalars are denoted by lower case letters ($a, b, \dots$), vectors by bold lower case letters ($\mathbf{a}, \mathbf{b} \dots$), matrices by bold upper-case letters ($\mathbf{A}, \mathbf{B} \dots$), and higher-order tensors by italicized bolded upper-case letters ($\mathcal{A}, \mathcal{B} \dots$) or upper-case caligraphic letters ($\mathcal{A}, \mathcal{B} \dots$).

In tensor terminology, column vectors can be referred to as mode-1 vectors, and row vectors are referred to as mode-2 vectors. Mode- n vectors of an N^{th} order tensor $\mathcal{A} \in IR^{I_1 \times I_2 \times \dots \times I_N}$ are the I_n -dimensional vectors obtained from the tensor $\mathcal{A}$ by varying index i_n while maintaining the other indices as fixed. The mode- n vectors are the column vectors of matrix $\mathbf{A}_{(n)} \in IR^{I_n \times (I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N)}$ that can result

from flattening the tensor $\mathcal{A}$, as shown in Figures 12A – 12F. The flattening procedure shall be described in further detail below. The n -rank of tensor $\mathcal{A} \in IR^{I_1 \times I_2 \times \dots \times I_N}$, denoted R_n , is defined as the dimension of the vector space generated by the mode- n vectors:

$$R_n = rank_n(\mathcal{A}) = rank(\mathbf{A}_{(n)}).$$

Figures 7A – 7C illustrate third order tensors 1200, 1210, 1220, respectively, each having dimensions $I_1 \times I_2 \times I_3$. Figure 7D shows the third order tensor 1200 after having been mode-1 flattened to obtain a matrix 1250 containing mode-1 vectors of the third order tensor 1200. The third order tensor 1200 of Fig. 12A is a cube type structure, while the matrix 1250 is a two dimensional type structure having one index, i.e., I_2 , embedded (to a certain degree) within the matrix 1250. Figure 7E shows a matrix 1260 containing mode-2 vectors of the third order tensor 1210 after it has been mode-2 flattened. This third order tensor 1210 is a cube type structure, while the matrix 1260 is a two dimensional type structure having one index, e.g., I_3 , imbedded (to a certain degree) with the data. Figure 7F shows the third order tensor 1220 after having been mode-3 flattened to obtain a matrix 1270 containing mode-3 vectors of the third order tensor 1220. Such third order tensor 1220 is a cube type structure, while the matrix 1270 organizes is a two dimensional type structure having one index, e.g., I_1 , imbedded (to a certain degree) with the data.

A generalization of the product of two matrices can be the product of the tensor and matrix. The mode- n product of tensor $\mathcal{A} \in IR^{I_1 \times I_2 \times \dots \times I_n \times \dots \times I_N}$ by a matrix $\mathbf{M} \in IR^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{M}$, is a tensor $\mathcal{B} \in IR^{I_1 \times \dots \times I_{n-1} \times J_n \times I_{n+1} \times \dots \times I_N}$, whose entries are $\mathcal{B}_{i_1 \dots i_{n-1} j_n i_{n+1} \dots i_N} = \sum_{i_n} a_{i_1 \dots i_{n-1} i_n i_{n+1} \dots i_N} m_{j_n i_n}$. The entries of the tensor $\mathcal{B}$ are computed by

$$(\mathcal{A} \times_n \mathbf{M})_{i_1 \dots i_{n-1} j_n i_{n+1} \dots i_N} = \sum_{i_n} a_{i_1 \dots i_{n-1} i_n i_{n+1} \dots i_N} m_{j_n i_n}.$$

The mode- n product can be expressed as $\mathcal{B} = \mathcal{A} \times_n \mathbf{M}$, or in terms of flattened matrices as $\mathbf{B}_{(n)} = \mathbf{M} \mathbf{A}_{(n)}$. The mode- n product of a tensor and a matrix is a special case of the inner product in multilinear algebra and tensor analysis. The mode- n

product is often denoted using Einstein summation notation, but for purposes of clarity, the mode- n product symbol can be used. The mode- n product may have the following properties:

1. Given a tensor $\mathcal{A} \in IR^{I_1 \times \dots \times I_n \times \dots \times I_m}$ and two matrices, $\mathbf{U} \in IR^{J_m \times I_m}$ and
- 5 $\mathbf{V} \in IR^{J_n \times I_n}$ the following property holds true:

$$\begin{aligned}\mathcal{A} \times_m \mathbf{U} \times_n \mathbf{V} &= (\mathcal{A} \times_m \mathbf{U}) \times_n \mathbf{V} \\ &= (\mathcal{A} \times_n \mathbf{V}) \times_m \mathbf{U} \\ &= \mathcal{A} \times_n \mathbf{V} \times_m \mathbf{U}\end{aligned}$$

2. Given a tensor $\mathcal{A} \in IR^{I_1 \times \dots \times I_n \times \dots \times I_N}$ and two matrices, $\mathbf{U} \in IR^{J_n \times I_n}$ and
- 10 $\mathbf{V} \in IR^{K_n \times J_n}$ the following property holds true:

$$(\mathcal{A} \times_n \mathbf{U}) \times_n \mathbf{V} = \mathcal{A} \times_n (\mathbf{V}\mathbf{U}).$$

- An N^{th} -order tensor $\mathcal{A} \in IR^{I_1 \times I_2 \times \dots \times I_N}$ has a rank-1 when it can be expressed as the outer product of N vectors: $\mathcal{A} = \mathbf{u}_1 \circ \mathbf{u}_2 \circ \dots \circ \mathbf{u}_N$. The tensor element is expressed as $a_{ij\dots m} = u_{1i}u_{2j}\dots u_{Nm}$, where u_{1i} is the i^{th} component of $\mathbf{u}_1$, etc. The rank
- 15 of a N^{th} order tensor $\mathcal{A}$, denoted $R = \text{rank}(\mathcal{A})$, is the minimal number of rank-1 tensors that yield $\mathcal{A}$ in a linear combination:

$$\mathcal{A} = \sum_{r=1}^R \sigma_r \mathbf{u}_1^{(r)} \circ \mathbf{u}_2^{(r)} \circ \dots \circ \mathbf{u}_N^{(r)}.$$

A singular value decomposition (SVD) can be expressed as a rank decomposition as is shown in the following simple example:

$$\begin{aligned}\mathbf{M} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} \begin{bmatrix} f & g \\ h & i \end{bmatrix} = \sigma_{11} \begin{bmatrix} a \\ c \end{bmatrix} \circ \begin{bmatrix} f \\ g \end{bmatrix} + \sigma_{22} \begin{bmatrix} b \\ d \end{bmatrix} \circ \begin{bmatrix} h \\ i \end{bmatrix} \\ &= \mathbf{U}_1 \Sigma \mathbf{U}_2^T \\ &= \begin{bmatrix} \mathbf{u}_1^{(1)} & \mathbf{u}_1^{(2)} \end{bmatrix} \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{u}_2^{(1)} & \mathbf{u}_2^{(2)} \end{bmatrix}^T \\ &= \sum_{i=1}^{R=2} \sum_{j=1}^{R=2} \sigma_{ij} \mathbf{u}_1^{(i)} \circ \mathbf{u}_2^{(j)}\end{aligned}$$

It should be noted that an SVD can be a combinatorial orthogonal rank decomposition, but that the reverse is not true. In general, rank decomposition is not necessarily singular value decomposition. Also, the N -mode SVD can be expressed
 5 as an expansion of mutually orthogonal rank-1 tensors, as follows:

$$\mathcal{D} = \sum_{i_1=1}^{R_1} \dots \sum_{i_n=1}^{R_n} \dots \sum_{i_N=1}^{R_N} \dots \mathcal{Z}_{i_1 \dots i_N} \mathbf{U}_1^{(i_1)} \circ \dots \circ \mathbf{U}_n^{(i_n)} \circ \dots \circ \mathbf{U}_N^{(i_N)},$$

where $\mathbf{U}_n^{(i_n)}$ is the i_n column vector of the matrix $\mathbf{U}_n$. This is analogous to the equation

$$\sum_{i=1}^{R=2} \sum_{j=1}^{R=2} \sigma_{ij} \mathbf{u}_1^{(i)} \circ \mathbf{u}_2^{(j)}.$$

For example, an order $N > 2$ tensor or N -way array $\mathcal{D}$ is an N -
 10 dimensional matrix comprising N spaces. N -mode SVD is a “generalization” of conventional matrix (i.e., 2-mode) SVD. It can orthogonalize these N spaces, and decompose the tensor as the *mode- n product*, denoted $\times_n$, of N -orthogonal spaces, as follows:

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \dots \times_n \mathbf{U}_n \dots \times_N \mathbf{U}_N. \quad (2)$$

15 Tensor $\mathcal{Z}$, known as the *core tensor*, is analogous to the diagonal singular value matrix in conventional matrix SVD (although it does not have a simple, diagonal structure). Using mode- n products, the conventional SVD in equation (1) can be rewritten as $\mathbf{D} = \mathbf{\Sigma} \times_1 \mathbf{U} \times_2 \mathbf{V}$. The core tensor governs the interaction between the *mode matrices* $\mathbf{U}_1, \dots, \mathbf{U}_N$. Mode matrix $\mathbf{U}_n$ contains the orthonormal vectors
 20 spanning the column space of matrix $\mathbf{D}_{(n)}$ resulting from the *mode- n flattening* of $\mathcal{D}$.

An N -mode SVD technique can be used for decomposing $\mathcal{D}$ according to equation (2):

1. For $n = 1, \dots, N$, compute matrix $\mathbf{U}_n$ in equation (2) by
 computing the SVD of the flattened matrix $\mathbf{D}_{(n)}$ and setting $\mathbf{U}_n$ to be the left matrix of
 25 the SVD. When $\mathbf{D}_{(n)}$ is a non-square matrix, the computation of $\mathbf{U}_n$ in the singular
 value decomposition (“SVD”) $\mathbf{D}_{(n)} = \mathbf{U}_n \mathbf{\Sigma}_n^T \mathbf{V}_n^T$ can be performed efficiently, depending
 on which dimension of $\mathbf{D}_{(n)}$ is smaller, by decomposing either $\mathbf{D}_{(n)} \mathbf{D}_{(n)}^T = \mathbf{U}_n \mathbf{\Sigma}_n^2 \mathbf{U}_n^T$ and

then computing $\mathbf{V}_n^T = \Sigma + \mathbf{U}_n^T \mathbf{D}_{(n)}$ or by decomposing $\mathbf{D}_{(n)}^T \mathbf{D}_{(n)} = \mathbf{V}_n \Sigma^2 \mathbf{V}_n^T$ and then computing $\mathbf{U}_n = \mathbf{D}_{(n)} \mathbf{V}_n \Sigma^+$.

2. Solve for the core tensor as follows:

$$\mathcal{Z} = \mathcal{D} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \dots \times_n \mathbf{U}_n^T \dots \times_N \mathbf{U}_N^T. \quad (3)$$

5 A dimensionality reduction in the linear case does not have a trivial multilinear counterpart. A useful generalization to tensors can involve an optimal rank- $(R_1, R_2, \dots, R_N)$ approximation which iteratively optimizes each of the modes of the given tensor, where each optimization step involves a best reduced-rank approximation of a positive semi-definite symmetric matrix. See L. de Lathauwer et al., "On the best rank-1 and rank- $(R_1, R_2, \dots, R_n)$ approximation of higher order tensors," SIAM Journal of Matrix Analysis and Applications, 21(4):1324-1342, 2000. This technique is a higher-order extension of the orthogonal iteration for matrices. Figure 2B shows a tensor basis 330 associated with multilinear PCA, as discussed further herein.

15 The independent component analysis ("ICA") technique of multivariate data looks for a sequence of projections such that the projected data look as far from Gaussian as possible. The ICA technique can be applied in two exemplary ways: Architecture I applies the ICA technique to $\mathbf{D}^T$, each of whose rows is a different vectorized image, and finds a spatially independent basis set that reflects the local properties of faces. On the other hand, architecture II applies the ICA technique to $\mathbf{D}$ and finds a set of coefficients that are statistically independent while the basis reflects the global properties of faces.

25 **Architecture I:** the ICA technique starts essentially from the factor analysis or PCA solution shown in equation (1) of a pre-whitened data set, and computes a rotation of the principal components such that they become *independent components*. See J. Friedman et al., "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," Springer, New York, 2001. Figure 2A illustrates independent components $\mathbf{C}_{\text{pixels}}$ for image data 320. The ICA technique can rotate the principal component directions $\mathbf{U}$ in (1) as follows:

$$\mathbf{D}^T = \mathbf{V}\Sigma\mathbf{U}^T \quad (4)$$

$$= (\mathbf{V}\Sigma\mathbf{W}^{-1})(\mathbf{W}\mathbf{U}^T) \quad (5)$$

$$= \mathbf{K}^T\mathbf{C}^T, \quad (6)$$

where every column of $\mathbf{D}$ is a different image, $\mathbf{W}$ is an invertible transformation matrix that is computed by the ICA technique, $\mathbf{C} = \mathbf{U}\mathbf{W}^T$ are the independent components shown in Figure 2A, and $\mathbf{K} = \mathbf{W}^{-T}\Sigma\mathbf{V}^T$ are the coefficients. Various objective functions, such as those based on mutual information, negentropy, higher-order cumulants, etc., can be presented using conventional procedures for computing the independent components along with different optimization methods for extremizing these objective functions. Dimensionality reduction with the ICA technique is usually performed in the PCA preprocessing stage. A mathematically equivalent way of representing the above expressions that will serve better in the next section is, as follows:

$$\mathbf{D} = (\mathbf{U}\mathbf{W}^T)(\mathbf{W}^{-T}\Sigma\mathbf{V}^T) \quad (7)$$

$$= \mathbf{C}\mathbf{K}. \quad (8)$$

Alternatively, in **architecture II**, ICA can be applied to $\mathbf{D}$ and it rotates the principal components directions such that the coefficients are statistically independent, as follows:

$$\mathbf{D} = \mathbf{U}\Sigma\mathbf{V}^T \quad (9)$$

$$= (\mathbf{U}\mathbf{W}^{-1})(\mathbf{W}\Sigma\mathbf{V}^T) \quad (10)$$

$$\mathbf{D} = \mathbf{C}\mathbf{K}. \quad (11)$$

where $\mathbf{C}$ are the basis and $\mathbf{K}$ are the independent coefficients. Note, that $\mathbf{C}$, $\mathbf{K}$ and $\mathbf{W}$ are computed differently in the two different architectures.

Figure 5 shows a block diagram of an exemplary embodiment of a computer system 100 according to the present invention which may be used to implement the MICA procedure according to the present invention. For example, the system 100 can include an image capturing system 120 in communication with a processor 130. The system may further include a storage arrangement (e.g., a

memory 110) in communication with the processor 130. The memory 110 may store a MICA application 112 that includes compute-executable instructions that are capable of being executed by the processor 130 on data captured by the image capturing system 120. The memory 110 may further store an image data base used to
 5 compare image data processed according to the MICA procedure.

Figure 6 shows an exemplary embodiment of a flow chart of the method 200 of the present invention. For example, as shown in Figure 6, a response of an image to a basis tensor is computed (step 210). The image response is flattened (step 220). A coefficient vector may be extracted from the image response (step 230).
 10 The extracted coefficient may be compared to a plurality of different parameters stored in coefficient rows of a matrix (step 240).

As indicated above, the exemplary MICA procedure according to the present invention may be implemented in the exemplary embodiments of the present invention. For example, architecture I can apply the MICA procedure, and may result
 15 in a factorial code. It can locate a representation in which each set of coefficients that encodes people, viewpoints, illuminations, etc., is statistically independent. Architecture II finds a set of independent bases across people, viewpoints, illuminations, etc.

Architecture I: Transposing the flattened data tensor $\mathcal{D}$ in the n th
 20 mode and computing the ICA as in (4)-(8):

$$\mathbf{D}_{(n)} = \mathbf{U}_n \mathbf{Z}_{(n)} \mathbf{V}_n^T \quad (12)$$

$$= (\mathbf{U}_n \mathbf{W}_n^T) \mathbf{W}_n^{-T} \mathbf{Z}_{(n)} \mathbf{V}_n^T \quad (13)$$

$$= \mathbf{C}_n \mathbf{W}_n^{-T} \mathbf{Z}_{(n)} \mathbf{V}_n^T, \quad (14)$$

where $\mathbf{C}_n = \mathbf{U}_n \mathbf{W}_n^T$. Thus, we can derive the N -mode ICA from N -mode SVD (2) as
 25 follows:

$$\begin{aligned} \mathcal{D} &= \mathbf{Z} \times_1 \mathbf{U}_1 \dots \times_N \mathbf{U}_N \\ &= \mathbf{Z} \times_1 \mathbf{U}_1 \mathbf{W}_1^T \mathbf{W}_1^{-T} \dots \times_N \mathbf{U}_N \mathbf{W}_N^T \mathbf{W}_N^{-T} \end{aligned}$$

$$\begin{aligned}
&= Z \times_1 \mathbf{C}_1 \mathbf{W}_1^{-T} \dots \times_N \mathbf{C}_N \mathbf{W}_N^{-T} \\
&= (Z \times_1 \mathbf{W}_1^{-T} \dots \times_N \mathbf{W}_N^{-T}) \times_1 \mathbf{C}_1 \dots \times_N \mathbf{C}_N \\
&= S \times_1 \mathbf{C}_1 \dots \times_N \mathbf{C}_N, \quad (15)
\end{aligned}$$

where the core tensor $S = Z \times_1 \mathbf{W}_1^{-T} \dots \times_N \mathbf{W}_N^{-T}$. The rows associated with each of the
5 mode matrices, $\mathbf{C}_i$ where $i = 1 \dots N$ are statistically independent.

A multilinear ICA decomposition can be performed of the tensor $\mathcal{D}$ of
vectorized training images $\mathbf{d}_d$,

$$\mathcal{D} = \mathcal{B} \times_1 \mathbf{C}_{\text{people}} \times_2 \mathbf{C}_{\text{views}} \times_3 \mathbf{C}_{\text{illums}}, \quad (16)$$

extracting a set of mode matrices—the matrix $\mathbf{C}_{\text{people}}$ containing row vectors of $\mathbf{c}_p^T$
10 coefficients for each person p , the matrix $\mathbf{C}_{\text{views}}$ containing row vectors $\mathbf{c}_v^T$ of
coefficients for each view direction v , the matrix $\mathbf{C}_{\text{illums}}$ containing row vectors $\mathbf{c}_l^T$ of
coefficients for each illumination direction l —and an MICA basis tensor $B =$
 $S \times_4 \mathbf{C}_{\text{pixels}}$ that governs the interaction between the different mode matrices as
illustrated by Figure 2B. Figure 2B shows a partial visualization 330 of the
15 $75 \times 6 \times 6 \times 8560$ MICA representation of $\mathcal{D}$, obtained as $\mathcal{B} = S \times_5 \mathbf{C}_{\text{pixels}}$.

For architecture I, each of the mode matrices contains a set of
statistically independent coefficients, while architecture II yields B , a set of
independent bases across people, viewpoints, illuminations, etc.

Architecture II: MICA has the same mathematical form as in
20 equation (15). However, the core tensor S and mode matrices $\mathbf{C}_1 \dots \mathbf{C}_N$ are computed
according to equations (9)-(11). This architecture results in a set of basis vectors that
are statistically independent across the different modes.

In the PCA or eigenface technique, a data matrix $\mathbf{D}$ of known
“training” facial images $\mathbf{d}_d$ is decomposed into a reduced-dimensional basis matrix
25 $\mathbf{B}_{\text{PCA}}$, and a matrix $\mathbf{C}$ containing a vector of coefficients $\mathbf{c}_d$ associated with each
vectorized image $\mathbf{d}_d$. Given an unknown facial image $\mathbf{d}_{\text{new}}$, the projection operator

$\mathbf{B}_{\text{PCA}}^{-1}$ linearly projects this new image into the reduced-dimensional space of image coefficients, $\mathbf{c}_{\text{new}} = \mathbf{B}_{\text{PCA}}^{-1} \mathbf{d}_{\text{new}}$.

The recognition procedure described in the reference immediately following was based on this linear projection approach, so it does not fully exploit the multilinear framework. See M.A.O. Vasilescu et al. "Multilinear analysis for facial image recognition," In Proc. Int. Conf on Pattern Recognition, Quebec City, August 2002. One exemplary embodiment of the present invention addresses the fundamental problem of inferring the identity, illumination, viewpoint, and expression labels of an unlabeled test image. Given a solution to this problem, a simple recognition algorithm is obtained that is based on the statistical independence properties of ICA and the multilinear structure of the tensor framework.

Figure 3A shows an exemplary image representation 340 of the set of coefficient vectors used by the MICA procedure to represent an image:

$\mathcal{d}_i = \mathcal{B} \times_1 \mathbf{c}_p^T \times_2 \mathbf{c}_v^T \times_3 \mathbf{c}_l^T$. The first coefficient vector encodes the person's identity, the second encodes the view point, the third encodes the illumination direction, etc.

Figure 3B shows an image response 350 of a new image to a basis tensor. Given a new image from which the person, view and illumination direction remain to be determined, the image response to the basis tensor $\mathcal{B} = \mathcal{S} \times_4 \mathbf{C}_{\text{pixels}}$ can be computed as

$$\mathbf{R}_{(\text{pixels})} = (\mathbf{C}_{(\text{pixels})} \mathbf{S}_{(\text{pixels})})^{-1} \mathbf{d}_{\text{new}}. \quad (17)$$

where $\mathcal{R}$ is a multi-modal response to the different factors that make up the image. This tensorial response has a particular structure shown in Figure 3B, which can be exploited in order to extract the people parameters associated with $\mathbf{d}_{\text{new}}$. For the sake of clarity, this image was constructed from the following unknown set of coefficient vectors: an illumination coefficient vector $\mathbf{c}_l$, a view point coefficient vector $\mathbf{c}_v$, and a person coefficient vector $\mathbf{c}_p$. The face recognition problem is the classification of the computed person coefficient vector.

$$\mathbf{R}_{(\text{people})} = [l_1 v_1 \mathbf{c}_p \dots l_1 v_n \mathbf{c}_p \dots l_n v_n \mathbf{c}_p \dots l_n v_n \mathbf{c}_p] \quad (18)$$

The image response $\mathcal{R}$ can be reorganized as a matrix whose columns are multiples of the people parameters $\mathbf{c}_p$ in equation (18). The reorganization of the image responses is achieved by flattening along the people mode. On closer inspection, the matrix $\mathbf{R}_{(\text{people})}$ has rank 1, it's column are multiples of $\mathbf{c}_p$. Therefore, the people rank of $\mathcal{R}$ is 1; hence, an SVD of $\mathbf{R}_{(\text{people})}$ can extract $\mathbf{c}_p$. Similarly, a flattening along the viewpoint mode or the illumination mode etc. results in the matrices $\mathbf{R}_{(\text{viewpoints})}$ and $\mathbf{R}_{(\text{illumination})}$ can be observed whose columns are multiples of $\mathbf{c}_v$ and $\mathbf{c}_l$ respectively. These coefficient vectors, the viewpoint coefficient vector, the illumination coefficient vector etc., are extracted by computing a singular value decomposition on the respective matrices.

Therefore, as indicated above, all the constituent factors, and all the coefficient vectors, associated with a test image can be extracted by computing the N-mode SVD on the multimodal response tensor $\mathcal{R}$, whose rank- $(R_1, R_2, \dots, R_n) = \text{rank}-(1, 1, \dots, 1)$.

Using the extracted $\mathbf{c}_p$, it is possible to perform individual recognition. It can be compared against the people parameters that are stored in the rows of $\mathbf{C}_{\text{people}}$ using a cosine function which is equivalent to a normalized nearest neighbor. Assuming one of the image factors is the facial expression one can similarly perform expression recognition. In this manner all the factors associated with an image can be recognized. In one exemplary embodiment of the present invention, the exemplary procedure can be applied to gray-level facial images of 75 subjects. Each subject can be imaged from 15 different viewpoints ($\theta = -35^\circ$ to $+35^\circ$ in 5° steps on the horizontal plane $\Phi = 0^\circ$) under 15 different illuminations ($\theta = -35^\circ$ to $+35^\circ$ in 5° steps on an inclined plane $\Phi = 45^\circ$). Figure 4A shows the full set 360 of 225 exemplary images for one of the subjects with viewpoints arrayed horizontally and illuminations arrayed vertically. The images boxed may be used for training. The image set 360 can be rendered from a 3D scan 370 of the subject shown as the top left scan in Figure 4B. The 75 scans shown in Figure 4B may be recorded using a Cyberware™ 3030PS laser

scanner and are part of the 3D morphable faces database created at the University of Freiburg.

Recognition Experiment	PCA	ICA	MICA
Training: 75 people, 6 viewpoints ($\theta = \pm 35, \pm 20, \pm 5, \Phi = 0$), 6 illuminations ($\theta = 45, \Phi = 90 + \delta, \delta = \pm 35, \pm 20, \pm 5$) Testing: 75 people, 9 viewpoints ($\theta = 0 \pm 10, \pm 15, \pm 25, \pm 30$), 9 illuminations ($\theta = 90 + \delta, \delta = \pm 35, \pm 20, \pm 5, \theta = 0$)	83.9%	89.5%	98.14%

5 In summary, the independent Component Analysis (“ICA”) technique may minimize the statistical dependence of the representational components of a training image ensemble. However, the ICA technique generally is not able to distinguish between the different factors related to scene structure, illumination and imaging, which are inherent to image formation. A nonlinear, multifactor ICA

10 procedure according to the present invention can be utilized (as described above) that generalizes the ICA technique. For example, the exemplary Multilinear ICA (“MICA”) procedure according to the present invention of image ensembles can learn the statistically independent components of multiple factors. Whereas the ICA technique employs linear (matrix) algebra, the MICA procedure according to the

15 present invention generally may exploit multilinear (tensor) algebra. In the context of facial image ensembles, we demonstrate that the statistical regularities learned by the exemplary MICA procedure can capture information that improves automatic face recognition can be addressed. In this context, we also address an issue fundamental to the multilinear framework for recognition, the inference of mode labels (person,

20 viewpoint, illumination, expression, etc.) of an unlabeled test image.

While the invention has been described in connection with preferred embodiments, it will be understood by those of ordinary skill in the art that other variations and modifications of the preferred embodiments described above may be made without departing from the scope of the invention. Other embodiments will be

apparent to those of ordinary skill in the art from a consideration of the specification or practice of the invention disclosed herein. It is intended that the specification and the described examples are considered as exemplary only, with the true scope and spirit of the invention indicated by the following claims. Additionally, all references
5 cited herein are hereby incorporated by this reference as though set forth fully herein.

CLAIMS

1. A method of processing data, comprising:
applying image data associated with an image to a basis tensor;
obtaining response data based on the application of the image data to
5 the basis tensor; and
extracting a coefficient vector from the response data.
2. The method of claim 1, wherein the response data relates to a multimodal response to factors that compose the image.
3. The method of claim 1, further comprising flattening an image section of the
10 image corresponding to the response data.
4. The method of claim 3, wherein the flattening step comprises flattening the image section using a predetermined mode to arrange the response data in a further matrix.
5. The method of claim 3, wherein the flattening step comprises flattening the
15 image section using a predetermined mode to arrange the response data in a further matrix having columns that are multiples of image parameters of the image.
6. The method of claim 3, wherein the flattening step comprises flattening the response data based on a person mode.
7. The method of claim 1, wherein the extracting step comprises extracting a
20 person coefficient vector from the response data.
8. The method of claim 1, wherein the extracting step comprises extracting the coefficient vector using a singular value decomposition procedure.
9. The method of claim 1, further comprising comparing the extracted coefficient vector to a plurality of parameters provided in a matrix of further data associated with
25 further images.

10. The method of claim 9, wherein the comparing step comprises comparing an extracted person coefficient vector to a plurality of people parameters stored in people coefficient rows of the matrix.
11. The method of claim 10, wherein the comparing step is performed using a cosine function that is approximately equivalent to a normalized nearest neighbor.
12. A computer system, comprising:
a processing arrangement capable of receiving data associated with an image; and
a storage provided in communication with the processing arrangement,
and having stored thereon computer-executable set of instructions, wherein the set of instructions comprise:
i. a first set configured to enable the processing arrangement to apply image data associated with an image to a basis tensor;
ii. a second set configured to enable the processing arrangement to obtain response data based on the application of the image data to the basis tensor; and
iii. a third set configured to enable the processing arrangement to extract a coefficient vector from the response data.
13. The computer system of claim 12, wherein the response data relates to is a multimodal response to factors that compose the image.
14. The computer system of claim 12, wherein the set of instruction further comprises a fourth set configured to enable the processing arrangement to flatten an image section of the image corresponding to the response data.
15. The computer system of claim 14, wherein the fourth set is capable of configuring the processing arrangement to flatten the image section using a predetermined mode to arrange the response data in a further matrix.
16. The computer system of claim 14, wherein the image section is flattened using a predetermined mode to arrange the response data in a further matrix having columns that are multiples of image parameters of the image.

17. The computer system of claim 14, wherein the image section is flattened using based on a person mode.
18. The computer system of claim 12, wherein the third set configures the processing arrangement to extract a person coefficient vector from the response data.
- 5 19. The computer system of claim 12, wherein the third set configures the processing arrangement to extract the coefficient vector using a singular value decomposition procedure.
20. The computer system of claim 12, wherein the set of instruction further comprises a fifth set configured to enable the processing arrangement to comparing
10 the extracted coefficient vector to a plurality of parameters provided in a matrix of further data associated with further images.
21. The computer system of claim 20, wherein the fourth set configures the processing arrangement to compare an extracted person coefficient vector to a plurality of people parameters stored in people coefficient rows of the matrix.
- 15 22. The computer system of claim 21, wherein the comparison is performed using a cosine function that is approximately equivalent to a normalized nearest neighbor.
23. A computer-readable medium having stored thereon computer-executable set of instructions, the set of instructions comprising:
- 20 i. a first set configured to enable a processing arrangement to apply image data associated with an image to a basis tensor;
 - ii. a second set configured to enable the processing arrangement to obtain response data based on the application of the image data to the basis tensor; and
 - 25 iii. a third set configured to enable the processing arrangement to extract a coefficient vector from the response data.
22. A method of processing data, comprising:
- applying a multilinear independent component analysis procedure to image data so as to generate a factorial code; and

based on the results of the applying step, producing a representation of the image data having a plurality of sets of coefficients that are associated with objects, viewpoints, and illuminations, wherein each of the sets is statistically independent from one another.

- 5 23. A data structure which is capable of being stored on a storage arrangement, comprising:

 a set of coefficient vectors associated with an image, wherein the vectors include an identifier of the image, a viewpoint of the image, and an illumination direction of the image.

- 10 24. The data structure of claim 21, wherein the image is an image of a person.

25. The data structure of claim 21, wherein the image is an image of a person's face.

26. The data structure of claim 21, wherein the set of coefficient vectors are arranged in a matrix.

- 15 27. A method for processing data, comprising:

 obtaining information associated with a plurality of images; and
 determining basis tensors associated with the information, wherein the basis tensors at least one of:

- 20 a. are statistically independent, and
 b. include set of coefficients, each of the sets is statistically independent.

28. A computer system, comprising:

 a processing arrangement capable of receiving data associated with an image; and

- 25 a storage provided in communication with the processing arrangement, and having stored thereon computer-executable set of instructions, wherein the set of instructions comprise:

- i. a first set configured to enable the processing arrangement to obtain information associated with a plurality of images;

- ii. a second set configured to enable the processing arrangement to determine basis tensors associated with the information, wherein the basis tensors at least one of:
 - a. are statistically independent, and
 - b. include set of coefficients, each of the sets is statistically independent.
- 5

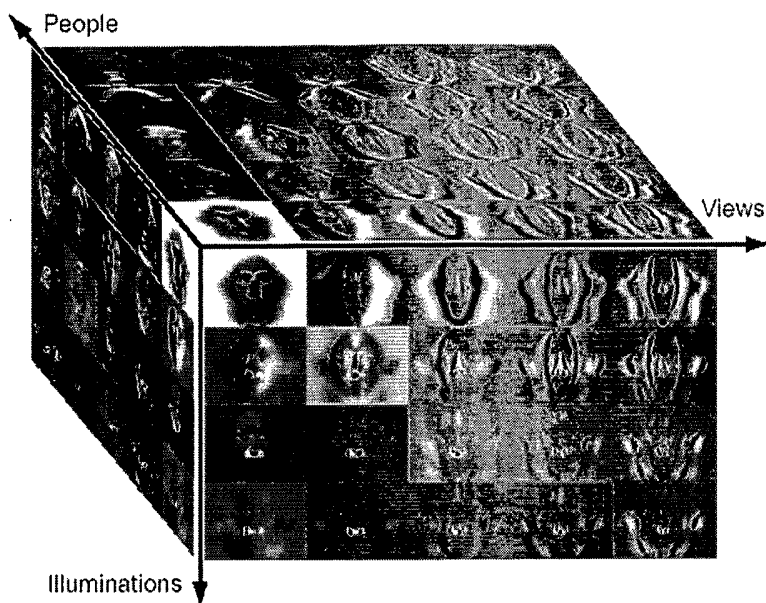
29. A computer-readable medium having stored thereon computer-executable set of instructions, the set of instructions comprising:

- i. a first set configured to enable a processing arrangement to obtain information associated with a plurality of images;
 - ii. a second set configured to enable the processing arrangement to determine basis tensors associated with the information, wherein the basis tensors at least one of:
 - a. are statistically independent, and
 - b. include set of coefficients, each of the sets is statistically independent.
- 10
- 15



300

FIG. 1A



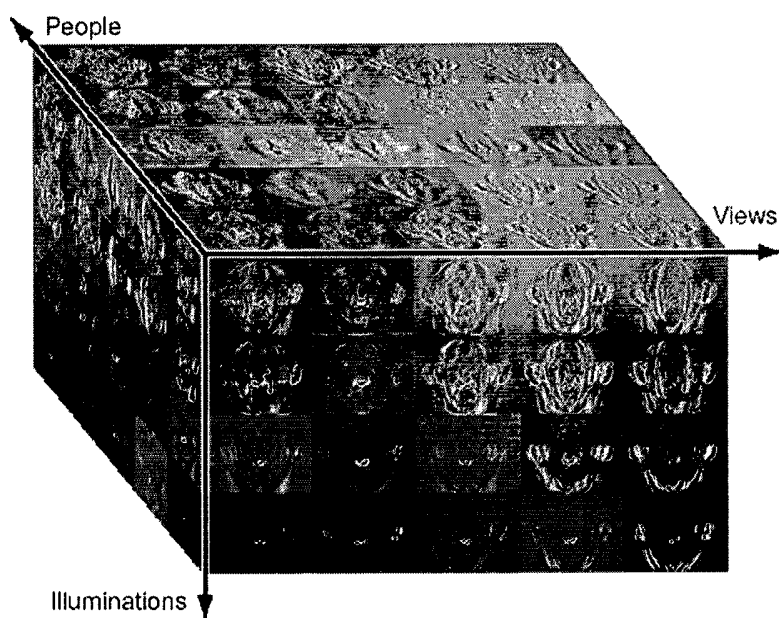
310

FIG. 1B



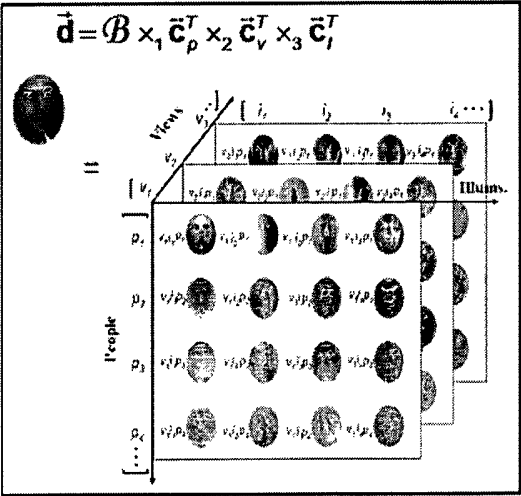
320

FIG. 2A

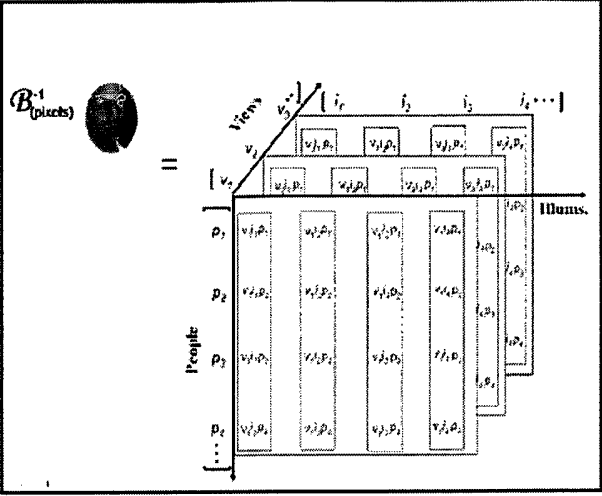


330

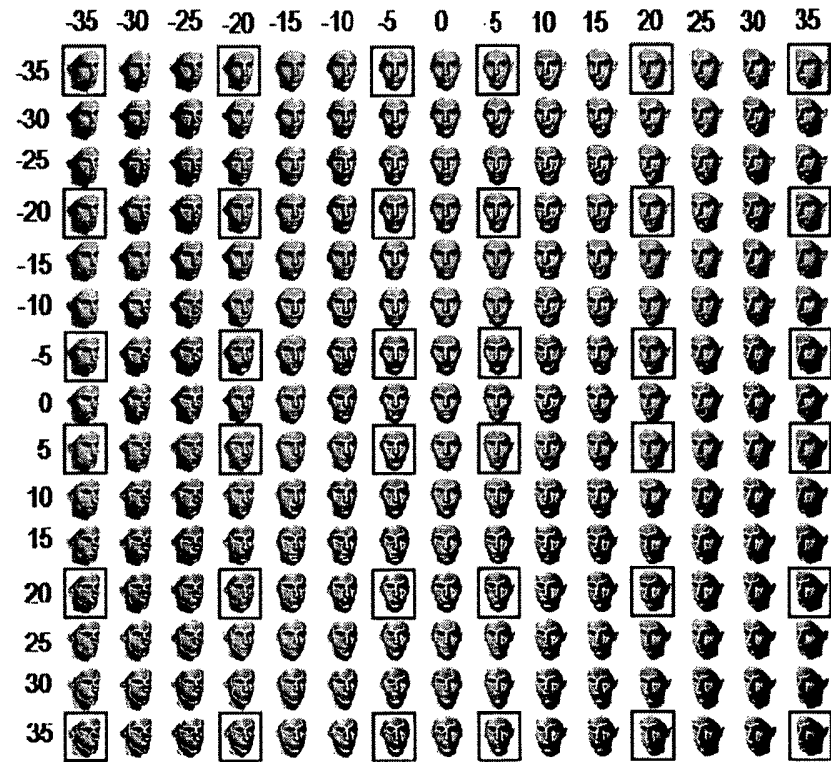
FIG. 2B



340
FIG. 3A



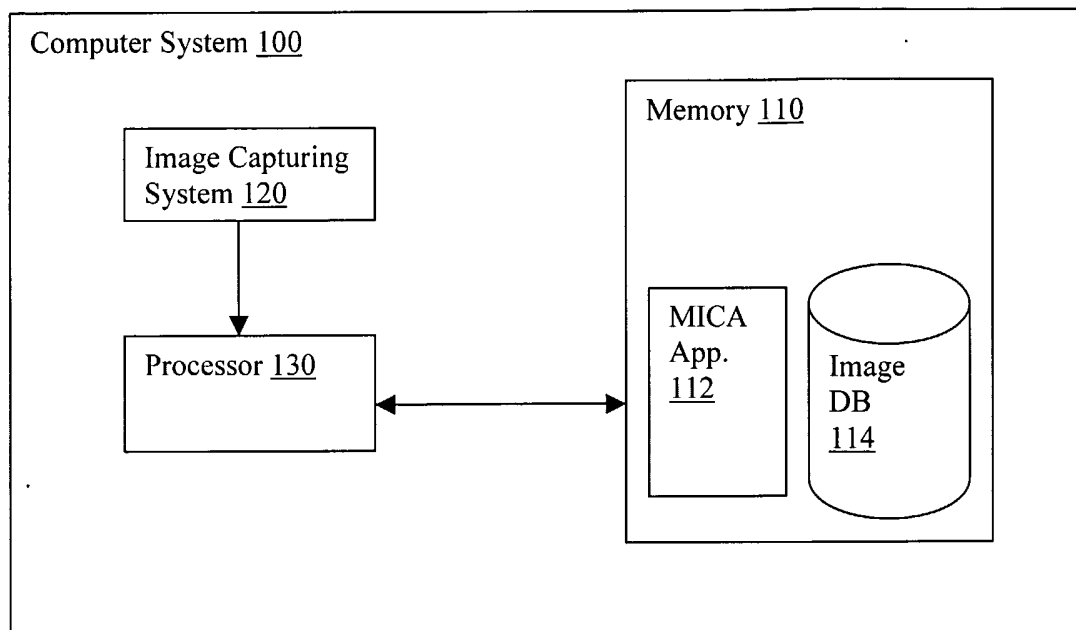
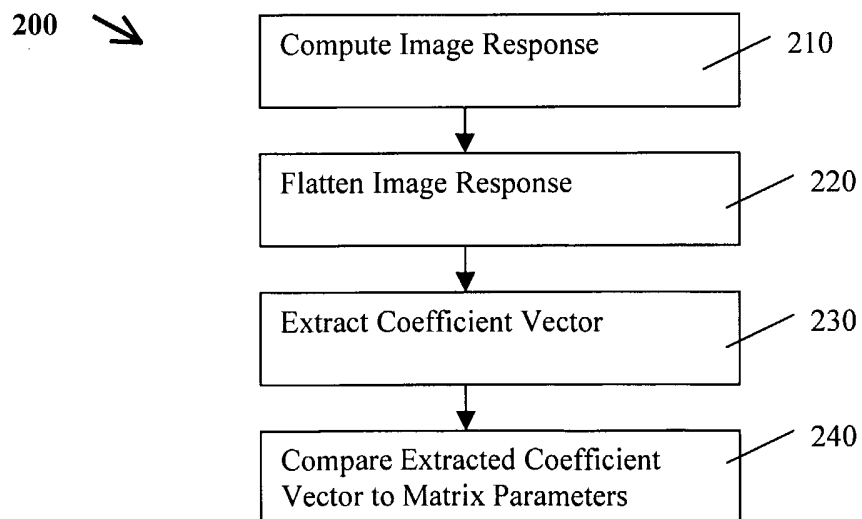
350
FIG. 3B



360
FIG. 4A



370
FIG. 4B

**FIG. 5****FIG. 6**

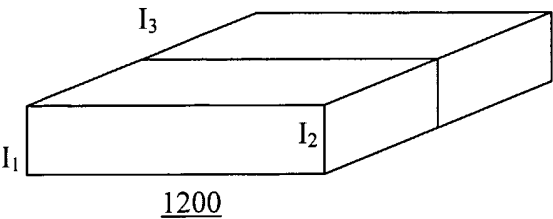


FIG. 7A

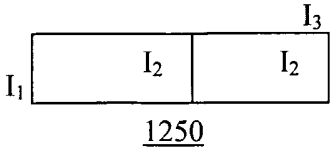


FIG. 7D

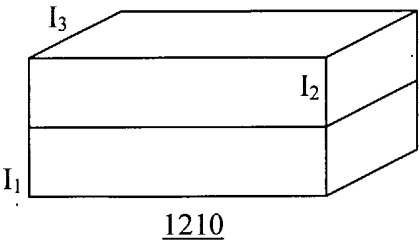


FIG. 7B

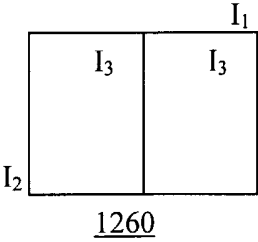


FIG. 7E

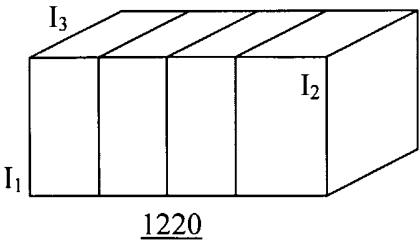


FIG. 7C

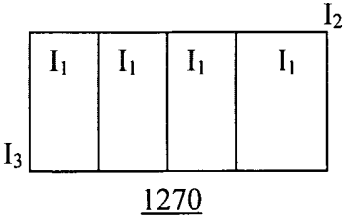


FIG. 7F