



- (51) **International Patent Classification:**
G06K 9/36 (2006.01) *G06K 9/46* (2006.01)
- (21) **International Application Number:**
PCT/US2014/057370
- (22) **International Filing Date:**
25 September 2014 (25.09.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/882,897 26 September 2013 (26.09.2013) US
- (71) **Applicants:** TOKITAE LLC [US/US]; 11235 SE 6th Street, Suite 200, Bellevue, Washington 98004-6481 (US). UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION [US/US]; 4311 11th Ave NE, Suite 500, Seattle, Washington 98105-4608 (US).
- (72) **Inventors:** BRUNTON, Bingni W.; c/o UW Center for Commercialization, 4311 11th Avenue NE, Suite 500, Seattle, Washington 98105-4608 (US). BRUNTON, Steven L.; c/o UW Center for Commercialization, 4311 11th Avenue NE, Suite 500, Seattle, Washington 98105-4608 (US). KUTZ, J. Nathan; c/o UW Center for Commercialization, 4311 11th Avenue NE, Suite 500, Seattle,

Washington 98105-4608 (US). PROCTOR, Joshua L.; 7742 16th Ave NE, Seattle, Washington 98115 (US).

(74) **Agents:** MALASKA, Stephen L. et al.; 11235 SE 6th Street, Suite 200, Bellevue, Washington 98004-6841 (US).

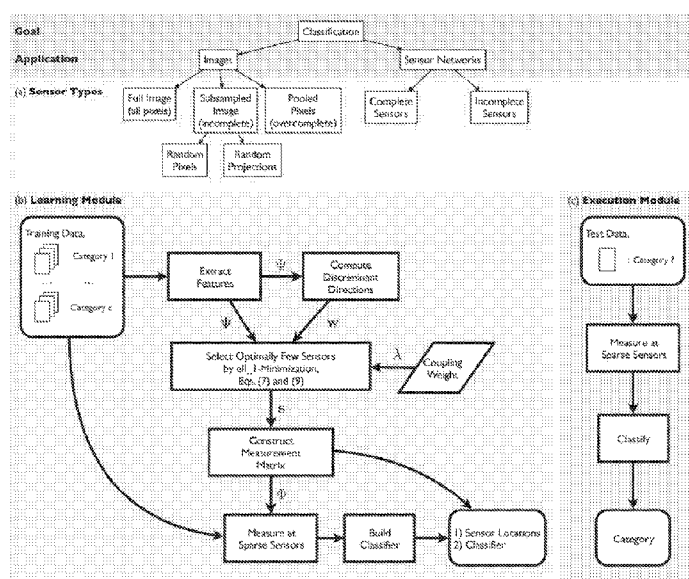
(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) **Title:** SYSTEMS, DEVICES, AND METHODS FOR CLASSIFICATION AND SENSOR IDENTIFICATION USING ENHANCED SPARSITY

FIG. 9



(57) **Abstract:** The goal of compressive sensing is efficient reconstruction of data from sparse measurements, often leading to a categorical decision. If only classification is required, reconstruction can be circumvented and the measurements needed are sparser still by orders-of-magnitude. "Enhanced sparsity" is used herein as the reduction in the number of measurements required for classification over reconstruction. Systems and methods exploit such enhanced sparsity and learn spatial sensor locations that optimally inform a categorical decision. Systems and methods include algorithms that solve an l1 minimization to find the fewest entries of the full measurement vector that exactly reconstruct the discriminant vector in feature space. Once the sensor locations have been identified from the training data, subsequent test samples are classified with remarkable efficiency, with performance comparable to that obtained by discrimination using the full image. Sensor locations may be learned from full images, or from a random subsample of pixels. For classification between more than two categories, we introduce a coupling parameter whose value tunes the number of sensors selected, trading accuracy for economy. Example datasets include those from image recognition using PCA for feature extraction and LDA for discrimination.

crimination; however, the system and method can be broadly applied to non-image data types and is easily adapted to work with other methods for feature extraction and discrimination.

Published:

— *with international search report (Art. 21(3))*

Systems, Devices, and Methods for Classification and Sensor Identification Using Enhanced Sparsity

SUMMARY

In an aspect, the present disclosure is directed to, among other things, an object
5 classification system. In an embodiment, the object classification system includes a
significant pixel identification module including circuitry for determining significant
pixels from one or more reference images. In an embodiment, the object classification
system includes a significant pixel identification module including circuitry for generating
optimal sparse pixel set information. In an embodiment, the object classification system
10 includes an object classification module including circuitry for comparing one or more
pixels of a first image indicative of one or more objects imaged in the first image to the
optimal sparse pixel set information. In an embodiment, the object classification system
includes an object classification module including circuitry for classifying at least one
object in the first image based on the comparison.

15 In an aspect, the present disclosure is directed to, among other things, an object
classification system including circuitry for determining an optimal sparse pixel set from
one or more reference images. In an embodiment, the object classification system
includes circuitry for classifying an object in at least one image based on a comparison of
one or more parameters associated with the optimal sparse pixel set to one or more pixels
20 associated with the at least one image.

In an aspect, the present disclosure is directed to, among other things, a method for
classifying images. In an embodiment, the method for classifying images includes
determining optimal sparse pixel set information from one or more reference images. In
an embodiment, the method for classifying images includes comparing one or more
25 parameters associated with the optimal sparse pixel set information to one or more pixels
associated with an image. In an embodiment, the method for classifying images includes
classifying at least one object in an image based on the comparison.

In an aspect, the present disclosure is directed to, among other things, an object
classification system including a significant pixel determination module, an optimal sparse
30 pixel set information generation module, and an object classification module.

In an aspect, the present disclosure is directed to, among other things, an object
classification system including a significant pixel identification module and an object

classification module. In an embodiment, the object classification module including circuitry for comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to optimal sparse pixel set information, and classifying at least one object in the first image based on the comparison.

5 In an aspect, the present disclosure is directed to, among other things, a system including a significant sensors identification module including circuitry for determining significant sensors associated with a sensor network. In an embodiment, the system includes an optimal sparse sensor module including circuitry for generating optimal sparse sensor set information from the significant sensors. In an embodiment, the system
10 includes a target decision task module including circuitry for classifying information associated with a target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.

 In an aspect, the present disclosure is directed to, among other things, a system for
15 classifying grid behavior. In an embodiment, system for classifying grid behavior includes circuitry for determining significant sensors from a sensor network. In an embodiment, the system for classifying grid behavior includes circuitry for generating optimal sparse sensor set information. In an embodiment, the system for classifying grid behavior includes circuitry for generating optimal sparse sensor set information from the
20 significant sensors responsive to one or more target decision task inputs. In an embodiment, the system for classifying grid behavior includes circuitry for classifying information associated with a target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.

25 In an aspect, the present disclosure is directed to, among other things, a method including determining significant sensors from a sensor network. In an embodiment, the method includes generating optimal sparse sensor set information. In an embodiment, the method includes generating optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs. In an embodiment, the
30 method includes classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.

In an aspect, the present disclosure is directed to, among other things, a system including a significant sensors determination from a sensor network module. In an embodiment, the system includes an optimal sparse sensor set information generation module. In an embodiment, the system includes a target decision task classification module.

In an aspect, the present disclosure is directed to, among other things, a system including a significant sensors identification module and an optimal sparse sensor module. In an embodiment, the optimal sparse sensor module includes circuitry for generating optimal sparse sensor set information from the significant sensors. In an embodiment, the system includes a target decision task module.

The foregoing summary is illustrative only and is not intended to be in any way limiting. In addition to the illustrative aspects, embodiments, and features described above, further aspects, embodiments, and features will become apparent by reference to the drawings and the following detailed description.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a schematic diagram of an extension of image classification procedure to a class of datasets.

Figure 2 depicts a visualization of the Fisherface and how sparse sensors approximate its dominant features.

Figure 3 shows a cross-validation of classification accuracy between images of cats and dogs.

Figure 4 illustrates mean sensor locations averaged over 400 random learning iterations.

Figure 5 shows classification accuracy of sensors.

Figure 6a depicts classification of 3 faces as coupling weight varies.

Figure 6b illustrates classification of 4 faces as coupling weight varies.

Figure 6c shows classification of 5 faces as coupling weight varies.

Figure 7 depicts increasing the coupling weight.

Figure 8 illustrates means sensor locations to discriminate between 3 faces.

Figure 9 is a schematic diagram of a system according one embodiment.

Figure 10a depicts examples of cats and dogs images used in the dataset.

Figure 10b illustrates the four PCA modes with the largest singular values of the full image dataset.

Figure 11a shows examples of images from the Yale Faces Database B.

Figure 11b depicts the six PCA modes with the largest singular values of the full
5 image dataset.

Figure 12 outlines exemplary sensor networks, pseudo-static sensor networks, grid systems, or the like that may benefit from the technologies and methodologies described herein.

Figure 13 shows a flow diagram of a method for classifying images according to
10 one embodiment.

Figure 14 shows a flow diagram of a method according to one embodiment.

DETAILED DESCRIPTION

In the following detailed description, reference is made to the accompanying drawings, which form a part hereof. In the drawings, similar symbols typically identify
15 similar components, unless context dictates otherwise. The illustrative embodiments described in the detailed description, drawings, and claims are not meant to be limiting. Other embodiments may be utilized, and other changes may be made, without departing from the spirit or scope of the subject matter presented here.

I. Introduction

20 Classification and detection based on measurements is a crucial capability for intelligent systems, both biological and engineered. It is desirable to perform such categorical tasks with limited or incomplete information, as the full state of the system may be inaccessible or expensive to measure. Fortunately, even when full measurement space is quite high-dimensional, the desired information extracted from the signal is often
25 inherently low-dimensional. For example, images may have a large number of pixels, but it is observed that natural images occupy a minuscule fraction of pixel space. Such an inherent sparsity in an appropriate transformed basis is the foundation of image compression (see M Elad, Sparse and redundant representations: from theory to applications in signal and image processing. Springer, New York, 2010, which is
30 incorporated by reference herein); natural images may be compressed in generic Fourier or wavelet bases.

The theory of compressive sensing (see: E. J. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Transactions on Information Theory*, 52(2):489–509, 2006; D. L. Donoho, Compressed sensing, *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006; R. G. Baraniuk, Compressive sensing, *IEEE Signal Processing Magazine*, 24(4):118–120, 2007; and J. A. Tropp and A. C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007, which are each incorporated by reference herein) takes further advantage of a signal's compressibility and demonstrates how a n -dimensional signal that is k -sparse in a transformed basis may be reconstructed exactly from $\mathcal{O}(k \log(n/k))$ measurements. Here, k -sparse means all but k coefficients are zero, and we assume $k \ll n$. The signal may be reconstructed from the known basis as an ℓ_0 sparse set of coefficients that best recover the measurements. The implementation of compressive sensing rests on the fact that the ℓ_0 sparse solution to an underdetermined linear system may almost certainly be found by relaxation to a convex ℓ_1 minimization (see: E. J. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Transactions on Information Theory*, 52(2):489–509, 2006; D. L. Donoho, Compressed sensing, *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006, which are each incorporated by reference herein), or alternatively, by a greedy algorithm (see J. A. Tropp and A. C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007, which are each incorporated by reference herein).

The compressive sensing strategy relies on the measurements being incoherent with respect to the known basis, so that measurement vectors are uncorrelated with basis directions. Incoherence holds between many pairs of bases, such as between delta functions and the Fourier basis. Therefore, it is possible to reconstruct images, which are sparse in the Fourier domain, from single pixel measurements, which may be viewed as discrete spatial delta functions. Surprisingly, a random Gaussian or Bernoulli matrix is incoherent with respect to any arbitrary basis with high probability (see: E. J. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Transactions on Information Theory*, 52(2):489–509, 2006; D. L. Donoho, Compressed sensing, *IEEE Transactions on*

Information Theory, 52(4):1289–1306, 2006, which are each incorporated by reference herein).

Compressive sensing is able to reconstruct a signal using surprisingly few measurements, but assigning a signal to one of a few categories may be accomplished with
 5 orders-of-magnitude fewer measurements (see for instance J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009, which is incorporated by reference herein). Typical natural images with n pixels can be recovered with significant savings, usually from $n/10$ to $n/3$ measurements (see J.
 10 Romberg, Imaging via compressive sampling, *Signal Processing Magazine, IEEE*, 25(2):14–20, 2008, which is incorporated by reference herein). For a 1-megapixel image, this means solving an ℓ_1 minimization with 100,000 constraints in a basis with one million vectors—a non-trivial computational task. In contrast, in the herein systems and methods, to classify a natural image and reduce it to its semantic content, only tens of measurement
 15 may be required.

Classification is often performed in a tailored low-dimensional basis extracted as hierarchical features of the data. One of the most ubiquitous methods in dimensionality reduction is principal components analysis (PCA) (see: K. Pearson, On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, 2(7–12):559–572, 1901;
 20 and R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, Wiley-Interscience, 2000, which are each incorporated by reference herein), which may be computed by a singular value decomposition (SVD) to yield an ordered orthonormal basis. Linear discriminant analysis (LDA) (see: C. M. Bishop, Pattern recognition and machine learning, Springer, New York, 2006; R. A. Fisher, The use of multiple measurements in
 25 taxonomic problems, *Annals of Eugenics*, 7(2):179–188, 1936; and C. R. Rao, The utilization of multiple measurements in problems of biological classification, *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):159–203, 1948, which are each incorporated by reference herein) is commonly applied in conjunction with PCA for discrimination tasks. Indeed, PCA-LDA is one of the classic approaches used to introduce
 30 machine learning methodologies (R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, Wiley-Interscience, 2000, which is incorporated by reference herein). High-variance PCA modes can account for common features shared across categories, which are not useful for discrimination. LDA produces a basis built with respect to the category

geometry to maximize between-class variance while minimizing within-class variance. PCA and LDA have both been used extensively for facial recognition (see for instance: L. Sirovich and M. Kirby. A low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519–524, 1987; M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991; D. L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 18(8):831–836, 1996; P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(7):711–720, 1997; H. Yu and J. Yang. A direct LDA algorithm for high-dimensional data– with application to face recognition. *Pattern Recognition*, 34:2067–2070, 2001; and A. M. Martinez and A. C. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(2):228–233, 2001, which are each incorporated by reference herein) and is a common benchmark against which novel classification techniques are compared.

Compared to signal reconstruction, classification is particularly efficient because of 1) the use of a tailored low-dimensional feature basis, and 2) the simplicity of deciding on a category rather than reconstructing exact details (J. N. Kutz, *Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data*. Oxford University Press, 2013, which is incorporated by reference herein).

A. Previous work on enhanced sparsity for classification

Combining ideas from compressive sensing with classification, there has been significant work relevant to exploring enhanced sparsity for classification. For example, the sparse representation algorithm has been applied to biometric recognition problems and demonstrated to be surprisingly robust (see J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma,. Robust face recognition via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009 and J. K. Pillai, V. M Patel, R. Chellappa, and N. K Ratha, Secure and robust iris recognition using random projections and sparse representations, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(9):1877–1893, 2011, which are each incorporated by reference herein). Sparse approximation (see: J. A. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004; and J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, Robust face recognition

via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009, which are each incorporated by reference herein) applied to semantic hierarchies (see M. Marszalek and C. Schmid, Semantic hierarchies for visual object recognition, In *Computer Vision and Pattern Recognition, 2007, CVPR'07. IEEE Conference on*, pages 1–7. IEEE, 2007, which is incorporated by reference herein) has been shown to be efficient in categorizing between large numbers of classes (see B. Kim, J. Y. Park, A. Mohan, A. C. Gilbert, and S. Savarese, Hierarchical classification of images by sparse approximation. In *BMVC*, pages 1–11. Citeseer, 2011, which is incorporated by reference herein). In another line of work, ideas from compressive sensing have been applied to dimensionality reduction tools to develop a theory of sketched SVD based on randomly projected data (see: J. E. Fowler. Compressive-projection principal component analysis, *IEEE Transactions on Image Processing*, 18(10):2230–2242, 2009; H. Qi and S. M. Hughes, Invariance of principal components under low-dimensional random projection of the data, *IEEE International Conference on Image Processing*, October 2012; and A. C. Gilbert, J. Y. Park, and M. B. Wakin, Sketched SVD: Recovering spectral features from compressive measurements, *ArXiv e-prints*, 2012, which are each incorporated by reference herein).

Of particular relevance to our work is the sparse representation for classification (SRC) algorithm as originally proposed for face recognition by Wright *et al.* (see J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009, which is incorporated by reference herein). In SRC, ℓ_1 minimization is used to find the sparsest representation of a test image in a dictionary composed of the training images. Each test image is then classified based on the category of dictionary elements whose sparse coefficients best reconstructs the test image, minimizing the residual in an ℓ_2 sense. Notably, in this framework, the classification is robust even using dictionaries of tens or hundreds of random projections measurements, due to the enhanced sparsity of classification.

These existing algorithms have all relied on random measurements, typically using dense or sparse random measurement matrices (A. C. Gilbert and P. Indyk, Sparse recovery using sparse matrices, *Proceedings of the IEEE*, 5=98(6):937–947, 2010, which is incorporated by reference herein). It remains unexplored how a refinement process may select an optimal subset of measurements to accomplish the classification, further

enhancing sparsity. Moreover, acquiring randomly projected measurements may be impractical where many individual point measurements are prohibitively expensive (e.g. ocean or atmospheric sensors). How can classification be accomplished with very few point measurements, and would the locations of those measurements illuminate coherent features of the data?

B. Contributions and perspectives herein

Methods and systems described herein include a novel framework to harness the enhanced sparsity in image recognition to classify images based on very few pixel measurements. Its distinct contribution consists of optimally selecting, within a large set of measurement locations, a smaller subset of key locations that serve the classification. Classification using very few learned pixel sensors with the systems and methods performs comparably with using the full image. Further, the algorithm has a parameter to tune the trade-off between fewer sensors and accuracy.

Datasets may be vast and data acquisition can be limited by bandwidth on data streams. Therefore, we also develop an intermediate technique, related to the sketched SVD (see: J. E. Fowler, Compressive-projection principal component analysis, *IEEE Transactions on Image Processing*, 18(10):2230–2242, 2009; A. C. Gilbert and P. Indyk, Sparse recovery using sparse matrices, *Proceedings of the IEEE*, 5=98(6):937–947, 2010; and A. C. Gilbert, J. Y. Park, and M. B. Wakin. Sketched SVD: Recovering spectral features from compressive measurements. *ArXiv e-prints*, 2012, which are each incorporated by reference herein), to start instead with a subsample of the original data. Starting with 10% of the original data, methods and systems described herein are still able to find *nearly* optimal sparse sensor locations.

In generic applications, this entire theory is identical if single pixels are replaced with random projection measurements. In the case of images, the use of single pixel measurements has a number of practical advantages, since pixels are the basic unit of measurement in images. Note that sampling with single-pixel measurements is effective when the image features are non-localized, so that measurements and basis are incoherent. Importantly, sparse sensor pixels identified by our algorithm cluster near coherent features in the images.

Our framework for sparse classification has two characteristics that distinguish it significantly from previous work: 1) instead of random measurements, spatial sensor locations are specifically selected; and 2) ℓ_1 minimization is applied *once* in sensor

learning, and classification of new images involves dot products involving a very small measurement vector. A schematic representation of the methods and systems is found in Fig. 1 and associated text. Equations (7) and (9) comprise primary theoretical contributions; Figs. 2 and 7 illustrate the sensor locations produced by the algorithm.

5 An engineering perspective of probing complex systems with underlying low-dimensional structure is taken herein, with the explicit goal of forming some categorical decision about the state of the system. The optimally sparse spatial sensors framework is particularly well suited for engineering applications, as an upfront investment in the learning phase allows remarkably efficient performance for all subsequent queries.

10 Abstractions of this method to more general data sets is discussed in more detail herein.

 Without being bound by theory, the principle of enhanced sparsity may also have relevance for biological organisms, which often need to make decisions based on very limited sensory information. Specifically, organisms interact with high-dimensional physical systems in nature but must rely on information gathered through a handful of
15 sensory organs. The methods and systems including the sparse sensor algorithm provides one approach to answer the question, Given a fixed budget of sensors, where should they be placed to optimally inform decision-making?

C. Organization

 The section below provides an overview of well-known techniques on which
20 contributions were built and establishes the notation used for the remainder herein. Summarized are compressive sensing, the PCA-LDA method for categorization, and how random projections may be used to form the sketched SVD. Following is described our algorithm for determining optimally sparse sensor locations by using ℓ_1 minimization and a dictionary of learned features. Sensor learning was applied to two image discrimination
25 tasks based on real-world data; the results of these experiments are presented in the Examples and the methods are discussed more generally herein.

II. Framework

 Methods and systems include 1) design of optimal sensor locations that take advantage of enhanced sparsity in classification, and 2) to design sensors from
30 substantially subsampled data. To put our contributions in context, below is a review from compressive sensing, which describes the theory of ℓ_1 reconstruction, sparse random projection matrices, and the sketched SVD.

Methods and systems make use of two well-established methods for systematically producing low-rank representations of a high-dimensional data set, principal components analysis (PCA) and linear discriminant analysis (LDA). Examples below demonstrate our algorithm by applying PCA in combination with LDA for face recognition. In general, the method can be implemented with many other dimensionality-reduction and discrimination algorithms, as drawn schematically in the top of Fig. 1.

Fig. 1 illustrates a schematic of the classification procedure based on full (top), randomly subsampled (middle), and sparsely sensed (bottom) data. Data measurements (left) are projected onto a PCA feature space (middle). LDA defines the projection $\mathbf{w}^T$ from PCA space to $\mathbb{R}^{c-1}$ (right), where classification occurs. The $\tilde{\Phi}$ matrix is a random projection that sub-samples the full image; the $\hat{\Phi}_1$ matrix is a learned projection that samples the full image at specific sensors as described below. Sparse sensors can also be learned from the randomly sub-sampled image, resulting in $\hat{\Phi}_2\tilde{\Phi}$, which is not the same as, but may approximate, $\hat{\Phi}_1$. The number of sensors q is at most $\bar{q}$, which is bounded by $r \leq \bar{q} \leq r(c-1)$; for two-way classification ($c = 2$), $\bar{q} = r$.

A. Compressive sensing and sparse representation

Compressive sensing theory states that if the information of a signal $x \in \mathbb{R}^n$ is k -sparse in a transformed basis Ψ (all but k coefficients are zero), it is possible to reconstruct the signal from very limited measurements $O(k \log(n/k))$ (see: E. J. Candès. Compressive sensing. *Proceedings of the International Congress of Mathematics*, 2006; E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006; E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, pages 21–30, 2008; and E. J. Candès, Y. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31:59–73, 2010, which are each incorporated by reference herein). This technique has widespread applications, including the fields of medical imaging (M. Lustig, D.L. Donoho, J.M. Santos, and J.M. Pauly. Compressed sensing mri: A look at how cs can improve on current imaging techniques. *IEEE Signal Processing Magazine*, 25(2):72–82, 2008, which is incorporated by reference herein), neuroscience (S. Ganguli and H. Sompolinsky. Compressed sensing, sparsity, and dimensionality in neuronal information processing and

data analysis. *Annual Review of Neuroscience*, 35:485–508, 2012, which is incorporated by reference herein), and engineering (S. Ganguli and H. Sompolinsky. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual Review of Neuroscience*, 35:485–508, 2012, which is incorporated by reference
 5 herein).

The standard framing of the compressive sensing problem is as follows. Let us suppose the measurement vector $\tilde{\mathbf{x}} \in \mathbb{R}^p$ is related to the desired signal $\mathbf{x} \in \mathbb{R}^n$ by $\tilde{\Phi}$, a known $p \times n$ measurement matrix: $\tilde{\mathbf{x}} = \tilde{\Phi}\mathbf{x}$. Compressive sensing seeks to recover $\mathbf{x}$ from $\tilde{\mathbf{x}}$ and applies to the underdetermined case where $p \ll n$.

10 The signal $\mathbf{x}$ has coefficients $\mathbf{a}$ in basis Ψ , so that

$$\tilde{\mathbf{x}} = \tilde{\Phi}\mathbf{x} = \Phi\Psi\mathbf{a} = \Theta\mathbf{a} \quad (1)$$

If $\mathbf{x}$ is sufficiently sparse in Ψ and the matrix Θ obeys the restricted isometry principle (RIP) (E. J. Candes and T. Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005, which is incorporated by
 15 reference herein), the search for $\mathbf{a}$ (and thus the reconstruction of $\mathbf{x}$) is possible.

We would like to solve for the sparsest $\mathbf{a}$,

$$\mathbf{a} = \underset{\mathbf{a}'}{\operatorname{argmin}} \|\mathbf{a}'\|_0, \text{ subject to } \tilde{\mathbf{x}} = \Theta\mathbf{a}', \quad (2)$$

but this involves an intractable combinatorial search. It has been shown that under certain
 20 conditions (related to the RIP), Eq. (2) may be solved by minimizing the ℓ_1 norm (E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006, D. L. Donoho. For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution.
 25 *Communications on pure and applied mathematics*, 59(6):797–829, 2006, which are each incorporated by reference herein),

$$\mathbf{a} = \underset{\mathbf{a}'}{\operatorname{argmin}} \|\mathbf{a}'\|_1, \text{ subject to } \tilde{\mathbf{x}} = \Theta\mathbf{a}', \quad (3)$$

Relaxation to a convex ℓ_1 minimization bypasses the combinatorially difficult
 30 problem. Solutions to Eq. (3) may be found through standard convex optimization routines, or by greedy algorithms such as orthogonal matching pursuit (J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006, J. A. Tropp and A. C. Gilbert.

Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007, which are each incorporated by reference herein).

An important aspect of compressive sensing is the choice of the measurement matrix $\tilde{\Phi}$. To achieve exact reconstruction, $\tilde{\Phi}$ must be incoherent with respect to Ψ (therefore satisfying the RIP). Many authors describe the properties of random matrix ensembles that fulfill these conditions with high probability: they include Gaussian, Bernoulli, and random partial Fourier matrices (E. J. Candès. Compressive sensing. *Proceedings of the International Congress of Mathematics*, 2006, E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006, D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006, which are each incorporated by reference herein). In addition, sparse random matrices, where many of the entries of the measurement vectors are zero, also allow reconstruction (as reviewed in A. C. Gilbert and P. Indyk. Sparse recovery using sparse matrices. *Proceedings of the IEEE*, 5=98(6):937–947, 2010, which is incorporated by reference herein).

In addition to the random measurement framework for reconstruction, we are concerned with the related question of the spectral properties of a data matrix under random projection. Consider a data matrix $X = [x_1 \ x_2 \ \dots \ x_m] \in \mathbb{R}^{n \times m}$. Let

$$\tilde{X} = \tilde{\Phi} X,$$

where $\tilde{\Phi} \in \mathbb{R}^{p \times n}$ and $p \ll n$, as above. Under what conditions are the spectral properties (i.e., the singular values and vectors) of $\tilde{X}$ and X approximately equal? Gilbert et al. (see A. C. Gilbert, J. Y. Park, and M. B. Wakin. Sketched SVD: Recovering spectral features from compressive measurements. *ArXiv e-prints*, 2012, which is incorporated by reference herein) found that spectral properties are approximately equal when the random measurement matrix $\tilde{\Phi}$ satisfies the distributed Johnson-Lindenstrauss (JL) property. This property regards the preservation of relative distances between points after random projection to a new space. Since the classification task depends on a geometric separation of data points between classes, the JL property is integral to the success of this model reduction step (dashed arrow between the full and subsampled PCA feature spaces in Fig. 1).

A related line of work on sparse representation for classification (SRC) (J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009, which is incorporated by reference herein) proposes to leverage
 5 sparsity promotion by ℓ_1 minimization for image recognition. In this formulation of the classification problem, each training image makes up a column of the dictionary Θ , and the test image $\mathbf{y}$ is represented as a linear combination of the dictionary elements weighted by coefficients $\mathbf{a}$. Thus posed, the solution to the following equation is the sparsest representation of the test image given the dictionary,

$$10 \quad \mathbf{a} = \underset{\mathbf{a}'}{\operatorname{argmin}} \|\mathbf{a}'\|_1, \text{ subject to } \mathbf{y} = \Theta \mathbf{a}.$$

Note that the above equation is identical to Eq. (3).

It is then possible to construct a sparse approximation of $\mathbf{y}$ using only coefficients in $\mathbf{a}$ associated with the i th class,

$$15 \quad \tilde{\mathbf{y}}_i = \Theta \delta_i(\mathbf{a}_i),$$

where $\delta_i(\mathbf{a})$ is a vector the same size as $\mathbf{a}$ whose only non-zero entries are the entries in $\mathbf{a}$ associated with training images in class i . Finally, $\mathbf{y}$ is assigned a category based on the class whose approximation minimizes the residual between $\mathbf{y}$ and $\tilde{\mathbf{y}}_i$:

$$20 \quad \text{category}(\mathbf{y}) = \underset{i}{\operatorname{argmin}} \|\mathbf{y} - \Theta \delta_i(\mathbf{a}_i)\|_2$$

The SRC framework exploits enhanced sparsity for classification, making use of the sparse subspace structure of face images. It has been demonstrated to work exceptionally well for many possible dictionaries Θ , including common features such as eigenfaces and Fisher faces, and uncommon features including downsampled faces and
 25 random projections.

It is worth noting that the performance of SRC for face recognition is achieved at the cost of one ℓ_1 minimization for each test image. This is distinct from the methods and systems including algorithm discussed herein, which uses ℓ_1 minimization to identify sparse sensor locations in the training phase only.

30 *B. PCA, LDA, and classification*

Two main dimensionality reduction techniques are used to demonstrate the pixel refinement algorithm discussed below. Consider a set of observations $\mathbf{x}_i$, $i = 1; \dots; m$, where $\mathbf{x}_i \in \mathbb{R}^n$; also suppose that $m \ll n$. Let us construct a data matrix $\mathbf{X}$ with columns $\mathbf{x}_i$,

as represented schematically in Fig. 1; we assume that rows of X have been mean subtracted. The PCA consists of performing a singular value decomposition (SVD) defined by the following modal decomposition:

$$X = \Psi \Sigma V^*. \quad (4)$$

5 The columns of Ψ are the left singular vectors of X ; they span the columns of X and are often referred to as the *principal components* or *features* of the data set. The matrix Σ is diagonal whose entries δ_{ii} are the *singular values* of X ; there are only m non-zero singular values, and we may write $\Sigma = [\Sigma_{m \times m} \ 0]^T$.

The columns of the matrix Ψ are eigenvectors of XX^T :

$$10 \quad XX^T \Psi = \Psi \Lambda, \quad \Lambda = \begin{bmatrix} \Sigma^2_{m \times m} & 0 \\ 0 & 0 \end{bmatrix}.$$

However, this is an extremely expensive eigendecomposition, since the dimension of XX^T is $n \times n$. In implementation, it is more practical to use the method of snapshots (L.

Sirovich. Turbulence and the dynamics of coherent structures, parts I-III. *Q. Appl. Math.*, XLV(3):561–590, 1987, which is incorporated by reference herein), whereby we solve the

15 $m \times m$ eigendecomposition for V :

$$X^T X V = V \Sigma^2_{m \times m}.$$

We may then obtain the m non-zero PCA modes by taking a linear combination of the columns of X :

$$\Psi_m = X V \Sigma_{m \times m}^{-1}.$$

20 The SVD provides a systematic approach to transform data into a lower-dimensional representation. For data sets that are inherently low-rank, the singular values Σ of Eq. (4), when ordered by magnitude, will contain many entries that are exactly zero, providing evidence for an exact low-rank representation. With most realistic applications and data sets, though, the singular values often exhibit a power-law decrease, where the

25 low-rank representation of the data is approximate. The singular values exhibiting a power-law decrease provides the opportunity for a heuristic (at least quantitatively informed) decision about the number of information-rich dimensions to retain for the reduced-order subspace. Taking the r columns of Ψ as basis vectors corresponding to the r largest singular values, a truncated basis Ψ_r can be formed. We may use this basis to

30 project data from the full measurement space into the reduced r -dimensional PCA space (known as *feature space*):

$$\Psi_r^T : \mathbb{R}^n \rightarrow \mathbb{R}^r, \quad x \mapsto a. \quad (5)$$

Thus, the principal components Ψ_r minimize the ℓ_2 projection error, $\|x - \Psi_r \Psi_r^T x\|_2$.

For categorical decisions, we apply the well known classifier linear discriminant analysis (LDA) in r -dimensional feature space. Despite the plethora of classifier techniques, we chose LDA for its simplicity and well-established behavior. We note once
5 again that PCA and LDA are serving only to illustrate the concept proposed in Fig. 1.

We start by noting that each observation x_i (and thus each $a_i = \Psi_r^T x_i$) belongs to one of c distinct categories C_j where $j = 1; \dots; c$. LDA attempts to find a set of directions in feature space $w \in \mathbb{R}^{r \times c-1}$ where the between-class variance is maximized and within-class variance is minimized. Specifically,

$$10 \quad w = \underset{w'}{\operatorname{argmax}} \frac{w'^T S_B w'}{w'^T S_W w'}. \quad (6)$$

Here S_W is the within-class scatter matrix,

$$S_W = \sum_{j=1}^c \sum_{i \in C_j} (a_i - \mu_j)(a_i - \mu_j)^T,$$

where μ_j is the mean of class j . S_B is the between-class scatter matrix,

$$15 \quad S_B = \sum_{j=1}^c N_j (\mu_j - \mu)(\mu_j - \mu)^T$$

where N_j is the number of observations in class j and μ is the mean of all observations in $A = \Psi_r^T X$ (in this case $\mu = 0$). LDA may also be performed on the full data X .

It follows that w are eigenvectors corresponding to the nonzero eigenvalues of $S_W^{-1} S_B$. It should be noted that the number of non-trivial LDA directions must be less than
20 or equal to the number of categories minus one ($c - 1$), and that at least $r + c$ samples are needed to guarantee S_W is not singular (C. M. Bishop. *Pattern recognition and machine learning*. Springer, New York, 2006, which is incorporated by reference herein).

In schematic form, the top “full image” row of Fig. 1 illustrates the PCA-LDA process for the classification task. In the case of $c = 2$ categories, the LDA projection is
25 $w^T : \mathbb{R}^r \rightarrow \mathbb{R}$, and we may apply a threshold value that separates the two categories. Assuming the two categories have equal covariances, we pick the threshold to be the midpoint between the means of the two classes, $w^T \mu_A$ and $w^T \mu_B$.

For decisions between $c \geq 2$ categories, LDA produces a projection $w^T : \mathbb{R}^r \rightarrow \mathbb{R}^{c-1}$. We use a nearest centroid (NC) method for classification, in which a new
30 measurement x_i is assigned to category j for which the distance between $w^T x_i$ and $w^T \mu_j$ is minimal. Many other classifier choices are possible in the reduced $\mathbb{R}^{c-1}$ space, including nearest neighbor (NN) (R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*.

Wiley-Interscience, 2000, which is incorporated by reference herein), nearest subspace (NS) (K. Lee, J. Ho, and D. J Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans Pattern Anal Mach Intell*, 27(5):684–98, May 2005, which is incorporated by reference herein), support vector machines (SVM) (A. Muñoz and J.M. Moguerza. Estimation of high-density regions using one-class neighbor machines. *IEEE Trans Pattern Anal Mach Intell*, 28(3):476–80, Mar 2006, which is incorporated by reference herein), and sparse representation (SRC) (J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009, which is incorporated by reference herein). While it is possible these classifiers may improve the accuracy of the decision, attention is restricted to NC since the focus is not on the specific decision algorithm, but rather on the method of learning sensor locations.

III. Obtaining Sparse Sensor Locations

The method and system to learn sparse sensor locations is described below, building upon the theory described above. We begin by addressing classification between $c = 2$ categories, reducing a full measurement vector to a small handful of key measurements. Also described below is an intermediate technique that starts instead with a random subsample of the full data. Next is developed an extension of the algorithm to classification between $c \geq 2$ categories. Also introduced is a coupling weight λ , which allows us to decrease the number of sensors required at the cost of slightly lower classification accuracy. Finally, below is demonstrated the possibility of transforming the full-dimensional feature/discrimination projection into an approximate projection from the learned sparse measurements to the decision space. Alternatively, it is possible to re-compute a new projection to decision space directly from the data matrix comprised of sparse measurements, as in the last row of Fig. 1. In the following, we refer to *features* Ψ_r (e.g., PCA modes) and *discrimination vectors* w (e.g., LDA directions) to emphasize the generality of the method and system.

A. *Deciding between two categories*

Let us first consider learning sparse sensors for a classification problem between $c = 2$ categories. The discrimination vector w encodes the direction in Ψ_r that is most informative in discriminating between categories of observations given by X . We seek a measurement vector s that satisfies $\Psi_r^T s = w$.

In particular, we seek the sparse solution $\mathbf{s}$: to find the measurements that best reconstruct $\mathbf{w}$, our approach is to solve for

$$\mathbf{s} = \operatorname{argmin} \|\mathbf{s}^l\|_1, \text{ subject to } \Psi_r^T \mathbf{s}^l = \mathbf{w}. \quad (7)$$

The equations for $\mathbf{s}$ are under constrained, so that we may use ℓ_1 minimization to
 5 find the sparse solution with at most r nonzero elements. It is important to remember that $\mathbf{s} \in \mathbb{R}^n$ and can be visualized as an image where most of the pixels are zero. The r non-zero elements of $\mathbf{s}$ represent locations of sensors that best recapture the discriminant projection vector $\Psi_r \mathbf{w}$.

To motivate this approach, consider that we have constructed a projection Ψ_r^T onto
 10 a set of features and a projection $\mathbf{w}^T$ from feature space to classification space; we may simplify this into a single projection:

$$\begin{aligned} \eta &= \mathbf{w}^T \Psi_r^T \mathbf{x}_{\text{test}} \\ &= \langle \mathbf{X}, \mathbf{x}_{\text{test}} \rangle \end{aligned}$$

where $\mathbf{X} \triangleq \Psi_r \mathbf{w}$. Therefore, Eq. (7) finds the sparse measurement vector $\mathbf{s}$ that projects to
 15 the same feature space coordinates as $\mathbf{X}$:

$$\begin{aligned} \Psi_r^T \mathbf{s} &= \Psi_r^T \mathbf{X} \\ &= \underbrace{\Psi_r^T \Psi_r}_{\|r \times r\|} \mathbf{w} = \mathbf{w} \end{aligned}$$

Finally, we see that

$$\mathbf{s} = \mathbf{X} + \xi \quad (8)$$

20 where $\xi \in \ker(\Psi_r^T)$. That is to say, the sparse measurement $\mathbf{s}$ is the same as the vector $\mathbf{X}$ (e.g., a Fisherface) plus some residual vector that does not contain any dominant features (i.e., is not in the span of dominant features).

Figure 2 depicts a visualization of the Fisherface $\mathbf{X} \triangleq \Psi_r \mathbf{w}$ and how sparse
 25 sensors, as described herein, approximate its dominant features. (a) $\mathbf{X}$ (black) and sparse approximation $\mathbf{s}$ (red). (b) An alternative visualization of $\mathbf{X}$ (grey) and locations of sparse sensors (non-zero elements of $\mathbf{s}$, in red). See below for specific details of cat/dog image recognition.

To compare $\mathbf{X}$ and $\mathbf{s}$, Fig. 2 depicts a visualization of these vectors for the dog/cat
 30 face recognition problem discussed below. Figure 2 (a) shows the magnitude $\mathbf{X}$ along with the 20 pixel measurement vector $\mathbf{s}$ obtained by ℓ_1 minimization of Eq. (7). The image of $\mathbf{X}$ (grey) and the sparse pixels (red) are shown in Fig. 2 (b). Importantly, it is not

possible to obtain sparse pixel measurements by thresholding $\mathbf{X}$ (Fig. 2 (a)); rather, $\mathbf{s}$ is a sparse image that exactly projects to $\mathbf{w}$ in Ψ_r space.

To implement this optimal sub-sampling at learned sensors, we construct a $q \times n$ projection matrix $\hat{\Phi}_1$ that maps $\mathbf{x} \mapsto \hat{\mathbf{x}}$. The number of sensors q is at most r , although in
 5 practice q is usually equal to r . The matrix $\hat{\Phi}_1$ are rows of the $n \times n$ identity matrix corresponding to the non-zero elements of $\mathbf{s}$.

D. Learning from randomly subsampled data

The identical approach applies when starting from subsampled data $\tilde{\mathbf{X}}$ (the middle “sub-sampled” row of Fig. 1). We now solve for $\tilde{\mathbf{s}}$ that best reconstructs the vector $\tilde{\mathbf{w}}$ in a
 10 space defined by the columns of $\tilde{\Psi}_r$. $\tilde{\Phi}$ is then refined to construct $\hat{\Phi}_2$, the r rows of the $p \times p$ identity matrix corresponding to non-zero elements of $\tilde{\mathbf{s}}$.

Note that $\hat{\Phi}_1$ and $\hat{\Phi}_2 \tilde{\Phi}$ are both $r \times n$ matrices, where typically $r \ll p \ll n$. Although the two sub-sampling matrices are not the same, we show below that they are quite similar when $p/n \geq 0.1$.

15 Finally, having learned sparse sensor locations, we project the data into $\mathbb{R}^r$,

$$\hat{\mathbf{X}} = \hat{\Phi} \mathbf{X}$$

where $\hat{\Phi}$ can be either $\hat{\Phi}_1$ or $\hat{\Phi}_2 \tilde{\Phi}$, as illustrated in Fig. 1. Possible methods to project these learned sparse measurements to classification space are described below.

E. Deciding between more than two categories

20 The simplest extension to classification between $c \geq 2$ categories can be implemented by considering the projection $\mathbf{w}^T: \mathbb{R}^r \rightarrow \mathbb{R}^{c-1}$ to decision space, followed by independently solving Eq. (7) for each column of $\mathbf{w}$. However, this approach scales badly with c ; in general, discriminating c categories of data by projection into r -dimensional feature space results in at most $q = r(c-1)$ learned sensors locations.

25 An alternative formulation of the convex optimization problem solves for columns of $\mathbf{s} \in \mathbb{R}^{n \times (c-1)}$ simultaneously; each column of $\mathbf{s}$ (image) projects to a column of $\mathbf{w}$ (discrimination vector) in feature space. We introduce a norm that penalizes the total number of non-zero rows in $\mathbf{s}$ (pixel measurements) to reconstruct the $c - 1$ columns of $\mathbf{w}$. Specifically,

30

$$\begin{aligned} \mathbf{s} = \underset{\mathbf{s}^l}{\operatorname{argmin}} \{ & \|\mathbf{s}^l\| + \lambda \|\mathbf{s}^l \mathbf{v}\|_1 \}, \\ \text{subject to } & \|\Psi_r^T \mathbf{s}^l - \mathbf{w}\|_F \leq \varepsilon, \end{aligned} \quad (9)$$

where $\|M\|_1 = \sum_{ij} |m_{ij}|$, $\mathbf{v}$ is a column vector of $(c - 1)$ ones, $\|M\|_F = \sqrt{\sum_{ij} |m_{ij}|^2}$ is the Frobenius norm, and ϵ is a small error tolerance ($\epsilon \approx 10^{-10}$ for examples as below).

Once again, we have an underdetermined system of equations for $\mathbf{s}$. The value of the coupling weight λ determines the number of non-zero rows of $\mathbf{s}$, so that the number of sensors q identified is at most $\bar{q}$ where $r \leq \bar{q} \leq r(c - 1)$.

In the limit where $\lambda = 0$, the solution to Eq. (9) is the same as obtained by the uncoupled, independent approach. In general, solving Eq. 9 with $\lambda = 0$ would result in at most $r(c - 1)$ sensor locations. As λ becomes larger, the coupling between columns of $\mathbf{s}$ becomes stronger, and the same pixel location can be shared between columns of $\mathbf{s}$ to approximately reconstruct columns of $\mathbf{w}$. In other words, the same pixel measurement is re-used to capture multiple linear discriminant projection vectors. In the limit $\lambda \rightarrow \infty$, a minimum of r sensors are found as non-zero rows of $\mathbf{s}$.

The optimization problem as formulated in Eq. (9) is closely related to the one solved by Simultaneous Orthogonal Matching Pursuit (S-OMP, J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006, J. A. Tropp. Algorithms for simultaneous sparse approximation. part ii: Convex relaxation. *Signal Processing*, 86(3):589–602, 2006, which are each incorporated by reference herein). S-OMP is a greedy algorithm, so instead of a coupling weight parameter λ , one would decide on a stopping criterion (for example, the desired number of iterations/sensors).

Following the same approach as in deciding between two categories, we implement sub-sampling at learned sensors to construct a projection matrix $\hat{\Phi}$.

F. Projecting sparse measurements to classification space

In the above procedure, the projections from the high-dimensional space into feature space (Ψ_r^T) and then into decision space ($\mathbf{w}^T$) are computed as a one-time upfront cost. It is then possible to re-use these projections to obtain an induced projection into the decision space starting only from learned sparse measurements. The alternative is to re-compute the discrimination vectors on the sparse measurement data $\hat{\mathbf{X}}$.

Given a new image $\mathbf{x}$, we obtain a vector η of decision space coordinates as follows:

$$\eta = \mathbf{w}^T \Psi_r^T \mathbf{x}. \quad (10)$$

We then substitute $\mathbf{w} = \Psi_r^T \mathbf{s}$ and $\mathbf{x} \approx \hat{\Phi}^T \hat{\mathbf{x}}$, where $\hat{\mathbf{x}} = \hat{\Phi} \mathbf{x}$, into Eq. (10):

$$\tilde{\eta} = \mathbf{s}^T \Psi_r^T \Psi_r^T \hat{\Phi}^T \hat{\mathbf{x}}. \quad (11)$$

where $\tilde{\eta} \approx \eta$. Define $\mathbf{z}$ as the measurement projection of $\mathbf{s}$, so that $\mathbf{z} = \hat{\Phi}\mathbf{s}$ and $\mathbf{s} \approx \hat{\Phi}^T\mathbf{z}$.

Substituting into Eq. (11) yields:

$$\begin{aligned}\tilde{\eta} &= \mathbf{z}^T \hat{\Phi} \Psi_r \Psi_r^T \hat{\Phi}^T \hat{\mathbf{x}} \\ &= \mathbf{z}^T \mathbf{T} \hat{\mathbf{x}}.\end{aligned}$$

- 5 The matrix $\mathbf{T} = \hat{\Phi} \Psi_r \Psi_r^T \hat{\Phi}^T$ defines a new inner-product on the space of sparse measurements that preserves the geometry of decision space. To compute $\mathbf{T}$ efficiently, compute $\hat{\Phi} \Psi_r$ and $\Psi_r^T \hat{\Phi}^T$ separately and then multiply the results.

- Re-computing the discrimination projection on the sparse data $\hat{\mathbf{X}} = \hat{\Phi}\mathbf{X}$ is typically inexpensive because of the small number of rows. Additionally, for multi-way
10 classification, this usually leads to better performance, as is discussed in the Examples.

- Figure 9 shows a system in which one or more methodologies or technologies can be implemented such as, for example, determining optimal sparse set information, determining significant sensors from a sensor network, determining significant pixel identification classifying information associated with a target decision task, classifying
15 imaged objects, refining a large set of measurement locations to learn a much smaller subset of key locations that best serve a classification task, or the like. Figure 9 illustrates the extension of image classification procedure to a broader class of datasets. In this case, single-pixel measurements become individual sensors in a sensor network. In some embodiments, single-pixel measurements become individual sensors in a sensor network.

- 20 The system includes an algorithm that refines a large set of measurement locations to learn a much smaller subset of key locations to best serve a classification task. The algorithm exploits enhanced sparsity for classification, when reconstruction may be bypassed. Enhanced sparsity provides an orders-of-magnitude reduction in number of measurements required when compared with standard compressive sensing strategies.
25 Measurements are projected directly into decision space; reconstruction from so few measurements is neither possible or needed.

- Our algorithm leverages the sparsity promoting ℓ_1 minimization as phrased in Equations (7) or (9). These convex optimization problems solve for a sparse set of sensor locations to closely approximate the linear discrimination vector in PCA space. The
30 approach may be generalized to a variety of other dimensionality reduction and discrimination algorithms. In addition, once a set of sensors have been obtained, a number of more sophisticated classifier algorithms may be applied to these measurements to improve the classification accuracy. These steps are shown in panels (d) and (e) of Fig. 9.

The approach was demonstrated on two examples of image recognition (see Examples). Classifiers built on learned sensors approached the performance of classifiers built on projections to principal components of the full images. These optimal sensor locations could be learned approximately even from already randomly subsampled images.

5 In either case, the learned sensors performed significantly better than the matched number of randomly chosen sensors. Further, the ensemble of learned sensor locations clustered around coherent features of the images. It is possible to think of this ensemble as a pixel mask for the faces in the training set, which may be of use when applied to engineered or biological systems where the important features may not be salient by inspection.

10 It is plausible that biological organisms may exploit enhanced sparsity for classification. Organisms interact with the external world with motor outputs, which are often discrete trajectories at specific moments in time, so that the transformation from sensory inputs to motor outputs can be thought of as a classification task. For example, a fly has no need to reconstruct the full flow velocity field around its body—it has only to
15 decide what to do in response to a gust. Sensory organs and data processing by the nervous system can be expensive; therefore, it is advantageous to place a smaller number of sensors at key locations on the body.

These convex optimization problems solve for a sparse set of sensor locations to closely approximate the linear discrimination vector in PCA space. The approach may be
20 generalized to a variety of other dimensionality reduction and discrimination algorithms. In addition, once a set of sensors have been obtained, a number of more sophisticated classifier algorithms may be applied to these measurements to improve the classification accuracy. Examples of these steps are shown in panels (d) and (e) of Fig. 9.

Although introduced for image discrimination, the system and algorithm may be
25 applied to a variety of non-image data types with more general sensor networks; this is shown schematically in Fig. 9. One such application is the detection of disease spread in epidemiological monitoring. We also envision these methods applying to mobile sensor networks in oceanographic and atmospheric sampling, as well as to detect and monitor various network behaviors in the electrical grid, internet packet routing, and
30 transportation.

Figure 12 illustrates applications for some embodiments of the methods and systems described herein. In some embodiments, the systems and methods are applicable for use with grid technologies. For example, methods and systems described herein can be

applied to epidemiological monitoring, power electrical grid systems, internet-related systems, and grids relating to the transport of goods and services. In some embodiments, the systems and methods are applicable for use with medical diagnostics. For example, methods and systems described herein can be used as part of medical diagnostic systems
5 such as those used in genetics, epigenetics, metabolites and natural sensing for prosthetics. In some embodiments, the systems and methods are applicable for use with engineering diagnostics. For example, methods and systems described herein can be used in process control for manufacturing and acoustic tone engineering diagnostics. In some
10 embodiments, the systems and methods described herein are applicable for use with other technologies. For example, methods and systems described herein can be used with RGB-D images and questionnaire refinement.

In an embodiment, the system includes an object classification system. In an embodiment, the object classification system includes one or more modules. For example, in an embodiment, the object classification system includes a significant sensor module
15 operable to determine significant sensor information associated with a sensor network. In an embodiment, the object classification system includes a significant pixel identification module.

In an embodiment, a module includes, among other things, one or more computing devices such as a processor (*e.g.*, a microprocessor), a central processing unit (CPU), a
20 digital signal processor (DSP), an application-specific integrated circuit (ASIC), a field programmable gate array (FPGA), or the like, or any combinations thereof, and can include discrete digital or analog circuit elements or electronics, or combinations thereof. In an embodiment, a module includes one or more ASICs having a plurality of predefined logic components. In an embodiment, a module includes one or more FPGAs, each
25 having a plurality of programmable logic components.

In an embodiment, a module includes one or more components operably coupled (*e.g.*, communicatively, electromagnetically, magnetically, ultrasonically, optically, inductively, electrically, capacitively coupled, or the like) to each other. In an embodiment, a module includes one or more remotely located components. In an
30 embodiment, remotely located components are operably coupled, for example, *via* wireless communication. In an embodiment, remotely located components are operably coupled, for example, *via* one or more receivers, transmitters, transceivers, antennas, or

the like. In an embodiment, the drive control module includes a module having one or more routines, components, data structures, interfaces, and the like.

In an embodiment, a module includes memory that, for example, stores instructions or information. For example, in an embodiment, a control module includes
5 memory that stores optimal sparse pixel set information, near optimal sparse pixel set information, reference image information, training image information, protocol information, sensor set information, optimal sparse sensor set information, near optimal sparse sensor set information, optimal sparse mote location information, etc. Non-limiting examples of memory include volatile memory (*e.g.*, Random Access Memory (RAM),
10 Dynamic Random Access Memory (DRAM), or the like), non-volatile memory (*e.g.*, Read-Only Memory (ROM), Electrically Erasable Programmable Read-Only Memory (EEPROM), Compact Disc Read-Only Memory (CD-ROM), or the like), persistent memory, or the like. Further non-limiting examples of memory include Erasable Programmable Read-Only Memory (EPROM), flash memory, or the like. In an
15 embodiment, the memory is coupled to, for example, one or more computing devices by one or more instructions, information, or power buses. For example, in an embodiment, the significant pixel identification module includes memory that, for example, stores optimal sparse pixel set information, near optimal sparse pixel set information, reference image information, training image information, protocol information, or the like.

20 In an embodiment, a module includes one or more computer-readable media drives, interface sockets, Universal Serial Bus (USB) ports, memory card slots, or the like, and one or more input/output components such as, for example, a graphical user interface, a display, a keyboard, a keypad, a trackball, a joystick, a touch-screen, a mouse, a switch, a dial, or the like, and any other peripheral device. In an embodiment, a module includes
25 one or more user input/output components, user interfaces, or the like, that are operably coupled to at least one computing device configured to control (electrical, electromechanical, software-implemented, firmware-implemented, or other control, or combinations thereof) at least one parameter associated with, for example, controlling one or more of activating, operating, or the like, object classification system.

30 In an embodiment, a module includes a computer-readable media drive or memory slot that is configured to accept signal-bearing medium (*e.g.*, computer-readable memory media, computer-readable recording media, or the like). In an embodiment, a program for causing a system to execute any of the disclosed methods can be stored on, for example, a

computer-readable recording medium (CRMM), a signal-bearing medium, or the like. Non-limiting examples of signal-bearing media include a recordable type medium such as a magnetic tape, floppy disk, a hard disk drive, a Compact Disc (CD), a Digital Video Disk (DVD), Blu-Ray Disc, a digital tape, a computer memory, or the like, as well as
5 transmission type medium such as a digital or an analog communication medium (*e.g.*, a fiber optic cable, a waveguide, a wired communications link, a wireless communication link (*e.g.*, receiver, transmitter, transceiver, transmission logic, reception logic, etc.). Further non-limiting examples of signal-bearing media include, but are not limited to, DVD-ROM, DVD-RAM, DVD+RW, DVD-RW, DVD-R, DVD+R, CD-ROM, Super
10 Audio CD, CD-R, CD+R, CD+RW, CD-RW, Video Compact Discs, Super Video Discs, flash memory, magnetic tape, magneto-optic disk, MINIDISC, non-volatile memory card, EEPROM, optical disk, optical storage, RAM, ROM, system memory, web server, or the like.

In an embodiment, the significant pixel identification module includes circuitry for
15 determining significant pixels from one or more reference images. For example, in an embodiment, the significant pixel identification module includes circuitry for determining significant pixels from one or more training images.

In an embodiment, the significant pixel identification module includes circuitry for generating optimal sparse pixel set information. In an embodiment, the significant pixel
20 identification module includes circuitry for generating a near optimal sparse pixel set information. In an embodiment, the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based one or more learning protocols.

In an embodiment, the significant pixel identification module includes circuitry for
25 generating the optimal sparse pixel set information based on a modal decomposition protocol. In an embodiment, the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based on a convex optimization protocol. In an embodiment, the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based on an ℓ_1 -minimization
30 protocol. In an embodiment, the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based on a subsampling using a sparse random matrices protocol. In an embodiment, the significant pixel identification

module includes circuitry for generating the optimal sparse pixel set information based on a subsampling using a dense random matrices protocol.

In an embodiment, the significant pixel identification module includes circuitry for determining one or more discrimination vectors associated with the significant pixels
5 based on a discrimination analysis protocol.

In an embodiment, the object classification system includes object classification module. For example, in an embodiment, the object classification system includes object classification module including circuitry for comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal sparse pixel set
10 information. In an embodiment, the object classification module includes circuitry for classifying at least one object in the first image based on the comparison. In an embodiment, the object classification module includes circuitry for classifying the at least one object in the first image based on comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal sparse pixel set
15 information and the one or more discrimination vectors. In an embodiment, the object classification module includes circuitry for determining reference data clusters associated with the significant pixels in a sparse sensor space.

In an embodiment, the object classification system includes circuitry for classifying an object in at least one image. For example, in an embodiment, the object
20 classification system includes circuitry for classifying an object in at least one image based on a comparison of one or more parameters associated with the optimal sparse pixel set to one or more pixels associated with the at least one image. In an embodiment, the object classification module includes circuitry for comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal sparse pixel set
25 information and generating classification information associated with at least one object in the first image based on a learning protocol.

In an embodiment, an object classification system includes circuitry for determining an optimal sparse pixel set. For example, in an embodiment, the object classification system includes circuitry for determining an optimal sparse pixel set from
30 one or more reference images. In an embodiment, the circuitry for determining the optimal sparse pixel set includes circuitry for determining one or more discrimination vectors from the one or more reference images. In an embodiment, the circuitry for determining the optimal sparse pixel set includes circuitry for determining one or more

sparse pixel locations. In an embodiment, the circuitry for determining the optimal sparse pixel set includes circuitry for generating a feature space transformation from the one or more reference images. In an embodiment, the circuitry for determining the optimal sparse pixel set includes circuitry for determining one or more weighting factors

5 associated with the optimal sparse pixel set.

In an embodiment, the object classification system includes circuitry for classifying an object. For example, in an embodiment, the object classification system includes circuitry for classifying an object in at least one image based on a comparison of one or more parameters associated with the optimal sparse pixel set to one or more pixels
10 associated with the at least one image. In an embodiment, the circuitry for classifying the object in at least one image includes circuitry for generating one or more classification categories associated with the object.

In an embodiment, an object classification system includes a significant pixel determination from one or more reference images module. In an embodiment, the object
15 classification system includes an optimal sparse pixel set information generation module.

In an embodiment, the object classification system includes an object classification module. In an embodiment, the object classification module includes one or more first image pixels comparison to optimal sparse pixel set information, the one or more first image pixels indicative of at least one object in the first image. In an embodiment, the
20 object classification module includes a first image object classification module configured to classify at least one object in the first image based on a comparison of one or more pixels of the first image to optimal sparse pixel set information.

In an embodiment, an object classification system includes a significant pixel identification module and an object classification module. In an embodiment, the
25 significant pixel identification module includes circuitry for determining significant pixels from one or more reference images, and generating optimal sparse pixel set information. In an embodiment, the object classification module includes circuitry for comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal sparse pixel set information. In an embodiment, the object classification
30 module includes circuitry for classifying at least one object in the first image based on the comparison. In an embodiment, the object classification module includes circuitry for comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal sparse pixel set information. In an embodiment, the object

classification module includes circuitry for classifying at least one object in the first image based on the comparison.

In an embodiment, a system includes a significant sensors identification module. For example, in an embodiment, the system includes a significant sensors identification
5 module including circuitry for determining significant sensors associated with a sensor network. In an embodiment, the significant sensors identification module includes circuitry for determining optimal sparse sensor location information associated with the sensor network. In an embodiment, the significant sensors identification module includes circuitry for determining nearly optimal sparse sensor location information associated with
10 the sensor network. In an embodiment, the significant sensors identification module includes circuitry for determining optimal sparse sensor location information associated with the sensor network based on at least one ℓ_1 -minimization protocol.

In an embodiment, the significant sensors identification module includes circuitry for determining significant nodes associated with a sensor network. In an embodiment,
15 the significant sensors identification module includes circuitry for determining significant nodes associated with a sensor network, and wherein the optimal sparse sensor module includes circuitry for generating optimal sparse node set information from the significant nodes responsive to one or more target decision task inputs. In an embodiment, the significant sensors identification module includes circuitry for determining optimal sparse
20 node location information associated with a node network.

In an embodiment, the system includes an optimal sparse sensor module. For example, in an embodiment, the system includes an optimal sparse sensor module including circuitry for generating optimal sparse sensor set information from the significant sensors. In an embodiment, the optimal sparse sensor module includes
25 circuitry for generating optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs.

In an embodiment, the system includes a target decision task module. For example, in an embodiment, the system includes a target decision task module including circuitry for classifying information associated with the target decision task. In an
30 embodiment, the system includes a target decision task module including circuitry for classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.

In an embodiment, the system for classifying grid behavior includes circuitry for determining significant sensors from a sensor network.

In an embodiment, the system for classifying grid behavior includes circuitry for generating optimal sparse sensor set information from the significant sensors. For
5 example, in an embodiment, the system for classifying grid behavior includes circuitry for generating optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs. In an embodiment, the circuitry for generating the optimal sparse sensor set information includes circuitry for generating one or more
10 discrimination vectors associated with the sensor network. In an embodiment, the circuitry for generating the optimal sparse sensor set information includes circuitry for generating one or more sparse sensor locations associated with the sensor network.

In an embodiment, the system for classifying grid behavior includes circuitry for classifying information associated with the target decision task. For example, in an
15 embodiment, the system for classifying grid behavior includes circuitry for classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information. In an embodiment, the circuitry for classifying information associated with the target decision task includes circuitry for generating one or more
20 classification categories associated with the target decision task. In an embodiment, the circuitry for classifying information associated with the target decision task includes circuitry for generating one or more classification parameters associated with the target decision task.

In an embodiment, a system includes one or more modules. For example, in an
25 embodiment, the system includes a significant sensors determination from a sensor network module. In an embodiment, the system includes an optimal sparse sensor set information generation module. In an embodiment, the optimal sparse sensor set information generation module is configured to generate optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs. In an embodiment, the system includes a target decision task classification
30 module. In an embodiment, the target decision task classification module is configured to classify information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.

In an embodiment, a system includes a significant sensors identification module and an optimal sparse sensor module. In an embodiment, the significant sensors identification module includes circuitry for determining significant sensors associated with a sensor network. In an embodiment, the optimal sparse sensor module includes circuitry
 5 for generating optimal sparse sensor set information from the significant sensors. In an embodiment, the optimal sparse sensor module includes circuitry for generating optimal sparse sensor set information from the significant sensors. In an embodiment, the system includes a target decision task module. In an embodiment, the target decision task module includes circuitry for classifying information associated with the target decision task based
 10 on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.

Figure 13 shows a method **300** for classifying images. At **310**, the method **300** includes determining optimal sparse pixel set information from one or more reference images. At **312**, determining the optimal sparse pixel set information includes
 15 determining the optimal sparse pixel set information based on at least one recovery protocol. At **314**, determining the optimal sparse pixel set information includes determining the optimal sparse pixel set information based on at least one ℓ_1 -minimization protocol. At **316**, determining the optimal sparse pixel set information includes determining optimal sparse pixel location information. At **318**, determining the optimal
 20 sparse pixel set information includes determining optimal sparse pixel weighting factor information. At **320**, determining the optimal sparse pixel set information includes determining enhanced sparsity classification information associated with the one or more reference images. At **322**, determining the optimal sparse pixel set information includes applying a coupling weight protocol.

25 At **330**, determining the optimal sparse pixel set information includes generating discrimination vector information associated with the one or more reference images. At **340**, the method **300** includes comparing one or more parameters associated with the optimal sparse pixel set information to one or more pixels associated with an image. At **350**, the method **300** includes classifying at least one object in the image based on the
 30 comparison.

Figure 14 shows a method **400**. At **410**, the method **400** includes determining significant sensors from a sensor network. At **412**, determining significant sensors from the sensor network includes determining one or more spatial sensor locations optimally

informative to a categorical discrimination based on enhanced sparsity protocol. At **414**, determining significant sensors from the sensor network includes determining one or more spatial sensor locations optimally informative to a categorical discrimination based on enhanced sparsity for classification protocol. At **416**, determining significant sensors from the sensor network includes determining sparse representation information of discriminant vectors in a feature space. At **418**, determining significant sensors from the sensor network includes determining sparse representation information of discriminant vectors in a feature space based on an ℓ_1 -minimization protocol. At **420**, determining significant sensors from the sensor network includes determining nearly optimal sparse sensor locations information. At **422**, determining the significant sensors from the sensor network includes determining significant nodes from a sensor network, and generating optimal the sparse sensor set information includes generating optimal sparse node set information from the significant nodes.

At **430**, the method **400** includes generating optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs. At **440**, the method **400** includes classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information. At **442**, classifying the information associated with the target decision task includes generating categorical decision information about a state of the sensor network.

It is noted that Figures 13 and 14 denotes “start” and “end” positions. However, nothing herein should be construed to indicate that these are limiting and it is contemplated that other or additional steps or functions can occur before or after those described in Figures 13 and 14.

The claims, description, and drawings of this application may describe one or more of the instant technologies in operational/functional language, for example as a set of operations to be performed by a computer. Such operational/functional description in most instances can be specifically-configured hardware (*e.g.*, because a general purpose computer in effect becomes a special purpose computer once it is programmed to perform particular functions pursuant to instructions from program software).

Importantly, although the operational/functional descriptions described herein are understandable by the human mind, they are not abstract ideas of the operations/functions divorced from computational implementation of those operations/functions. Rather, the

operations/functions represent a specification for the massively complex computational machines or other means. As discussed in detail below, the operational/functional language must be read in its proper technological context, i.e., as concrete specifications for physical implementations.

5 The logical operations/functions described herein are a distillation of machine specifications or other physical mechanisms specified by the operations/functions such that the otherwise inscrutable machine specifications may be comprehensible to the human mind. The distillation also allows one of skill in the art to adapt the operational/functional description of the technology across many different specific vendors' hardware
10 configurations or platforms, without being limited to specific vendors' hardware configurations or platforms.

 Some of the present technical description (*e.g.*, detailed description, drawings, claims, etc.) may be set forth in terms of logical operations/functions. As described in more detail in the following paragraphs, these logical operations/functions are not
15 representations of abstract ideas, but rather representative of static or sequenced specifications of various hardware elements. Differently stated, unless context dictates otherwise, the logical operations/functions are representative of static or sequenced specifications of various hardware elements. This is true because tools available to
20 implement technical disclosures set forth in operational/functional formats—tools in the form of a high-level programming language (*e.g.*, C, java, visual basic), etc.), or tools in the form of Very high speed Hardware Description Language (“VIDAL,” which is a language that uses text to describe logic circuits—) are generators of static or sequenced specifications of various hardware configurations. This fact is sometimes obscured by the broad term “software,” but, as shown by the following explanation, what is termed
25 “software” is a shorthand for a massively complex interchanging/specification of ordered-matter elements. The term “ordered-matter elements” may refer to physical components of computation, such as assemblies of electronic logic gates, molecular computing logic constituents, quantum computing mechanisms, etc.

 For example, a high-level programming language is a programming language with
30 strong abstraction, *e.g.*, multiple levels of abstraction, from the details of the sequential organizations, states, inputs, outputs, etc., of the machines that a high-level programming language actually specifies. *See, e.g.*, Wikipedia, High-level programming language, available at the website en.wikipedia.org/wiki/High-level_programming_language (as of

June 5, 2012, 21:00 GMT). In order to facilitate human comprehension, in many instances, high-level programming languages resemble or even share symbols with natural languages. *See, e.g.,* Wikipedia, Natural language, available at the website en.wikipedia.org/wiki/Natural_language (as of June 5, 2012, 21:00 GMT).

5 It has been argued that because high-level programming languages use strong abstraction (*e.g.*, that they may resemble or share symbols with natural languages), they are therefore a “purely mental construct” (*e.g.*, that “software” – a computer program or computer –programming—is somehow an ineffable mental construct, because at a high level of abstraction, it can be conceived and understood in the human mind). This
10 argument has been used to characterize technical description in the form of functions/operations as somehow “abstract ideas.” In fact, in technological arts (*e.g.*, the information and communication technologies) this is not true.

 The fact that high-level programming languages use strong abstraction to facilitate human understanding should not be taken as an indication that what is expressed is an
15 abstract idea. In an embodiment, if a high-level programming language is the tool used to implement a technical disclosure in the form of functions/operations, it can be understood that, far from being abstract, imprecise, “fuzzy,” or “mental” in any significant semantic sense, such a tool is instead a near incomprehensibly precise sequential specification of specific computational –machines—the parts of which are built up by activating/selecting
20 such parts from typically more general computational machines over time (*e.g.*, clocked time). This fact is sometimes obscured by the superficial similarities between high-level programming languages and natural languages. These superficial similarities also may cause a glossing over of the fact that high-level programming language implementations ultimately perform valuable work by creating/controlling many different computational
25 machines.

 The many different computational machines that a high-level programming language specifies are almost unimaginably complex. At base, the hardware used in the computational machines typically consists of some type of ordered matter (*e.g.*, traditional electronic devices (*e.g.*, transistors), deoxyribonucleic acid (DNA), quantum devices,
30 mechanical switches, optics, fluidics, pneumatics, optical devices (*e.g.*, optical interference devices), molecules, etc.) that are arranged to form logic gates. Logic gates are typically physical devices that may be electrically, mechanically, chemically, or

otherwise driven to change physical state in order to create a physical reality of Boolean logic.

Logic gates may be arranged to form logic circuits, which are typically physical devices that may be electrically, mechanically, chemically, or otherwise driven to create a physical reality of certain logical functions. Types of logic circuits include such devices as multiplexers, registers, arithmetic logic units (ALUs), computer memory devices, etc., each type of which may be combined to form yet other types of physical devices, such as a central processing unit (CPU)—the best known of which is the microprocessor. A modern microprocessor will often contain more than one hundred million logic gates in its many logic circuits (and often more than a billion transistors). See, *e.g.*, Wikipedia, Logic gates, available at the website en.wikipedia.org/wiki/Logic_gates (as of June 5, 2012, 21:03 GMT).

The logic circuits forming the microprocessor are arranged to provide a microarchitecture that will carry out the instructions defined by that microprocessor's defined Instruction Set Architecture. The Instruction Set Architecture is the part of the microprocessor architecture related to programming, including the native data types, instructions, registers, addressing modes, memory architecture, interrupt and exception handling, and external Input/Output. See, *e.g.*, Wikipedia, Computer architecture, available at the website en.wikipedia.org/wiki/Computer_architecture (as of June 5, 2012, 21:03 GMT).

The Instruction Set Architecture includes a specification of the machine language that can be used by programmers to use/control the microprocessor. Since the machine language instructions are such that they may be executed directly by the microprocessor, typically they consist of strings of binary digits, or bits. For example, a typical machine language instruction might be many bits long (*e.g.*, 32, 64, or 128 bit strings are currently common). A typical machine language instruction might take the form “11110000101011110000111100111111” (a 32 bit instruction).

It is significant here that, although the machine language instructions are written as sequences of binary digits, in actuality those binary digits specify physical reality. For example, if certain semiconductors are used to make the operations of Boolean logic a physical reality, the apparently mathematical bits “1” and “0” in a machine language instruction actually constitute a shorthand that specifies the application of specific voltages to specific wires. For example, in some semiconductor technologies, the binary number

“1” (*e.g.*, logical “1”) in a machine language instruction specifies around +5 volts applied to a specific “wire” (*e.g.*, metallic traces on a printed circuit board) and the binary number “0” (*e.g.*, logical “0”) in a machine language instruction specifies around -5 volts applied to a specific “wire.” In addition to specifying voltages of the machines’ configuration,
5 such machine language instructions also select out and activate specific groupings of logic gates from the millions of logic gates of the more general machine. Thus, far from abstract mathematical expressions, machine language instruction programs, even though written as a string of zeros and ones, specify many, many constructed physical machines or physical machine states.

10 Machine language is typically incomprehensible by most humans (*e.g.*, the above example was just ONE instruction, and some personal computers execute more than two billion instructions every second). See, *e.g.*, Wikipedia, Instructions per second, [available at the website en.wikipedia.org/wiki/Instructions_per_second](http://en.wikipedia.org/wiki/Instructions_per_second) (as of June 5, 2012, 21:04 GMT).

15 Thus, programs written in machine language—which may be tens of millions of machine language instructions long—are incomprehensible. In view of this, early assembly languages were developed that used mnemonic codes to refer to machine language instructions, rather than using the machine language instructions’ numeric values directly (*e.g.*, for performing a multiplication operation, programmers coded the
20 abbreviation “mult,” which represents the binary number “011000” in MIPS machine code). While assembly languages were initially a great aid to humans controlling the microprocessors to perform work, in time the complexity of the work that needed to be done by the humans outstripped the ability of humans to control the microprocessors using merely assembly languages.

25 At this point, it was noted that the same tasks needed to be done over and over, and the machine language necessary to do those repetitive tasks was the same. In view of this, compilers were created. A compiler is a device that takes a statement that is more comprehensible to a human than either machine or assembly language, such as “add 2 + 2 and output the result,” and translates that human understandable statement into a
30 complicated, tedious, and immense machine language code (*e.g.*, millions of 32, 64, or 128 bit length strings). Compilers thus translate high-level programming language into machine language.

This compiled machine language, as described above, is then used as the technical specification which sequentially constructs and causes the interoperation of many different computational machines such that humanly useful, tangible, and concrete work is done. For example, as indicated above, such machine language– the compiled version of the
5 higher-level language– functions as a technical specification which selects out hardware logic gates, specifies voltage levels, voltage transition timings, etc., such that the humanly useful work is accomplished by the hardware.

Thus, a functional/operational technical description, when viewed by one of skill in the art, is far from an abstract idea. Rather, such a functional/operational technical
10 description, when understood through the tools available in the art such as those just described, is instead understood to be a humanly understandable representation of a hardware specification, the complexity and specificity of which far exceeds the comprehension of most any one human. Accordingly, any such operational/functional technical descriptions may be understood as operations made into physical reality by (a)
15 one or more interchained physical machines, (b) interchained logic gates configured to create one or more physical machine(s) representative of sequential/combinatorial logic(s), (c) interchained ordered matter making up logic gates (*e.g.*, interchained electronic devices (*e.g.*, transistors), DNA, quantum devices, mechanical switches, optics, fluidics, pneumatics, molecules, etc.) that create physical reality representative of logic(s), or (d)
20 virtually any combination of the foregoing. Indeed, any physical object which has a stable, measurable, and changeable state may be used to construct a machine based on the above technical description. Charles Babbage, for example, constructed the first computer out of wood and powered by cranking a handle.

Thus, far from being understood as an abstract idea, it can be recognizes that a
25 functional/operational technical description as a humanly-understandable representation of one or more almost unimaginably complex and time sequenced hardware instantiations. The fact that functional/operational technical descriptions might lend themselves readily to high-level computing languages (or high-level block diagrams for that matter) that share some words, structures, phrases, etc. with natural language simply cannot be taken as an
30 indication that such functional/operational technical descriptions are abstract ideas, or mere expressions of abstract ideas. In fact, as outlined herein, in the technological arts this is simply not true. When viewed through the tools available to those of skill in the art,

such functional/operational technical descriptions are seen as specifying hardware configurations of almost unimaginable complexity.

As outlined above, the reason for the use of functional/operational technical descriptions is at least twofold. First, the use of functional/operational technical descriptions allows near-infinitely complex machines and machine operations arising from interchained hardware elements to be described in a manner that the human mind can process (*e.g.*, by mimicking natural language and logical narrative flow). Second, the use of functional/operational technical descriptions assists the person of skill in the art in understanding the described subject matter by providing a description that is more or less independent of any specific vendor's piece(s) of hardware.

The use of functional/operational technical descriptions assists the person of skill in the art in understanding the described subject matter since, as is evident from the above discussion, one could easily, although not quickly, transcribe the technical descriptions set forth in this document as trillions of ones and zeroes, billions of single lines of assembly-level machine code, millions of logic gates, thousands of gate arrays, or any number of intermediate levels of abstractions. However, if any such low-level technical descriptions were to replace the present technical description, a person of skill in the art could encounter undue difficulty in implementing the disclosure, because such a low-level technical description would likely add complexity without a corresponding benefit (*e.g.*, by describing the subject matter utilizing the conventions of one or more vendor-specific pieces of hardware). Thus, the use of functional/operational technical descriptions assists those of skill in the art by separating the technical descriptions from the conventions of any vendor-specific piece of hardware.

In view of the foregoing, the logical operations/functions set forth in the present technical description are representative of static or sequenced specifications of various ordered-matter elements, in order that such specifications may be comprehensible to the human mind and adaptable to create many various hardware configurations. The logical operations/functions disclosed herein should be treated as such, and should not be disparagingly characterized as abstract ideas merely because the specifications they represent are presented in a manner that one of skill in the art can readily understand and apply in a manner independent of a specific vendor's hardware implementation.

At least a portion of the devices or processes described herein can be integrated into an information processing system. An information processing system generally

includes one or more of a system unit housing, a video display device, memory, such as volatile or non-volatile memory, processors such as microprocessors or digital signal processors, computational entities such as operating systems, drivers, graphical user interfaces, and applications programs, one or more interaction devices (*e.g.*, a touch pad, a touch screen, an antenna, etc.), or control systems including feedback loops and control motors (*e.g.*, feedback for detecting position or velocity, control motors for moving or adjusting components or quantities). An information processing system can be implemented utilizing suitable commercially available components, such as those typically found in data computing/communication or network computing/communication systems.

10 The state of the art has progressed to the point where there is little distinction left between hardware and software implementations of aspects of systems; the use of hardware or software is generally (but not always, in that in certain contexts the choice between hardware and software can become significant) a design choice representing cost vs. efficiency tradeoffs. Various vehicles by which processes or systems or other technologies described herein can be effected (*e.g.*, hardware, software, firmware, etc., in one or more machines or articles of manufacture), and that the preferred vehicle will vary with the context in which the processes, systems, other technologies, etc., are deployed. For example, if an implementer determines that speed and accuracy are paramount, the implementer may opt for a mainly hardware or firmware vehicle; alternatively, if flexibility is paramount, the implementer may opt for a mainly software implementation that is implemented in one or more machines or articles of manufacture; or, yet again alternatively, the implementer may opt for some combination of hardware, software, firmware, etc. in one or more machines or articles of manufacture. Hence, there are several possible vehicles by which the processes, devices, other technologies, etc., described herein may be effected, none of which is inherently superior to the other in that any vehicle to be utilized is a choice dependent upon the context in which the vehicle will be deployed and the specific concerns (*e.g.*, speed, flexibility, or predictability) of the implementer, any of which may vary. In an embodiment, optical aspects of implementations will typically employ optically-oriented hardware, software, firmware, etc., in one or more machines or articles of manufacture.

The herein described subject matter sometimes illustrates different components contained within, or connected with, different other components. It is to be understood that such depicted architectures are merely examples, and that in fact, many other

architectures can be implemented that achieve the same functionality. In a conceptual sense, any arrangement of components to achieve the same functionality is effectively “associated” such that the desired functionality is achieved. Hence, any two components herein combined to achieve a particular functionality can be seen as “associated with”
5 each other such that the desired functionality is achieved, irrespective of architectures or intermedial components. Likewise, any two components so associated can also be viewed as being “operably connected”, or “operably coupled,” to each other to achieve the desired functionality, and any two components capable of being so associated can also be viewed as being “operably coupleable,” to each other to achieve the desired functionality.
10 Specific examples of operably coupleable include, but are not limited to, physically mateable, physically interacting components, wirelessly interactable, wirelessly interacting components, logically interacting, logically interactable components, etc.

In an embodiment, one or more components may be referred to herein as “configured to,” “configurable to,” “operable/operative to,” “adapted/adaptable,” “able to,” “conformable/conformed to,” etc. Such terms (*e.g.*, “configured to”) can generally
15 encompass active-state components, or inactive-state components, or standby-state components, unless context requires otherwise.

The foregoing detailed description has set forth various embodiments of the devices or processes *via* the use of block diagrams, flowcharts, or examples. Insofar as
20 such block diagrams, flowcharts, or examples contain one or more functions or operations, it will be understood by the reader that each function or operation within such block diagrams, flowcharts, or examples can be implemented, individually or collectively, by a wide range of hardware, software, firmware in one or more machines or articles of manufacture, or virtually any combination thereof. Further, the use of “Start,” “End,” or
25 “Stop” blocks in the block diagrams is not intended to indicate a limitation on the beginning or end of any functions in the diagram. Such flowcharts or diagrams may be incorporated into other flowcharts or diagrams where additional functions are performed before or after the functions shown in the diagrams of this application. In an embodiment, several portions of the subject matter described herein is implemented *via* Application
30 Specific Integrated Circuits (ASICs), Field Programmable Gate Arrays (FPGAs), digital signal processors (DSPs), or other integrated formats. However, some aspects of the embodiments disclosed herein, in whole or in part, can be equivalently implemented in integrated circuits, as one or more computer programs running on one or more computers

(*e.g.*, as one or more programs running on one or more computer systems), as one or more programs running on one or more processors (*e.g.*, as one or more programs running on one or more microprocessors), as firmware, or as virtually any combination thereof, and that designing the circuitry or writing the code for the software and or firmware would be well within the skill of one of skill in the art in light of this disclosure. In addition, the mechanisms of the subject matter described herein are capable of being distributed as a program product in a variety of forms, and that an illustrative embodiment of the subject matter described herein applies regardless of the particular type of signal-bearing medium used to actually carry out the distribution. Non-limiting examples of a signal-bearing medium include the following: a recordable type medium such as a floppy disk, a hard disk drive, a Compact Disc (CD), a Digital Video Disk (DVD), a digital tape, a computer memory, etc.; and a transmission type medium such as a digital or an analog communication medium (*e.g.*, a fiber optic cable, a waveguide, a wired communications link, a wireless communication link (*e.g.*, transmitter, receiver, transmission logic, reception logic, etc.), etc.).

While particular aspects of the present subject matter described herein have been shown and described, it will be apparent to the reader that, based upon the teachings herein, changes and modifications can be made without departing from the subject matter described herein and its broader aspects and, therefore, the appended claims are to encompass within their scope all such changes and modifications as are within the true spirit and scope of the subject matter described herein. In general, terms used herein, and especially in the appended claims (*e.g.*, bodies of the appended claims) are generally intended as “open” terms (*e.g.*, the term “including” should be interpreted as “including but not limited to,” the term “having” should be interpreted as “having at least,” the term “includes” should be interpreted as “includes but is not limited to,” etc.). Further, if a specific number of an introduced claim recitation is intended, such an intent will be explicitly recited in the claim, and in the absence of such recitation no such intent is present. For example, as an aid to understanding, the following appended claims may contain usage of the introductory phrases “at least one” and “one or more” to introduce claim recitations. However, the use of such phrases should not be construed to imply that the introduction of a claim recitation by the indefinite articles “a” or “an” limits any particular claim containing such introduced claim recitation to claims containing only one such recitation, even when the same claim includes the introductory phrases “one or more”

or “at least one” and indefinite articles such as “a” or “an” (*e.g.*, “a” and/or “an” should typically be interpreted to mean “at least one” or “one or more”); the same holds true for the use of definite articles used to introduce claim recitations. In addition, even if a specific number of an introduced claim recitation *is* explicitly recited, such recitation
5 should typically be interpreted to mean *at least* the recited number (*e.g.*, the bare recitation of “two recitations, ” without other modifiers, typically means *at least* two recitations, or *two or more* recitations). Furthermore, in those instances where a convention analogous to “at least one of A, B, and C, etc.” is used, in general such a construction is intended in the sense of the convention (*e.g.*, “ a system having at least one of A, B, and C” would include
10 but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). In those instances where a convention analogous to “at least one of A, B, or C, etc.” is used, in general such a construction is intended in the sense of the convention (*e.g.*, “ a system having at least one of A, B, or C” would include but not be limited to systems that have A alone, B alone, C
15 alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). Typically a disjunctive word or phrase presenting two or more alternative terms, whether in the description, claims, or drawings, should be understood to contemplate the possibilities of including one of the terms, either of the terms, or both terms unless context dictates otherwise. For example, the phrase “A or B” will be typically understood to
20 include the possibilities of “A” or “B” or “A and B.”

With respect to the appended claims, the operations recited therein generally may be performed in any order. Also, although various operational flows are presented in a sequence(s), it should be understood that the various operations may be performed in orders other than those that are illustrated, or may be performed concurrently. Examples
25 of such alternate orderings includes overlapping, interleaved, interrupted, reordered, incremental, preparatory, supplemental, simultaneous, reverse, or other variant orderings, unless context dictates otherwise. Furthermore, terms like “responsive to,” “related to,” or other past-tense adjectives are generally not intended to exclude such variants, unless context dictates otherwise.

30 We demonstrated our approach on two examples of image recognition. Classifiers built on learned sensors approached the performance of classifiers built on projections to principal components of the full images. These optimal sensor locations could be learned approximately even from already randomly subsampled images. In either case, the learned

sensors performed significantly better than the matched number of randomly chosen sensors. Further, the ensemble of learned sensor locations clustered around coherent features of the images. It is possible to think of this ensemble as a pixel mask for the faces in the training set, which may be of use when applied to engineered or biological systems where the important features may not be salient by inspection.

It is plausible that biological organisms may exploit enhanced sparsity for classification. Organisms interact with the external world with motor outputs, which are often discrete trajectories at specific moments in time, so that the transformation from sensory inputs to motor outputs can be thought of as a classification task. For example, a fly has no need to reconstruct the full flow velocity field around its body—it has only to decide what to do in response to a gust. Sensory organs and data processing by the nervous system can be expensive; therefore, it is advantageous to place a smaller number of sensors at key locations on the body.

Examples

We apply the algorithm for learning sparse sensor locations, as described above, on two examples of face recognition based on publicly available image datasets. Both data sets were presented and described (B. W. Brunton, S. L. Brunton, J. L. Proctor, J. N. Kutz, Optimal Sensor Placement and Enhanced Sparsity for Classification, arXiv:1310.2417 [cs.CV] , which is incorporated by reference herein). In each experiment, we demonstrate that a classifier built on a few optimally placed sparse sensors performs comparably to a classifier built on PCA features of the full image. Moreover, approximately optimal sparse sensor locations may be learned from pixels randomly subsampled from 10% of the full image. Ensembles of the learned sensors cluster at locations consistent with the coherent features in the faces.

A. Experiment 1 – Cats and Dogs

Our first example seeks to classify images of cats versus images of dogs. Each image belongs to a distinct species, and there is considerable variability between individuals of the same species (Fig. 10). Notably, members of each species take on a large range of colorations in fur, markings, and ear postures. We used 121 images each of cats and dogs; images have $n = 64 \times 64 = 4096$ pixels and are stacked into columns of X . Figure 10 (b) shows the first four PCA modes of X .

We compared classification accuracy using learned measurements as well as using the principal components of the full image and using random pixel measurements. The

accuracy for all of these methods improved with larger numbers of sensors/features; r learned sensors performed almost as well as r principal components and consistently better than r random pixel sensors. Figure 3 shows the results of experiments where classifiers were trained on a random 90% of the images and assessed on the remaining 10% of images. The mean and standard deviation of the cross-validated accuracy over 400 random iterations of training/test images are plotted. Classifiers built on less than 1% of the total pixels (green and red lines in Fig. 3 (a)) performed nearly as well as projections to principal components of the full image.

Figure 3 depicts a cross-validation of classification accuracy between images of cats and dogs. Panel (a) compares using r learned sensors/features (solid red and green lines) against using r random pixel sensors (dashed blue line) and projections onto the first r principal components of the full image (solid blue line). Each data point summarizes 400 random iterations. At each iteration, a different 90% subsample was used to train the classifier, whose accuracy was assessed on the remaining 10% of images. Error bars are standard deviations. Panel (b) shows a summary of mean cross-validated accuracy varying p , the number of pixels used in the random subsample, and r , the number of features/sensors used to construct the classifier.

Accuracy reached a plateau at around $r = 15$ features/ sensors (Fig. 3 (b)) of around 80%, suggesting the discriminant vector between cats and dogs may be over-fit when more than 15 features were used. Figure 3 (b) shows mean classification performance using r sensors learned from p subsampled measurements and r features, where p and r are varied systematically. Sparse sensors learned from $p = 400$ pixels did almost as well as those learned from $p = n = 4096$ pixels. In other words, we were able to learn a nearly optimally sparse set of sensors starting from an already massively undersampled set of pixels.

The performance of all of these approaches is limited by the large variability in appearances within the categories of cats and dogs. In fact, the images that are most often misclassified include animals that are predominantly one color (black or white) and dogs with pointy as opposed to droopy ears. In other words, some dog images may not lie close to the bulk of the other dog images in feature space, so that a linear classifier is not able to capture the complex geometry separating the two categories.

It is possible to build classifiers from random pixels that perform significantly better than chance. The improvement in accuracy of using computed features or learned

sensors over random pixels became smaller as r increases (Fig. 3 (a)). The perhaps surprising performance of random pixels is consistent with the notion of enhanced sparsity and Wright et al.'s (J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009, which is incorporated by reference herein) observation that using random features, as long as there is enough of them, serves classification as well as using engineered features.

Figure 4 shows the mean sensor locations averaged over 400 random learning iterations (top row), compared to images of the mean cat and the mean dog (bottom row). For the top row, $r = 15$ sensors locations are learned at each iteration from a random 50% of images designated as the training set, and the colormap from black to white represents the probability a sensors is at that location. In other words, we are visualizing the distribution of sensor locations, obtained by summing the rows of either $\hat{\Phi}_1^T$ or $(\hat{\Phi}_2\tilde{\Phi})^T$. Panel (a) shows sensors learned with no subsampling; panel (b) shows sensors learned from 400 randomly sampled pixels; panel (c) shows sensors learned from 400 random projections. The bottom row shows the centroid of the raw data in panel (c), which was subtracted from each image to obtain $\mathbf{X}$. Panels (e) and (f) show the average cat and the average dog after mean subtraction. Comparing the top row to the difference between (e) and (f), it is apparently that sensors cluster around the forehead, eyes, mouth, and the tops of ears.

If the sparse sensors represent pixels that are key to the decision, then they should be clustered at locations that are maximally informative about the difference between cats and dogs. Indeed, an average of sensor locations learned over 400 random iterations shows clustering around the animals' mouth, forehead, eyes, and tops of ears (Fig. 4 (a)). When sparse sensors were learned from an already subsampled dataset, a qualitatively similar distribution of sensor locations was found, where each cluster of sensors showed a slightly larger spatial variance (Fig. 4 (b)). Such an ensemble of sensor placements can be thought of as a mask for facial features particularly relevant for the classification.

Our algorithm applies equally well when the subsampled dataset (middle row in Fig. 1) is not a set of random pixels but a set of random projections. Figure 5 shows that classifiers built on r sensors learned from $p = 400$ pixels or $p = 400$ random projections perform identically. Random projections used to produce this result consisted of ensembles of Bernoulli random variables with mean of 0.5. Even so, Fig. 4 (c) makes clear

that using random projections instead of pixels is a poor engineering choice. The ensemble of sensors computed from random projections does not illuminate coherent facial features.

Figure 5 shows that the classification accuracy of sensors learned from random projections (black dashed line) is the same as sensors learned from single pixels (red solid line). The experiments performed were identical to those show in Fig. 3. Random projections were ensembles of Bernoulli random variables with mean of 0.5.

B. Experiment 2 – Yale B Faces

We extended sparse sensor learning to classification between more than two categories, applying our approach to human face recognition. We used the Yale Faces Database B extended (A.S. Georgiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):643–660, 2001, and K. Lee, J. Ho, and D. J Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans Pattern Anal Mach Intell*, 27(5):684–98, May 2005, which are each incorporated by reference herein), which contains images of individual faces captured under various lighting conditions (example images in Fig. 11 (a)). Images have been perviously aligned and cropped; each has $n = 192 \times 168 = 32,256$ pixels. Figure 11 (b) shows the first six eigenfaces corresponding to the three example sets of faces. Comparing the faces dataset with the cat/dog dataset above, there is significantly less variability within each category in facial features, although large portions of the faces were effectively occluded by lack of illumination.

Figure 6 depicts classification of 3, 4, and 5 faces as the coupling weight λ varies between 0 and 100. Number of sensors and the cross-validated performance are shown, comparing sensors learned with no subsampling (solid green lines), sensors learned from 1/10 of the pixels (solid red lines), and random sensors of the same number (dashed blue lines) against the accuracy by using r PCA features of the full images (blue square). Each instance of the classifier was trained on a random 75% of the images and evaluated on the remaining 25%; error bars are standard deviations.

We demonstrated our sensor learning approach to learn sensor locations that categorize $c = 3, 4$, and 5 faces and compared the classification accuracy of learned sensors to projections to PCA features of the full image and to random sensors of the same number. Figure 6 shows that, as the coupling weight λ is increased, the number of learned sensors decreases. When $\lambda = 0$, a maximum of $r(c - 1)$ pixels locations were found. In the

limit $\lambda \rightarrow \infty$, the number of sensors approached r , the number of PCA features included in the discrimination. By adjusting the magnitude of λ , we gain control over the number of sensor locations used in the classification. When individual sensors are expensive, it may be desirable to use fewer sensors at the cost of slightly lower performance.

5 The cross-validated accuracy shown in the bottom row of Fig. 6 are results obtained from re-computing the LDA projection $\hat{w}^T$ to decision space after learning the sensor locations. This strategy is not typically an expensive computation, as the number of rows in the measurement vectors is small. Further, a LDA classifier tailor-made for the specific learned pixels can out-perform the PCA approach when the number of sensors
10 exceeds the number of PCA features. In contrast, the induced projection method described above cannot perform better than the PCA-LDA approach. For classification between $c > 2$ categories, especially for larger λ coupling weights, re-computing the LDA projection usually leads to more accurate results.

15 Interestingly, sensors learned from randomly subsampled pixels (10% of the original pixels) performed exactly as well as sensors learned from the full image (red and green lines in Fig. 6). Obtaining sparse sensors from already subsampled data presents significant savings in the learning procedure, both in the SVD to extract a feature space and in the convex optimization to solve for a sparse s .

20 Figure 7 shows that increasing the coupling weight λ results in fewer sensor locations that capture approximately the same features. In both panels, the underlying image is a visualization of the decision vector $X = \Psi_r w$ for an example of categorizing $c = 3$ faces using $r = 10$ features. Panel (a) shows the locations of sparse sensors that independently reconstruct the $c - 1 = 2$ columns of w (yellow and green dots, 20 total). The sensors boxed in red are aggregated in panel (b), where the coupling weight λ brought the
25 total number of sensors down to 15.

30 Figure 7 illustrates an example of increasing λ on the number and locations of sensors identified by the solution to the optimization in Eq. (9). For this example, $c = 3$ faces were in the training set so w and s each has $c - 1 = 2$ columns. When $\lambda = 0$, the columns of w (linear discriminant vectors in $r = 10$ dimensional PCA feature space) are treated independently, and $r(c - 1) = 20$ total sensors are found. Notice, however, in Fig. 7 (a) that certain pairs of sensor locations are in close proximity (red boxes) and likely carry information about the same non-local facial feature. As λ increases and the total number of sensors is penalized, these sensor pairs appear to collapse onto single sensors in Fig. 7 (b).

Comparing the two sets of sensor locations, we can see that, in addition to aggregated sensors, some sensors have remained the same (for example, the tops of each eyebrow), some sensors have disappeared (on the cheek, at bottom right), while some entirely new sensors have appeared (lower right corner of eye).

5 Figure 8 illustrates mean sensor locations (top left) to discriminate between 3 faces (mean faces b-d), averaged over 400 learning iterations where a random 75% subset of the images was used as the training set. The mean sensor locations map was shrunk by a factor of 4 (using a cubic kernel) to emphasize the sensors converging around the eyes, nose, corner of mouth, and arches of eyebrows.

10 Sparse sensors for face recognition cluster at major facial features: the eyes, the nose, and corners of the mouth (Fig. 8). The nose is particularly prominent in these masks, consistent with the fact that, due to the eccentricity in illumination, the nose is the only facial feature reliably visible in all images. Interestingly, this data-driven algorithm identifies the same features that are favored by humans. As first noted by Yarbus (A.L.
15 Yarbus, B. Haigh, and L.A. Riggs. Eye movements and vision. *Plenum press* New York, 1967, which is incorporated by reference herein), humans examining an image of a face spend a preponderance of time fixating at the eyes, the nose, and the mouth.

Aspects of the subject matter described herein are set out in the following numbered clauses:

- 20 1. In some embodiments, an object classification system includes: a significant pixel identification module including circuitry for determining significant pixels from one or more reference images, and generating optimal sparse pixel set information; and an object classification module including circuitry for comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal
25 sparse pixel set information; and classifying at least one object in the first image based on the comparison.
2. Some embodiments include an object classification system such as in paragraph 1, wherein the significant pixel identification module includes circuitry for generating a near optimal sparse pixel set information.
- 30 3. Some embodiments include an object classification system such as in paragraph 1, wherein the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based one or more learning protocols.

4. Some embodiments include an object classification system such as in paragraph 1, wherein the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based on a modal decomposition protocol.
5. Some embodiments include an object classification system such as in paragraph 1, wherein the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based on a convex optimization protocol.
6. Some embodiments include an object classification system such as in paragraph 1, wherein the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based on an ℓ_1 -minimization protocol.
- 10 7. Some embodiments include an object classification system such as in paragraph 1, wherein the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based on a subsampling using a sparse random matrices protocol.
8. Some embodiments include an object classification system such as in paragraph 1, wherein the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based on a subsampling using a dense random matrices protocol.
- 15 9. Some embodiments include an object classification system such as in paragraph 1, wherein the significant pixel identification module includes circuitry for determining one or more discrimination vectors associated with the significant pixels based on a discrimination analysis protocol.
- 20 10. Some embodiments include an object classification system such as in paragraph 9, wherein the object classification module includes circuitry for classifying the at least one object in the first image based on comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal sparse pixel set information and the one or more discrimination vectors.
- 25 11. Some embodiments include an object classification system such as in paragraph 1, wherein the object classification module includes circuitry for determining reference data clusters associated with the significant pixels in a sparse sensor space.
- 30 12. Some embodiments include an object classification system such as in paragraph 1, wherein the object classification module includes circuitry for comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the

- optimal sparse pixel set information and generating classification information associated with at least one object in the first image based on a learning protocol.
13. Some embodiments include an object classification system such as in paragraph 1, wherein the one or more reference images includes one or more training images.
- 5 14. In some embodiments, an object classification system includes: circuitry for determining an optimal sparse pixel set from one or more reference images; and circuitry for classifying an object in at least one image based on a comparison of one or more parameters associated with the optimal sparse pixel set to one or more pixels associated with the at least one image.
- 10 15. Some embodiments include an object classification system such as in paragraph 14, wherein the circuitry for determining the optimal sparse pixel set includes circuitry for determining one or more discrimination vectors from the one or more reference images.
16. Some embodiments include an object classification system such as in paragraph 14, wherein the circuitry for determining the optimal sparse pixel set includes circuitry for determining one or more sparse pixel locations.
- 15 17. Some embodiments include an object classification system such as in paragraph 14, wherein the circuitry for determining the optimal sparse pixel set includes circuitry for generating a feature space transformation from the one or more reference images.
- 20 18. Some embodiments include an object classification system such as in paragraph 14, wherein the circuitry for determining the optimal sparse pixel set includes circuitry for determining one or more weighting factors associated with the optimal sparse pixel set.
19. Some embodiments include an object classification system such as in paragraph 14, wherein the circuitry for classifying the object in at least one image includes circuitry for generating one or more classification categories associated with the object.
- 25 20. In some embodiments, a method for classifying images includes: determining optimal sparse pixel set information from one or more reference images; comparing one or more parameters associated with the optimal sparse pixel set information to one or more pixels associated with an image; and classifying at least one object in the image based on the comparison.
- 30 21. Some embodiments include a method for classifying images such as in paragraph 20, wherein determining the optimal sparse pixel set information includes determining the optimal sparse pixel set information based on at least one recovery protocol.

22. Some embodiments include a method for classifying images such as in paragraph 20, wherein determining the optimal sparse pixel set information includes determining the optimal sparse pixel set information based on at least one ℓ_1 -minimization protocol.
23. Some embodiments include a method for classifying images such as in paragraph 20,
5 wherein determining the optimal sparse pixel set information includes determining optimal sparse pixel location information.
24. Some embodiments include a method for classifying images such as in paragraph 20, wherein determining the optimal sparse pixel set information includes determining optimal sparse pixel weighting factor information.
- 10 25. Some embodiments include a method for classifying images such as in paragraph 20, wherein determining the optimal sparse pixel set information includes determining enhanced sparsity classification information associated with the one or more reference images.
26. Some embodiments include a method for classifying images such as in paragraph 20,
15 wherein determining the optimal sparse pixel set information includes applying a coupling weight protocol.
27. Some embodiments include a method for classifying images such as in paragraph 20, wherein determining the optimal sparse pixel set information includes generating discrimination vector information associated with the one or more reference images.
- 20 28. In some embodiments, an object classification system includes: a significant pixel determination from one or more reference images module; an optimal sparse pixel set information generation module; and an object classification module.
29. Some embodiments include an object classification system such as in paragraph 28,
25 wherein the object classification module includes: one or more first image pixels comparison to optimal sparse pixel set information, the one or more first image pixels indicative of at least one object in the first image; and a first image object classification module configured to classify at least one object in the first image based on a comparison of one or more pixels of the first image to optimal sparse pixel set information.
- 30 30. In some embodiments, an object classification system includes: a significant pixel identification module; and an object classification module including circuitry for comparing one or more pixels of a first image indicative of one or more objects

imaged in the first image to the optimal sparse pixel set information; and classifying at least one object in the first image based on the comparison.

31. Some embodiments include an object classification system such as in paragraph 30, wherein the significant pixel identification module includes circuitry for determining
5 significant pixels from one or more reference images, and generating optimal sparse pixel set information.
32. Some embodiments include an object classification system such as in paragraph 30, wherein the object classification module includes circuitry for comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the
10 optimal sparse pixel set information.
33. Some embodiments include an object classification system such as in paragraph 32, wherein the object classification module includes circuitry for classifying at least one object in the first image based on the comparison.
34. In some embodiments, a system includes: a significant sensors identification module
15 including circuitry for determining significant sensors associated with a sensor network; an optimal sparse sensor module including circuitry for generating optimal sparse sensor set information from the significant sensors; and a target decision task module including circuitry for classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and
20 one or more parameters associated with the optimal sparse sensor set information.
35. Some embodiments include a system such in paragraph 34, wherein the significant sensors identification module includes circuitry for determining optimal sparse sensor location information associated with the sensor network.
36. Some embodiments include a system such in paragraph 34, wherein the significant
25 sensors identification module includes circuitry for determining nearly optimal sparse sensor location information associated with the sensor network.
37. Some embodiments include a system such in paragraph 34, wherein the significant sensors identification module includes circuitry for determining optimal sparse sensor location information associated with the sensor network based on at least one ℓ_1 -
30 minimization protocol.
38. Some embodiments include a system such in paragraph 34, wherein the significant sensors identification module includes circuitry for determining significant notes associated with a sensor network.

39. Some embodiments include a system such in paragraph 34, wherein the significant sensors identification module includes circuitry for determining significant motes associated with a sensor network, and wherein the optimal sparse sensor module includes circuitry for generating optimal sparse mote set information from the significant motes responsive to one or more target decision task inputs.
40. Some embodiments include a system such in paragraph 34, wherein the significant sensors identification module includes circuitry for determining optimal sparse mote location information associated with a mote network.
41. Some embodiments include a system such in paragraph 34, wherein the optimal sparse sensor module includes circuitry for generating optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs.
42. In some embodiments, a system for classifying grid behavior includes: circuitry for determining significant sensors from a sensor network; circuitry for generating optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs; and circuitry for classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.
43. Some embodiments include a system for classifying grid behavior such in paragraph 42, wherein the circuitry for generating the optimal sparse sensor set information includes circuitry for generating one or more discrimination vectors associated with the sensor network.
44. Some embodiments include a system for classifying grid behavior such in paragraph 42, wherein the circuitry for generating the optimal sparse sensor set information includes circuitry for generating one or more sparse sensor locations associated with the sensor network.
45. Some embodiments include a system for classifying grid behavior such in paragraph 42, wherein the circuitry for classifying information associated with the target decision task includes circuitry for generating one or more classification categories associated with the target decision task.
46. Some embodiments include a system for classifying grid behavior such in paragraph 42, wherein the circuitry for classifying information associated with the target decision

task includes circuitry for generating one or more classification parameters associated with the target decision task.

47. In some embodiments, a method includes: determining significant sensors from a sensor network; generating optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs; and classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.
48. Some embodiments include a method as in paragraph 47, wherein determining significant sensors from the sensor network includes determining one or more spatial sensor locations optimally informative to a categorical discrimination based on enhanced sparsity protocol.
49. Some embodiments include a method as in paragraph 47, wherein determining significant sensors from the sensor network includes determining one or more spatial sensor locations optimally informative to a categorical discrimination based on enhanced sparsity for classification protocol.
50. Some embodiments include a method as in paragraph 47, wherein determining significant sensors from the sensor network includes determining sparse representation information of discriminant vectors in a feature space.
51. Some embodiments include a method as in paragraph 47, wherein determining significant sensors from the sensor network includes determining sparse representation information of discriminant vectors in a feature space based on an ℓ_1 -minimization protocol.
52. Some embodiments include a method as in paragraph 47, wherein determining significant sensors from the sensor network includes determining nearly optimal sparse sensor locations information.
53. Some embodiments include a method as in paragraph 47, wherein classifying the information associated with the target decision task includes generating categorical decision information about a state of the sensor network.
54. Some embodiments include a method as in paragraph 47, wherein determining the significant sensors from the sensor network includes determining significant nodes from a sensor network; and wherein generating optimal the sparse sensor set

information includes generating optimal sparse mote set information from the significant motes.

55. In some embodiments, a system includes: a significant sensors determination from a sensor network module; an optimal sparse sensor set information generation module;
- 5 and a target decision task classification module.
56. Some embodiments include a system such as in paragraph 55, wherein the optimal sparse sensor set information generation module is configured to generate optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs.
- 10 57. Some embodiments include a system such as in paragraph 55, wherein the target decision task classification module is configured to classify information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.
- 15 58. In some embodiments, a system includes: a significant sensors identification module; and an optimal sparse sensor module including circuitry for generating optimal sparse sensor set information from the significant sensors.
59. Some embodiments include a system such as in paragraph 58, further including a target decision task module.
- 20 60. Some embodiments include a system such as in paragraph 59, wherein the target decision task module includes circuitry for classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.
- 25 61. Some embodiments include a system such as in paragraph 58, wherein the significant sensors identification module includes circuitry for determining significant sensors associated with a sensor network.
62. Some embodiments include a system such as in paragraph 58, wherein the optimal sparse sensor module includes circuitry for generating optimal sparse sensor set
- 30 information from the significant sensors.

All of the above U.S. patents, U.S. patent application publications, U.S. patent applications, foreign patents, foreign patent applications and non-patent publications

referred to in this specification and/or listed in any Application Data Sheet, are incorporated herein by reference, to the extent not inconsistent herewith.

- While various aspects and embodiments have been disclosed herein, other aspects and embodiments are contemplated. The various aspects and embodiments disclosed
- 5 herein are for purposes of illustration and are not intended to be limiting, with the true scope and spirit being indicated by the following claims.

CLAIMS

What is claimed is:

1. An object classification system, comprising:
a significant pixel identification module including circuitry for determining
5 significant pixels from one or more reference images, and generating
optimal sparse pixel set information; and
an object classification module including circuitry for
comparing one or more pixels of a first image indicative of one or more
objects imaged in the first image to the optimal sparse pixel set
10 information; and
classifying at least one object in the first image based on the comparison.
2. The object classification system of claim 1, wherein the significant pixel
identification module includes circuitry for generating a near optimal sparse pixel
set information.
- 15 3. The object classification system of claim 1, wherein the significant pixel
identification module includes circuitry for generating the optimal sparse pixel set
information based one or more learning protocols.
4. The object classification system of claim 1, wherein the significant pixel
identification module includes circuitry for generating the optimal sparse pixel set
20 information based on a modal decomposition protocol.
5. The object classification system of claim 1, wherein the significant pixel
identification module includes circuitry for generating the optimal sparse pixel set
information based on a convex optimization protocol.
6. The object classification system of claim 1, wherein the significant pixel
25 identification module includes circuitry for generating the optimal sparse pixel set
information based on an ℓ_1 -minimization protocol.
7. The object classification system of claim 1, wherein the significant pixel
identification module includes circuitry for generating the optimal sparse pixel set
information based on a subsampling using a sparse random matrices protocol.

8. The object classification system of claim 1, wherein the significant pixel identification module includes circuitry for generating the optimal sparse pixel set information based on a subsampling using a dense random matrices protocol.
9. The object classification system of claim 1, wherein the significant pixel identification module includes circuitry for determining one or more discrimination vectors associated with the significant pixels based on a discrimination analysis protocol.
10. The object classification system of claim 9, wherein the object classification module includes circuitry for classifying the at least one object in the first image based on comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal sparse pixel set information and the one or more discrimination vectors.
11. The object classification system of claim 1, wherein the object classification module includes circuitry for determining reference data clusters associated with the significant pixels in a sparse sensor space.
12. The object classification system of claim 1, wherein the object classification module includes circuitry for comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal sparse pixel set information and generating classification information associated with at least one object in the first image based on a learning protocol.
13. The object classification system of claim 1, wherein the one or more reference images includes one or more training images.
14. An object classification system, comprising:
circuitry for determining an optimal sparse pixel set from one or more reference images; and
circuitry for classifying an object in at least one image based on a comparison of one or more parameters associated with the optimal sparse pixel set to one or more pixels associated with the at least one image.

15. The object classification system of claim 14, wherein the circuitry for determining the optimal sparse pixel set includes circuitry for determining one or more discrimination vectors from the one or more reference images.
- 5 16. The object classification system of claim 14, wherein the circuitry for determining the optimal sparse pixel set includes circuitry for determining one or more sparse pixel locations.
17. The object classification system of claim 14, wherein the circuitry for determining the optimal sparse pixel set includes circuitry for generating a feature space transformation from the one or more reference images.
- 10 18. The object classification system of claim 14, wherein the circuitry for determining the optimal sparse pixel set includes circuitry for determining one or more weighting factors associated with the optimal sparse pixel set.
19. The object classification system of claim 14, wherein the circuitry for classifying the object in at least one image includes circuitry for generating one or more classification categories associated with the object.
- 15 20. A method for classifying images, comprising:
determining optimal sparse pixel set information from one or more reference images;
comparing one or more parameters associated with the optimal sparse pixel set
20 information to one or more pixels associated with an image; and
classifying at least one object in the image based on the comparison.
21. The method for classifying images of claim 20, wherein determining the optimal sparse pixel set information includes determining the optimal sparse pixel set information based on at least one recovery protocol.
- 25 22. The method for classifying images of claim 20, wherein determining the optimal sparse pixel set information includes determining the optimal sparse pixel set information based on at least one ℓ_1 -minimization protocol.

23. The method for classifying images of claim 20, wherein determining the optimal sparse pixel set information includes determining optimal sparse pixel location information.
24. The method for classifying images of claim 20, wherein determining the optimal sparse pixel set information includes determining optimal sparse pixel weighting factor information.
25. The method for classifying images of claim 20, wherein determining the optimal sparse pixel set information includes determining enhanced sparsity classification information associated with the one or more reference images.
26. The method for classifying images of claim 20, wherein determining the optimal sparse pixel set information includes applying a coupling weight protocol.
27. The method for classifying images of claim 20, wherein determining the optimal sparse pixel set information includes generating discrimination vector information associated with the one or more reference images.
28. An object classification system, comprising:
a significant pixel determination from one or more reference images module;
an optimal sparse pixel set information generation module; and
an object classification module.
29. The object classification system of claim 28, wherein the object classification module includes
one or more first image pixels comparison to optimal sparse pixel set information,
the one or more first image pixels indicative of at least one object in the first image; and
a first image object classification module configured to classify at least one object in the first image based on a comparison of one or more pixels of the first image to optimal sparse pixel set information.
30. An object classification system, comprising:
a significant pixel identification module; and
an object classification module including circuitry for

comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal sparse pixel set information; and
classifying at least one object in the first image based on the comparison.

- 5 31. The object classification system of claim 30, wherein the significant pixel identification module includes circuitry for determining significant pixels from one or more reference images, and generating optimal sparse pixel set information.
- 10 32. The object classification system of claim 30, wherein the object classification module includes circuitry for comparing one or more pixels of a first image indicative of one or more objects imaged in the first image to the optimal sparse pixel set information.
- 15 33. The object classification system of claim 32, wherein the object classification module includes circuitry for classifying at least one object in the first image based on the comparison.
- 20 34. A system, comprising:
a significant sensors identification module including circuitry for determining significant sensors associated with a sensor network;
an optimal sparse sensor module including circuitry for generating optimal sparse sensor set information from the significant sensors; and
a target decision task module including circuitry for classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.
- 25 35. The system of claim 34, wherein the significant sensors identification module includes circuitry for determining optimal sparse sensor location information associated with the sensor network.
36. The system of claim 34, wherein the significant sensors identification module includes circuitry for determining nearly optimal sparse sensor location information associated with the sensor network.

37. The system of claim 34, wherein the significant sensors identification module includes circuitry for determining optimal sparse sensor location information associated with the sensor network based on at least one ℓ_1 -minimization protocol.
38. The system of claim 34, wherein the significant sensors identification module includes circuitry for determining significant motes associated with a sensor network.
39. The system of claim 34, wherein the significant sensors identification module includes circuitry for determining significant motes associated with a sensor network, and wherein the optimal sparse sensor module includes circuitry for generating optimal sparse mote set information from the significant motes responsive to one or more target decision task inputs.
40. The system of claim 34, wherein the significant sensors identification module includes circuitry for determining optimal sparse mote location information associated with a mote network.
41. The system of claim 34, wherein the optimal sparse sensor module includes circuitry for generating optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs.
42. A system for classifying grid behavior, comprising:
 circuitry for determining significant sensors from a sensor network;
 circuitry for generating optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs; and
 circuitry for classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.
43. The system for classifying grid behavior of claim 42, wherein the circuitry for generating the optimal sparse sensor set information includes circuitry for generating one or more discrimination vectors associated with the sensor network.

44. The system for classifying grid behavior of claim 42, wherein the circuitry for generating the optimal sparse sensor set information includes circuitry for generating one or more sparse sensor locations associated with the sensor network.
- 5 45. The system for classifying grid behavior of claim 42, wherein the circuitry for classifying information associated with the target decision task includes circuitry for generating one or more classification categories associated with the target decision task.
- 10 46. The system for classifying grid behavior of claim 42, wherein the circuitry for classifying information associated with the target decision task includes circuitry for generating one or more classification parameters associated with the target decision task.
- 15 47. A method, comprising:
determining significant sensors from a sensor network;
generating optimal sparse sensor set information from the significant sensors
responsive to one or more target decision task inputs; and
classifying information associated with the target decision task based on a
comparison of the one or more target decision task inputs and one or more
parameters associated with the optimal sparse sensor set information.
- 20 48. The method of claim 47, wherein determining significant sensors from the sensor network includes determining one or more spatial sensor locations optimally informative to a categorical discrimination based on enhanced sparsity protocol.
- 25 49. The method of claim 47, wherein determining significant sensors from the sensor network includes determining one or more spatial sensor locations optimally informative to a categorical discrimination based on enhanced sparsity for classification protocol.
50. The method of claim 47, wherein determining significant sensors from the sensor network includes determining sparse representation information of discriminant vectors in a feature space.

51. The method of claim 47, wherein determining significant sensors from the sensor network includes determining sparse representation information of discriminant vectors in a feature space based on an ℓ_1 -minimization protocol.
52. The method of claim 47, wherein determining significant sensors from the sensor network includes determining nearly optimal sparse sensor locations information.
53. The method of claim 47, wherein classifying the information associated with the target decision task includes generating categorical decision information about a state of the sensor network.
54. The method of claim 47, wherein determining the significant sensors from the sensor network includes determining significant nodes from a sensor network; and wherein generating optimal the sparse sensor set information includes generating optimal sparse node set information from the significant nodes.
55. A system, comprising:
a significant sensors determination from a sensor network module;
an optimal sparse sensor set information generation module; and
a target decision task classification module.
56. The system of claim 55, wherein the optimal sparse sensor set information generation module is configured to generate optimal sparse sensor set information from the significant sensors responsive to one or more target decision task inputs.
57. The system of claim 55, wherein the target decision task classification module is configured to classify information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.
58. A system, comprising:
a significant sensors identification module; and
an optimal sparse sensor module including circuitry for generating optimal sparse sensor set information from the significant sensors.
59. The system of claim 58, further comprising:

a target decision task module.

60. The system of claim 59, wherein the target decision task module includes circuitry for classifying information associated with the target decision task based on a comparison of the one or more target decision task inputs and one or more parameters associated with the optimal sparse sensor set information.
61. The system of claim 58, wherein the significant sensors identification module includes circuitry for determining significant sensors associated with a sensor network.
62. The system of claim 58, wherein the optimal sparse sensor module includes circuitry for generating optimal sparse sensor set information from the significant sensors.

FIG. 1

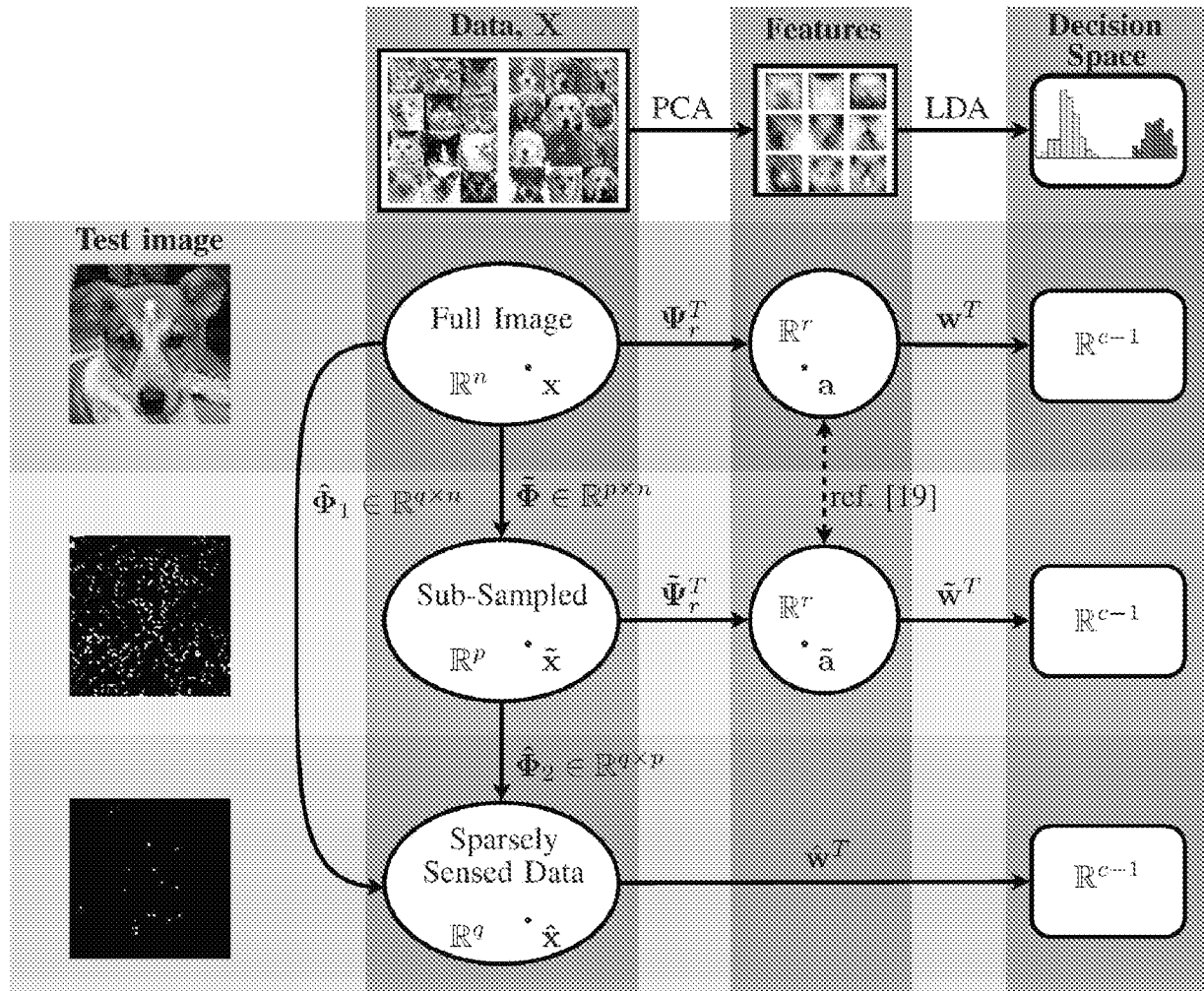


FIG. 2

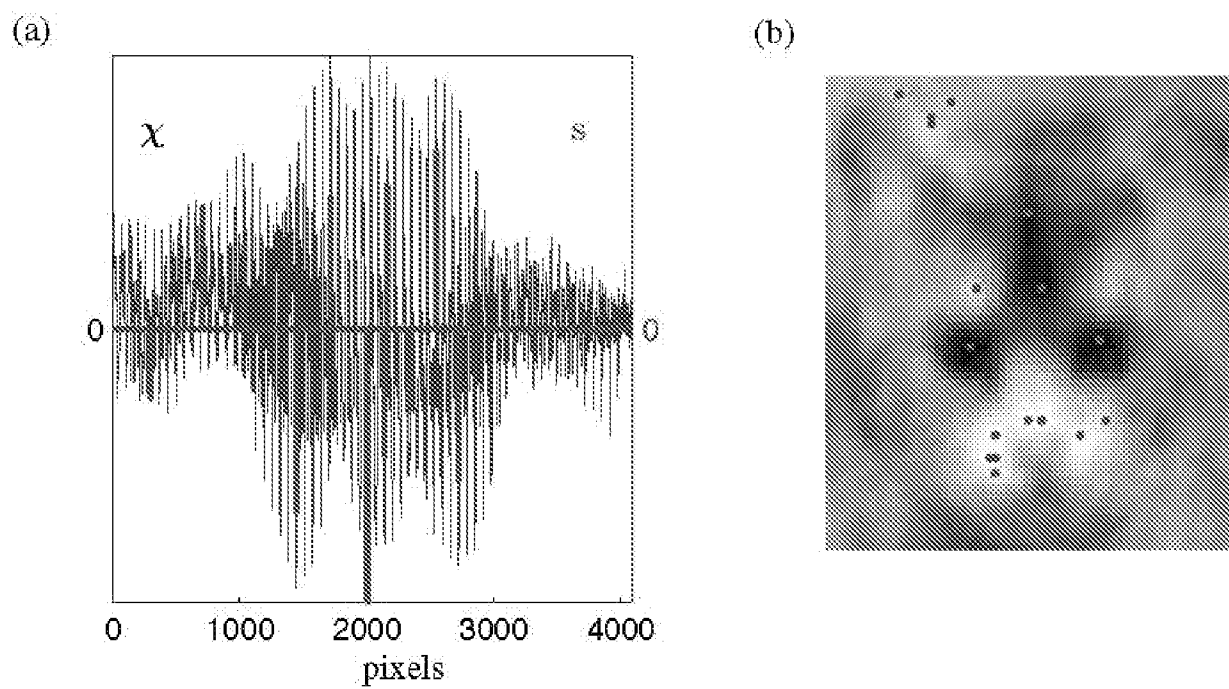


FIG. 3

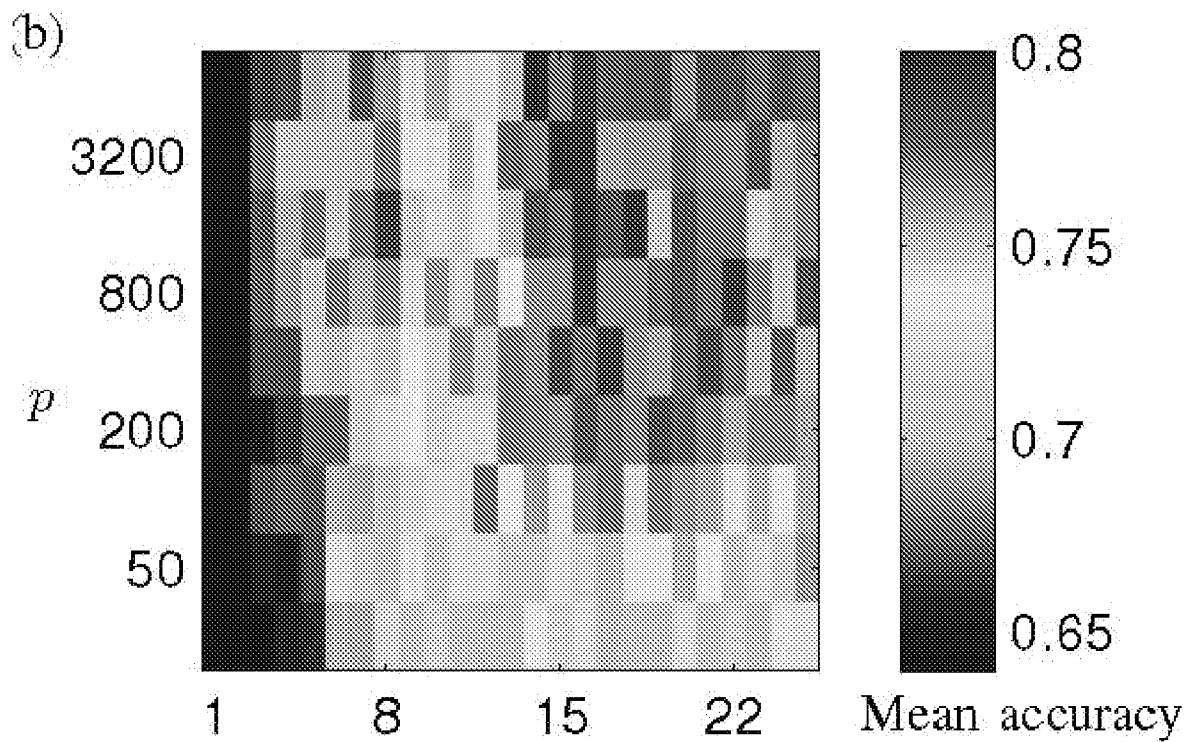
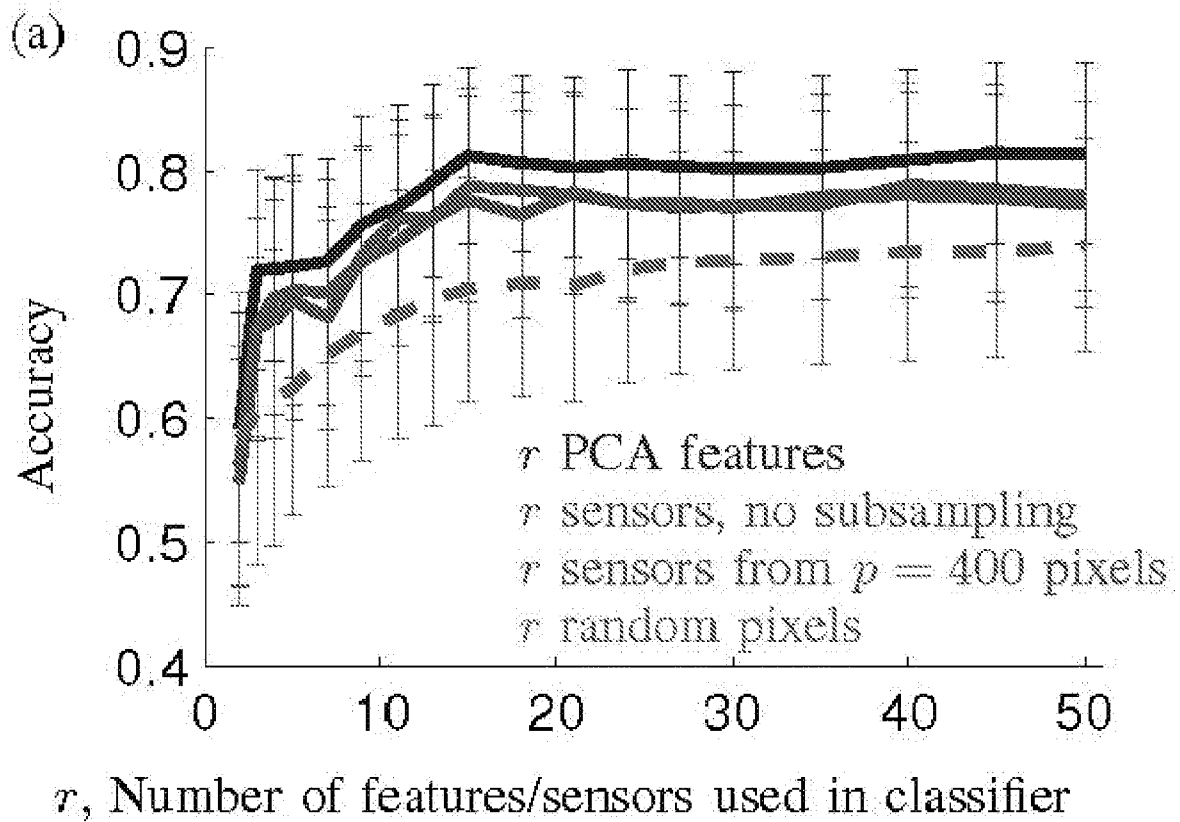
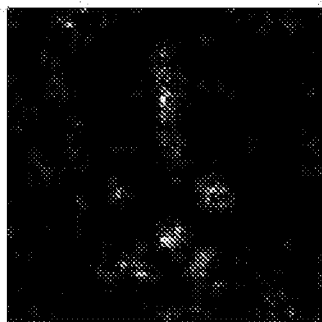
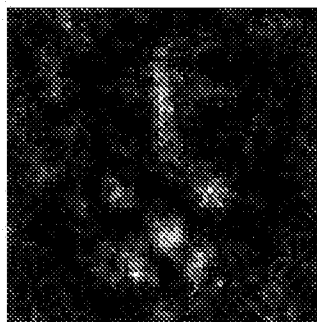
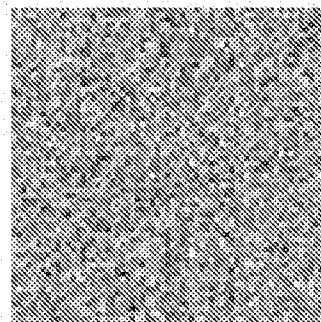
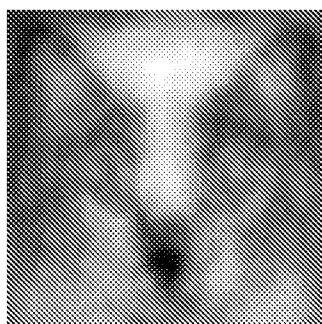


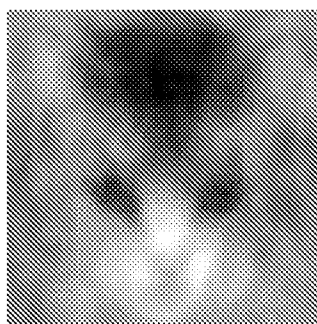
FIG. 4

(a) mean sensors
(no subsampling)(b) mean sensors
($p = 400$ pixels)(c) mean sensors
($p = 400$ r.p.)

(d) mean of raw data



(e) mean cat



(f) mean dog

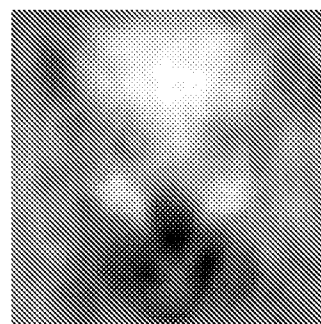


FIG. 5

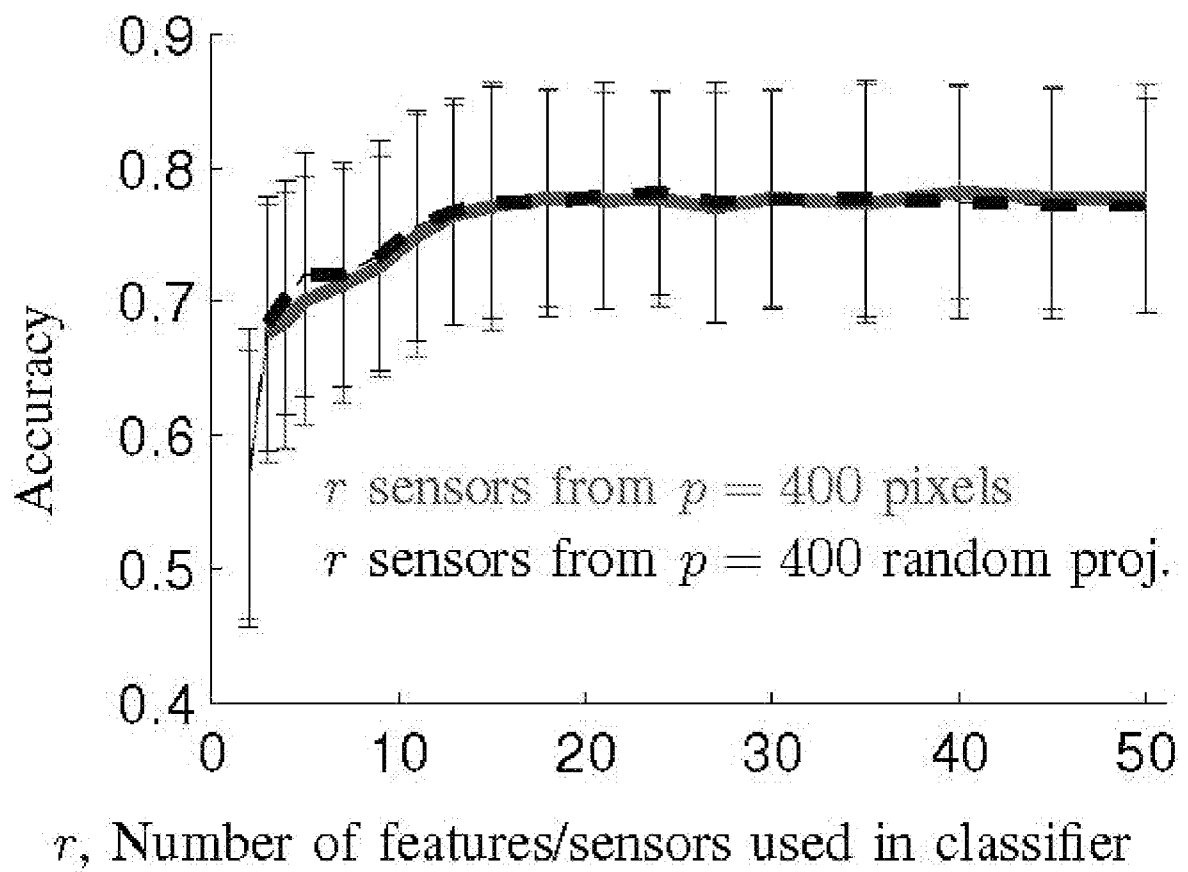


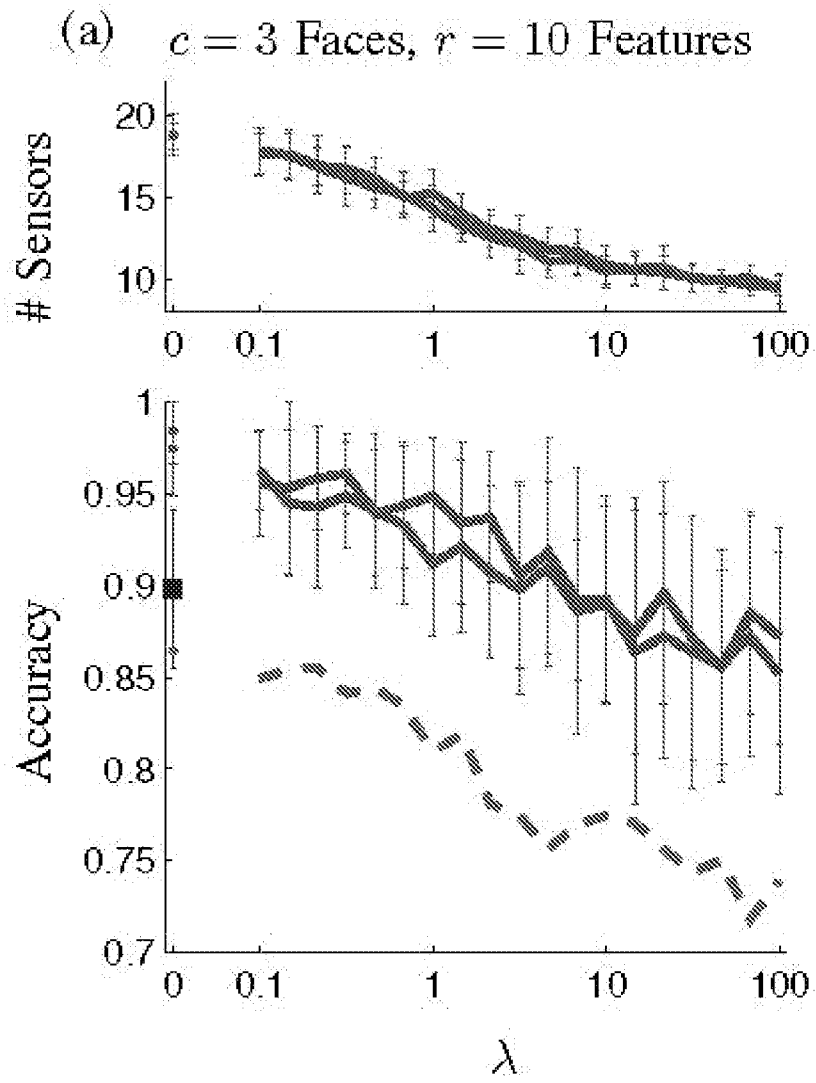
FIG. 6

FIG. 6

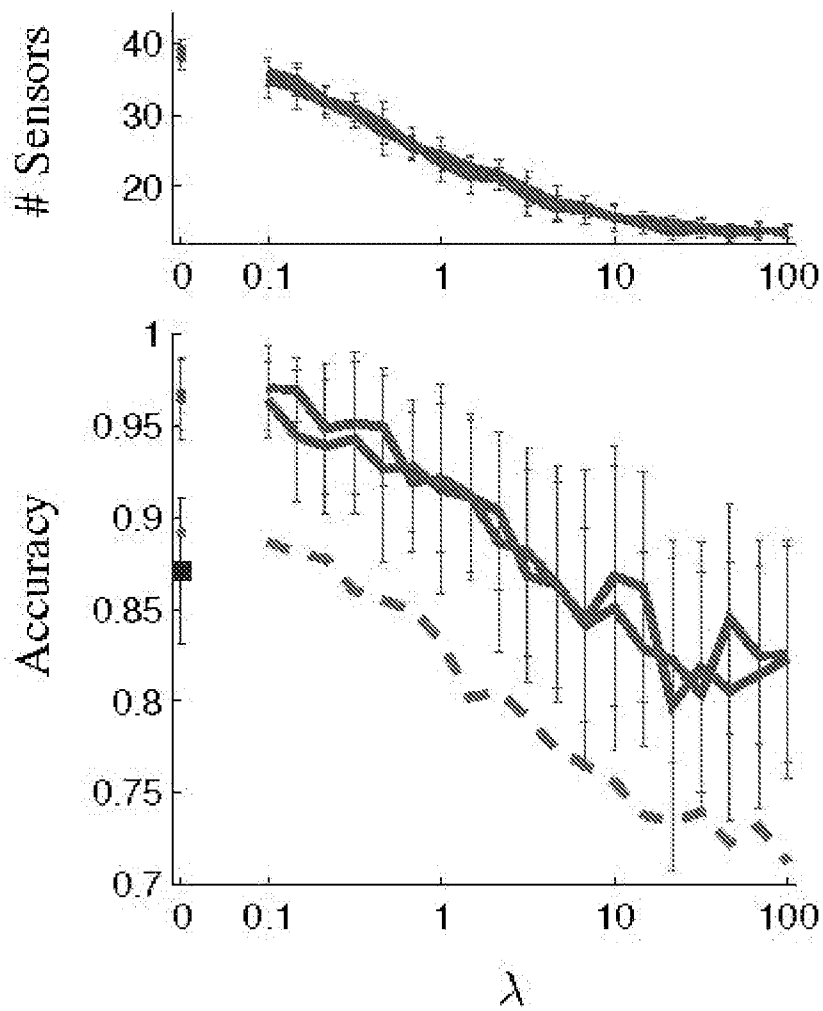
(b) $c = 4$ Faces, $r = 14$ Features

FIG. 6

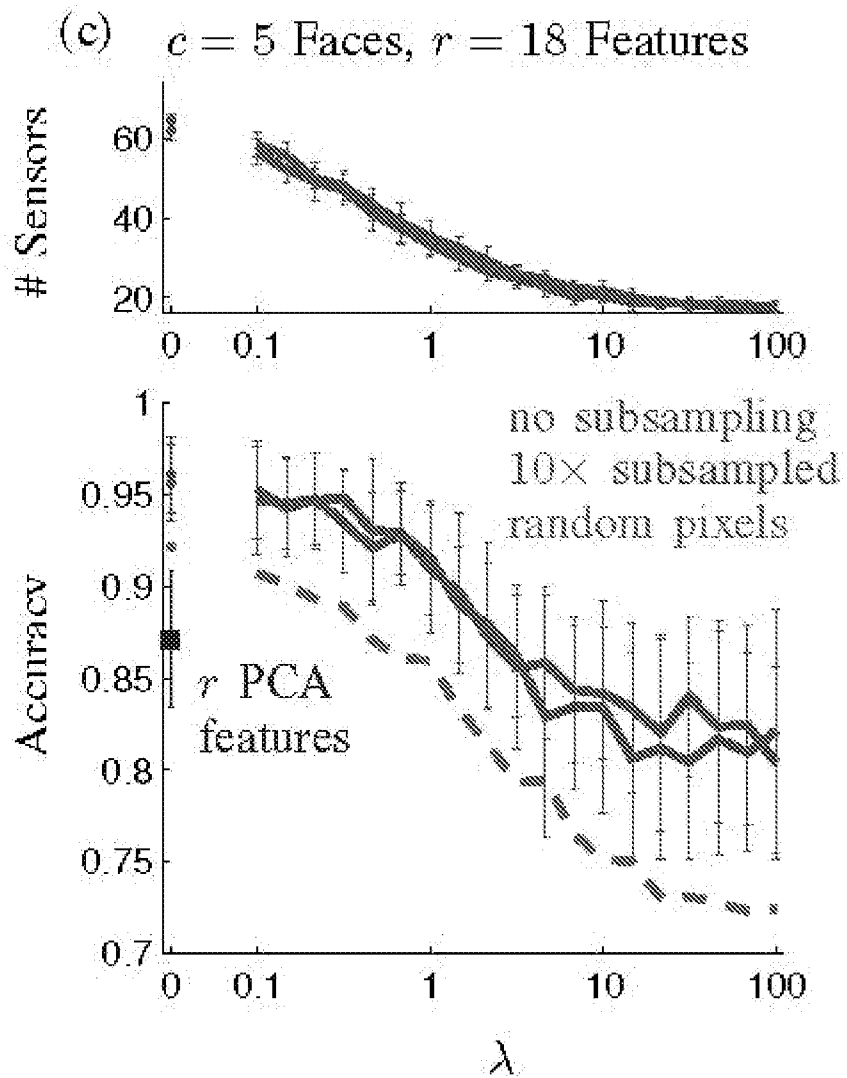


FIG. 7

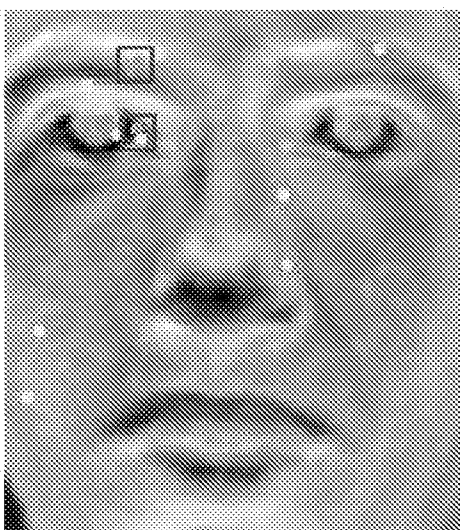
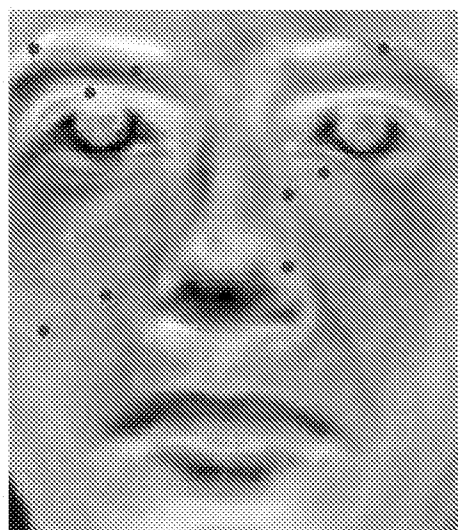
(a) $\lambda = 0$ (b) $\lambda = 0.29$ 

FIG. 8

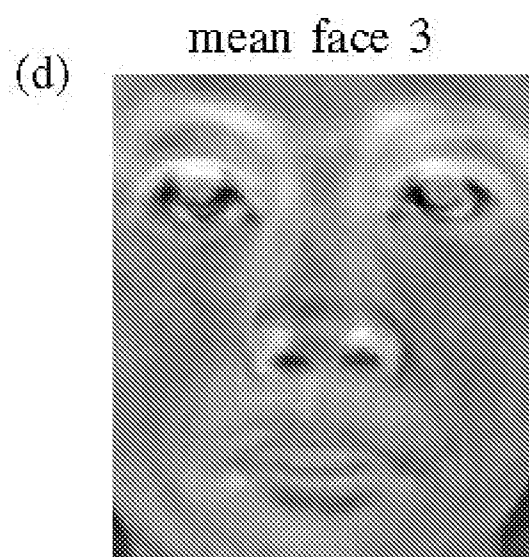
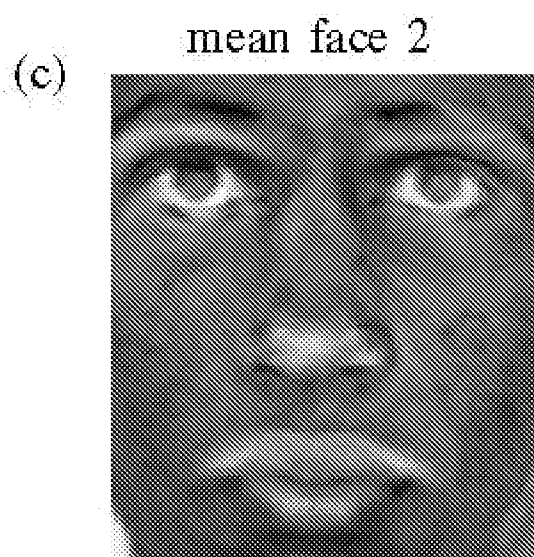
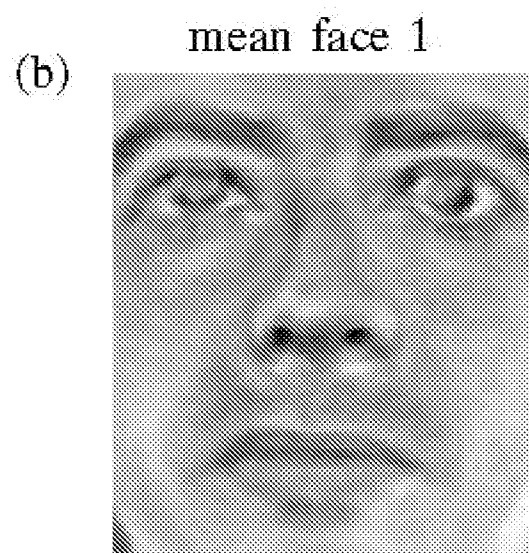
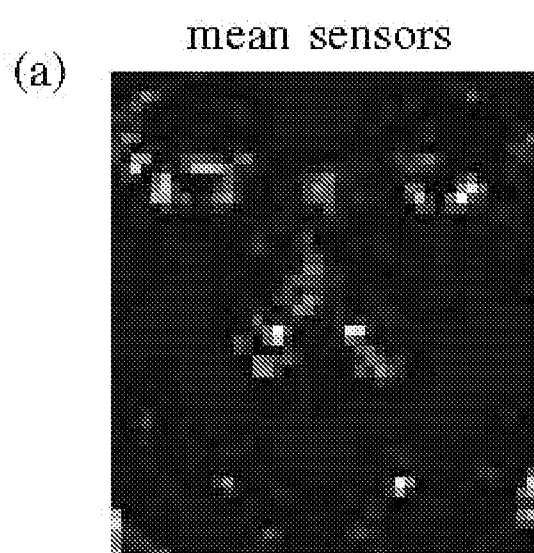


FIG. 9

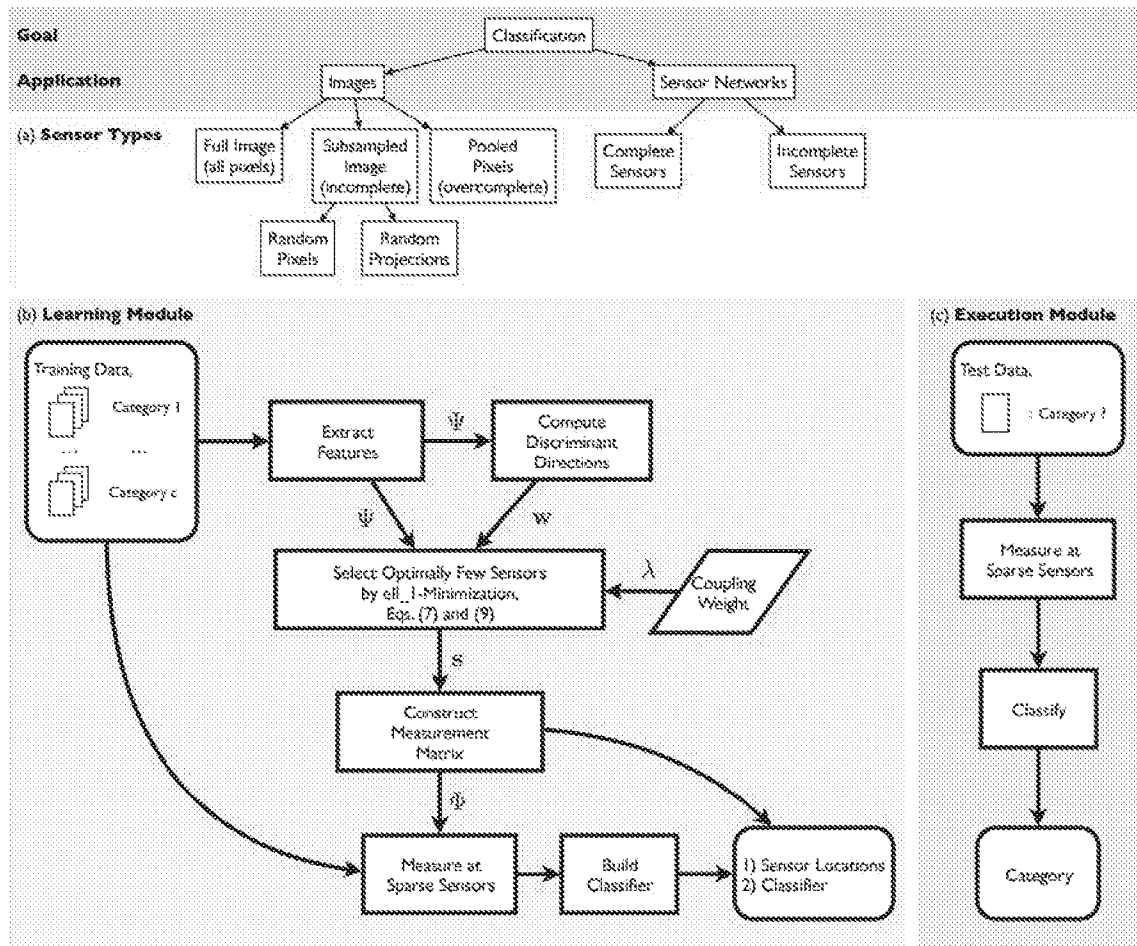


FIG. 10

(a)

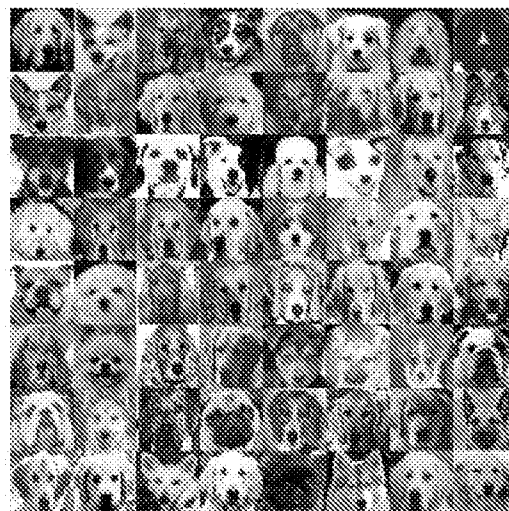
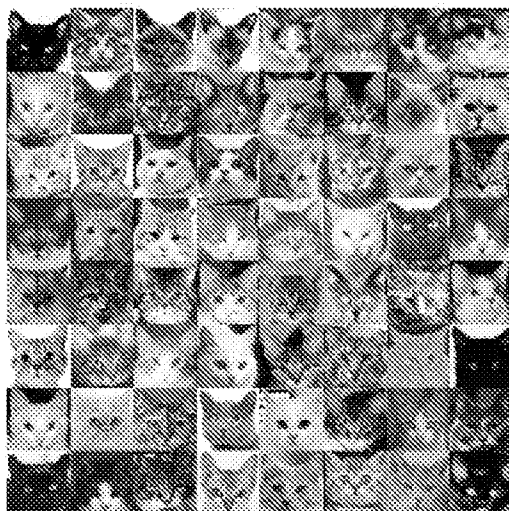


FIG. 10

(b)

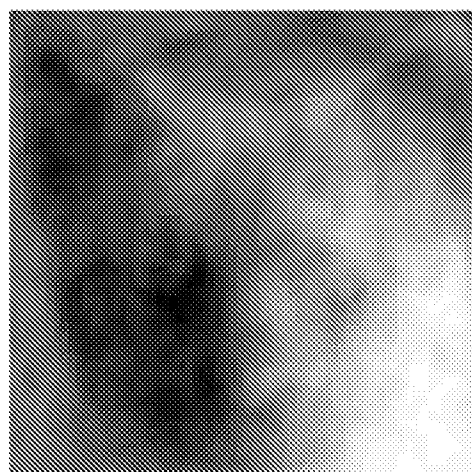
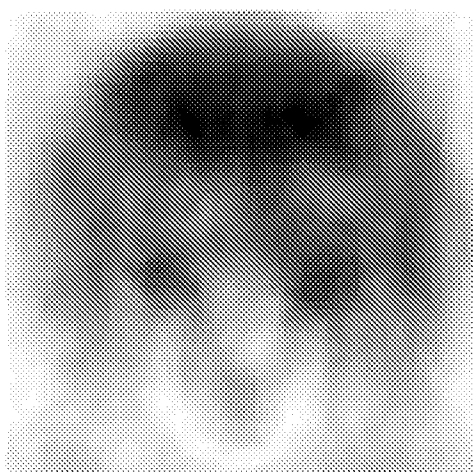
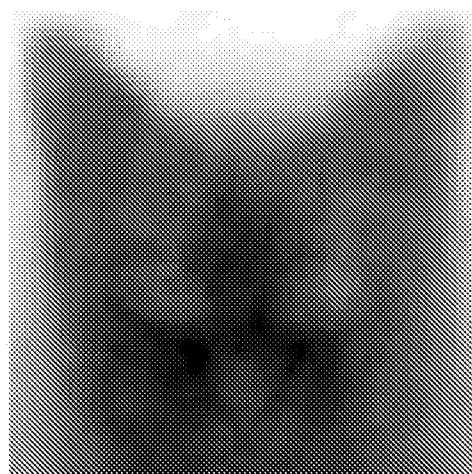
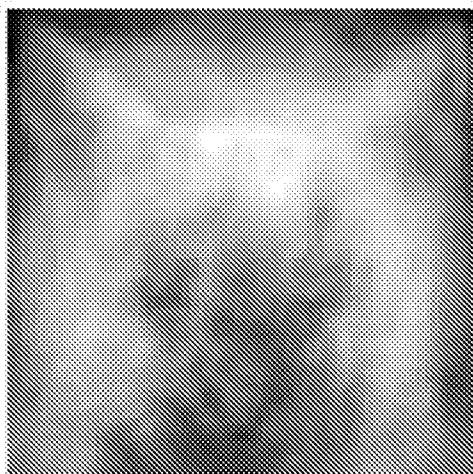


FIG. 11

(a)

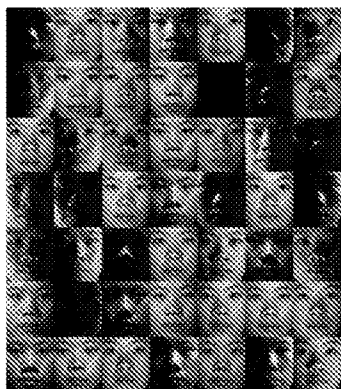


FIG. 11

(b)

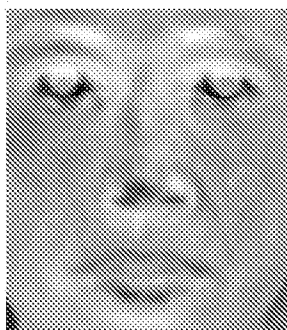
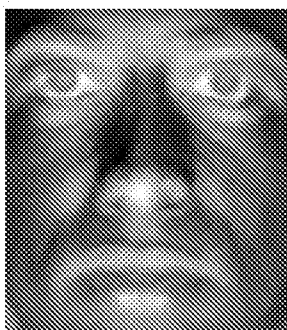


FIG. 12

1. Grids
 - Epidemiological monitoring
 - Power electrical grid
 - Internet
 - Transportation of goods and services
2. Medical diagnostics
 - Genetics
 - Epigenetics
 - Metabolites
 - Natural Sensing for prosthetics
3. Engineering diagnostics
 - Process control in manufacturing
 - Acoustic tone
4. Other, miscellaneous
 - RGB-D images
 - Questionnaire refinement

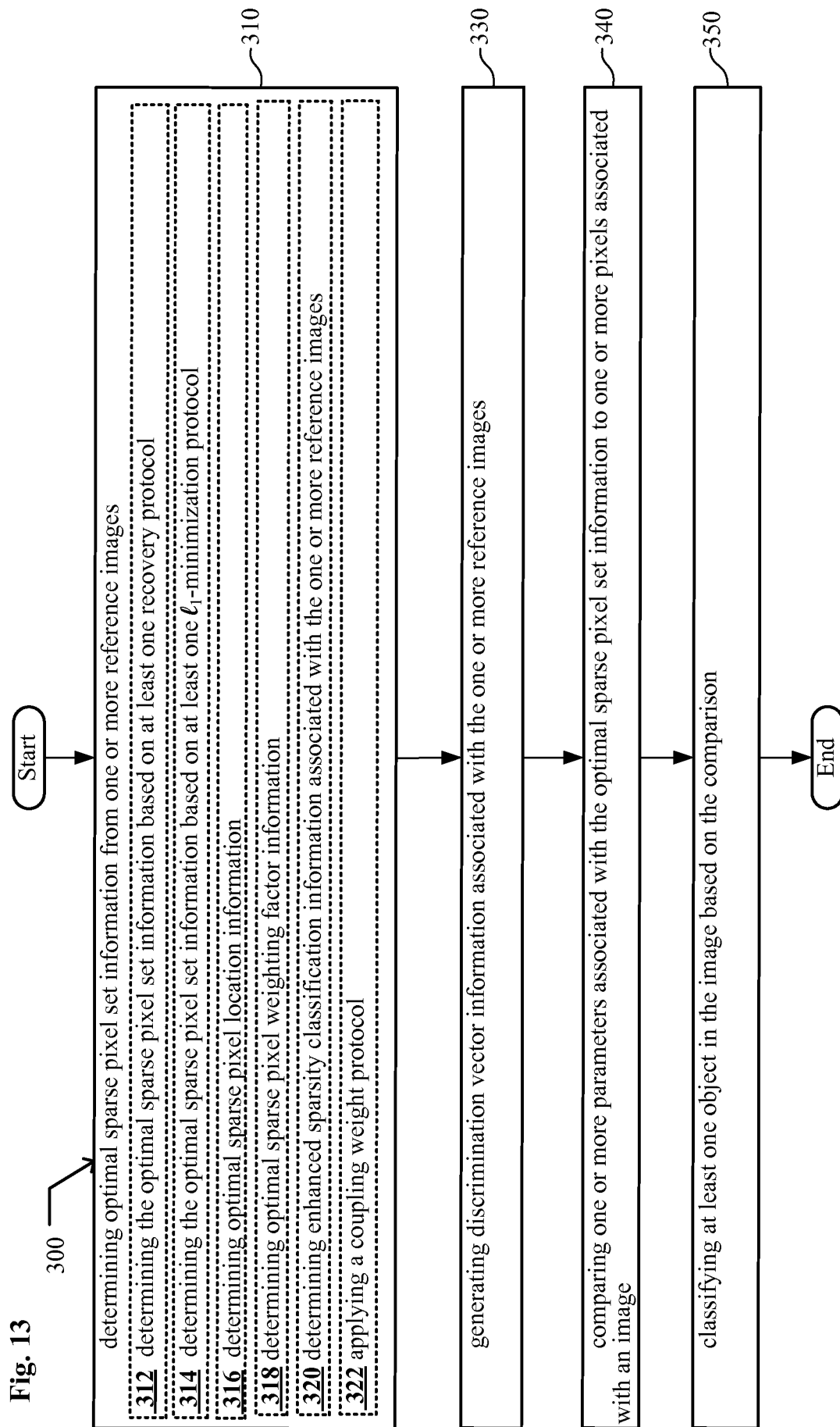
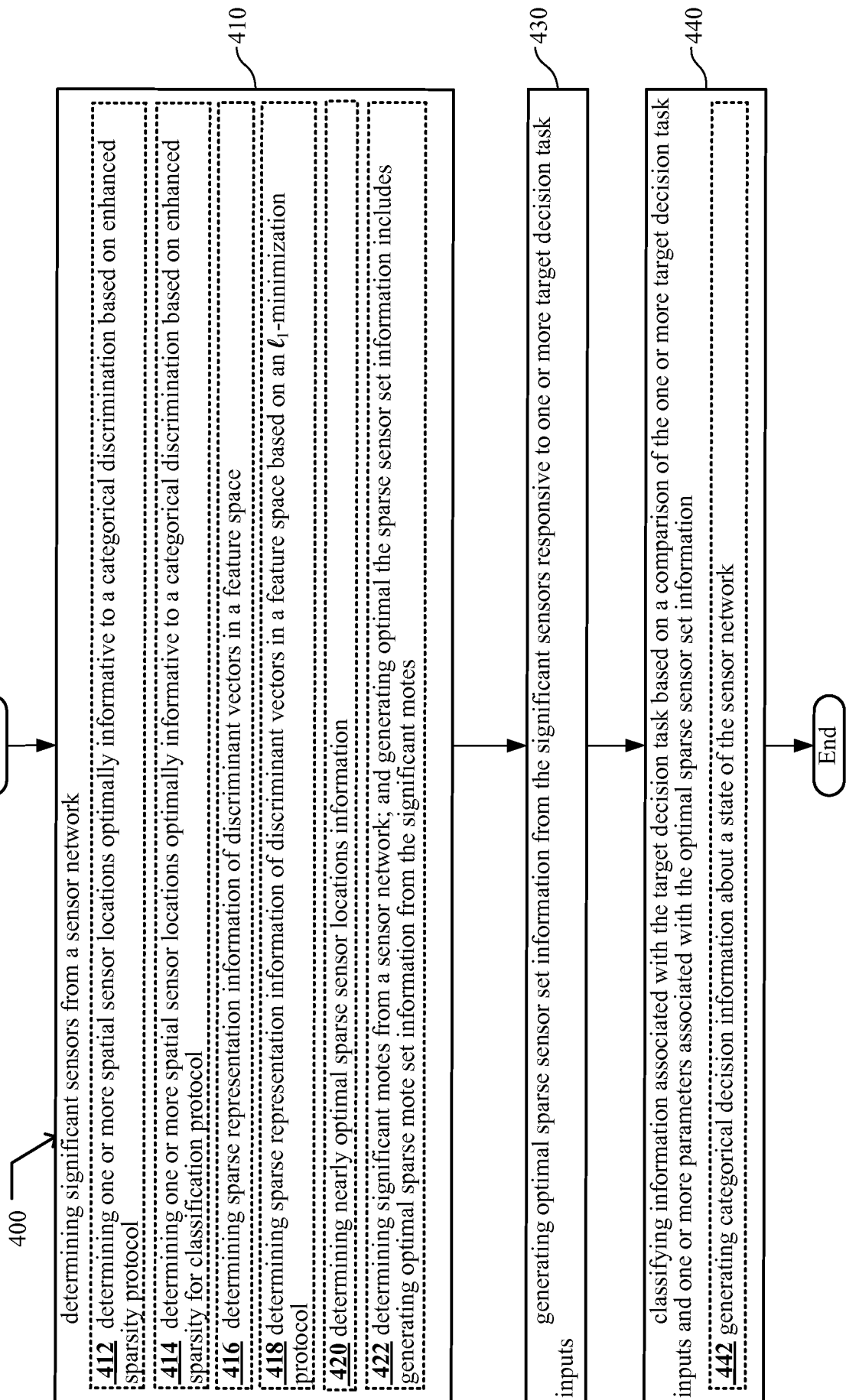


Fig. 14



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/057370**A. CLASSIFICATION OF SUBJECT MATTER****G06K 9/36(2006.01)i, G06K 9/46(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K 9/36; G06K 9/68; G06K 9/46; G08G 1/123; G06F 17/30; G06K 9/00; G06K 9/62; G06T 7/20

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords:object, classification, pixel, sparse, reference, image, comparison

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2011-0286628 A1 (LUIS F. GONCALVES et al.) 24 November 2011 See paragraphs [0027], [0042], [0129]-[0131], [0139], [0147]; and figures 6, 10, 12.	1-33
A		34-62
Y	US 2010-0215254 A1 (DANIL V. PROKHOROV) 26 August 2010 See paragraphs [0027]-[0029], [0033]; claim 1; and figure 5.	1-33
Y	US 2010-0189354 A1 (TEOFILO E. DE CAMPOS et al.) 29 July 2010 See paragraph [0038]; and figure 3	18, 24, 26
A	WO 2013-029674 A1 (METAIO GMBH) 07 March 2013 See page 5, lines 20-34; page 17, lines 21-31; page 18, lines 9-12; and figure 3.	1-62
A	JP 2002-183206 A (MITSUBISHI ELECTRIC CORP) 28 June 2002 See paragraphs [0009]-[0010], [0050]; and figures 2-3, 6.	1-62

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

29 December 2014 (29.12.2014)

Date of mailing of the international search report

29 December 2014 (29.12.2014)

Name and mailing address of the ISA/KR



International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

KANG, Sung Chul

Telephone No. +82-42-481-8405



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/057370

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011-0286628 A1	24/11/2011	CN 103003814 A EP 2569721 A2 EP 2569721 A4 WO 2011-143633 A2 WO 2011-143633 A3	27/03/2013 20/03/2013 27/11/2013 17/11/2011 16/02/2012
US 2010-0215254 A1	26/08/2010	US 8254670 B2	28/08/2012
US 2010-0189354 A1	29/07/2010	EP 2230628 A2 EP 2230628 A3 US 8774498 B2	22/09/2010 09/03/2011 08/07/2014
WO 2013-029674 A1	07/03/2013	CN 103959308 A EP 2751742 A1 TW 201319955 A	30/07/2014 09/07/2014 16/05/2013
JP 2002-183206 A	28/06/2002	JP 4215386 B2	28/01/2009