(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 February 2013 (28.02.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/029008 A1

(51) International Patent Classification: *G06K 9/20* (2006.01)

(21) International Application Number: PCT/US2012/052348

(22) International Filing Date: 24 August 2012 (24.08.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/527,493 25 August 2011 (25.08.2011) US
61/657,406 8 June 2012 (08.06.2012) US

(71) Applicant (*for all designated States except US*): CORNELL UNIVERSITY [US/US]; 395 Pine Tree Road, Suite 310, Ithaca, New York 14850 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): NIRENBERG, Sheila [US/US]; 430 East 63rd Street, Apt. 2-G, New York, New York 10065 (US). BOMASH, Illya [US/US]; 46 Wyckoff Street, Floor 4, Brooklyn, New York 11201 (US).

(74) Agents: EWING, James F. et al.; Foley & Lardner, LLP, 111 Huntington Avenue, Boston, MA 02199 (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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, 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, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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

(54) Title: RETINAL ENCODER FOR MACHINE VISION

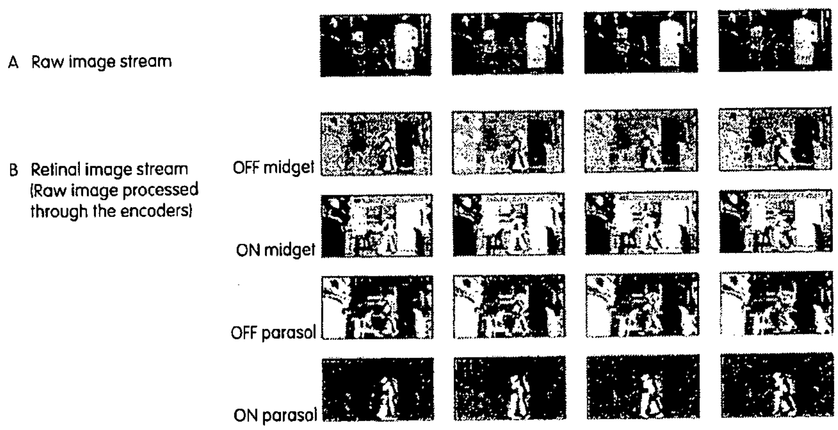


FIG. 3A

(57) Abstract: A method is disclosed including: receiving raw image data corresponding to a series of raw images; processing the raw image data with an encoder to generate encoded data, where the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a vertebrate retina; and applying a first machine vision algorithm to data generated based at least in part on the encoded data.

WO 2013/029008 A1

RETINAL ENCODER FOR MACHINE VISION

Cross Reference to Related Application

This application claims the benefit of U.S. Provisional Application Nos. 61/527493 (filed August 25, 2011) and 61/657406 (filed June 8, 2012). The contents of each of the forgoing applications are incorporated by reference in their entirety.

This application is also related to U.S. Provisional Application Nos. 61/308,681 (filed on February 26, 2010), 61/359,188 (filed on June 28, 2010), 61/378,793 (filed on August 31, 2010), and 61/382,280 (filed on September 13, 2010); to U.S. Patent Application 13/230,488, (filed on September 12, 2011); and to International Patent Application Nos. PCT/US2011/026526 (filed on February 28, 2011) and PCT/US2011/049188 (filed August 25, 2011). The contents of each of the forgoing applications are incorporated by reference in their entirety.

Statement Regarding Federally Sponsored Research or Development

This invention was made with U.S. Government support under is R01 EY12978 awarded by the National Eye Institute of the National Institute of Health (NIH). The U.S. Government has certain rights in the invention.

Field

The present disclosure relates to methods and devices for use in machine vision. In particular, the present disclosure relates to methods and devices for processing images using encoders that mimic the performance of an animal retina, and using the processed images in machine vision applications.

Background

Machine vision (or computer vision) refers to technology that allows a computer to use visual information, e.g., to extract information from an image, to solve some task, or perhaps "understand" the scene in either a broad or limited sense. In general, machine vision is concerned with the extraction of information from image data. The image data can take many forms, such as single images, video sequences, views from multiple cameras, or higher dimensional data (e.g., three dimensional images from a medical scanner).

Machine vision has numerous applications, ranging from relatively simple tasks, such as industrial systems used to count objects passing by on a production line, to more complicated tasks such as facial recognition, and perceptual tasks (e.g., to allow robots to navigate complex environments). A non-limiting list of examples of applications of machine vision include systems for controlling processes (e.g., an industrial robot or an autonomous vehicle), detecting events (e.g., for visual surveillance or people counting), organizing information (e.g., for indexing databases of images and image sequences), modeling objects or environments (e.g., industrial inspection, medical image analysis or topographical modeling), and interaction (e.g., as the input to a device for computer-human interaction).

In many applications, machine vision involves highly computationally expensive tasks. A single color digital image may be composed of millions of pixels or more, each pixel having an associate value, such as a multiple (e.g., 8 or 24) bit value defining the coordinates of the pixel in a color space (e.g., the familiar RGB color space, the YCbCr space, the HSV space, etc.). Video streams may include sequences of such images at frame rates of, e.g., dozens of frames per second, corresponding to bit rates of hundreds of megabits per second or more. Many machine vision applications require quick processing of such images or video streams (e.g., to track and react to the motion of an object, to identify or classify an object as it moves along an assembly line, to allow a robot to react in real time to its environment, etc.).

Processing such a large volume of data under such time constraints can be extremely challenging. Accordingly, it would be desirable to find techniques for processing image data to reduce the raw amount of information while retaining (or even accentuating) the features of the image data that are salient for the machine vision task at hand. This pre-processed image data, rather than the raw data, could then be input to a machine vision system, reducing the processing burden on the system and allowing for sufficiently speedy response and potentially improved performance.

It has been recognized that the retina of the vertebrate eye provides image processing of this just this nature, taking in a visual stimulus and converting the stimulus into a form that can be understood by the brain. This system (developed over the course of millions of years of evolution) is remarkably efficient and effective, as evidenced by high level of complex visual perception in mammals (particularly monkeys and humans).

Several approaches have been proposed for developing image data pre-processing schemes for machine vision based on abstract models of the operations of the retina. However, these models have been based on rough approximations to the actual performance of the retina.

Portions of this Background section are adapted from the Wikipedia article on computer visions available at http://en.wikipedia.org/wiki/Computer_vision and used pursuant to the Creative Commons Attribution-ShareAlike License.

Summary

Embodiments described in the present disclosure utilize an encoder that provides a near-complete replication of the operations performed by the retina. As described in detail in International Patent Applications, incorporated by reference above (henceforth the "Prosthesis Applications") this encoder may be used to develop a highly effective retinal prosthetic. In the present disclosure, the encoder is applied to machine vision.

When used as a preprocessing step (in particular, a dimension-reduction step), the encoder substantially enhances the performance of machine vision algorithms. In some embodiments, the encoder allows the machine vision algorithm to extract information very effectively in a broad range of environments and lighting conditions, including information that could not be extracted by other methods. In cases where existing machine vision algorithms are in part effective, this dimension reduction may serve as a strong enhancer. The encoder may allow the extraction to be carried out more effectively (higher performance), as well as faster and more efficiently.

As described in detail in the Prosthesis Applications the applicants have developed a prosthetic device that receives a stimulus, and transforms the stimulus into a set of codes with a set of encoders, transforms the codes into signals with an interface, which then activate a plurality of retinal cells with a high resolution transducer driven by the signals from the interface. Activation of the plurality of retinal cells results in retinal ganglion cell responses, to a broad range of stimuli, which are substantially similar to the time dependent responses of retinal ganglion cells from a normal retina to the same stimuli. The applicants have realized that the encoders used in such devices may be adapted to process image data for use in machine vision applications.

The retina prosthesis described in the Prosthesis Applications, like the normal retina, is an image processor - it extracts essential information from the stimuli it receives, and reformats the information into patterns of action potentials the brain can understand. The patterns of action potentials produced by the normal retinal are in what is referred to as the retina's code or the ganglion cell's code. The retina prosthesis converts visual stimuli into this same code, or a close proxy of it, so that the damaged or degenerated retina can produce normal or near-normal output. Because the retina prosthesis uses the same code as the normal retina or a close proxy of it, the firing patterns of the ganglion cells in the damaged or degenerated retina, that is, their patterns of action potentials are the same, or substantially similar, to those produced by normal ganglion cells. Thus, this prosthetic allows the retina to send to the brain the same signals about the visual world as the normal retina.

As detailed in the Prosthesis Application, the encoders use input/output models for retinal cells which were generated using data obtained from studies of the input/output response of actual retinal cells to a variety of stimuli, e.g., both white noise (WN) and natural scene (NS) movies. In some embodiments, the encoders are based on a linear nonlinear cascade model that includes a spatiotemporal transformation characterized by a number of parameters. These parameters are optimized based on data obtained through experiments in the real retina, resulting in transformation that closely mimics the response of the actual cells to a broad range of stimuli. The result is a model that captures the input/output relations for natural images (static or spatiotemporally-varying), such as faces, landscapes, people walking, children playing, etc., not just for white noise stimuli or stimuli with Gaussian statistics. The effectiveness on a broad range of stimuli is shown in the Prosthesis Applications, and in Figs. 18A-18F discussed in detail below.

Because this approach leverages data obtained through experiments, the generated encoders can accurately simulate retinal processing, without requiring a detailed abstract understanding of the retina's underlying processing schemes. For example, it is believed that retinal processing in primates and humans highlights features in the visual stimulus useful for pattern recognition tasks (e.g., facial recognition) while de-emphasizing or eliminating other features (e.g., redundant information or noise) to allow for efficient processing in the brain. As of yet, there is no complete abstract understanding of the details of this processing scheme, which developed as the result natural selection over the course of eons. However, despite this lack of abstract understanding, the devices and techniques described herein can capture the benefit of this processing, by accurately mimicking the retinal response.

In other words, in various embodiments described herein, the approach is data-driven – that is, it uses a data-driven model of retinal input/output relations, and thus provide realistic image pre-processing. This gives downstream machine vision algorithms a pre-processing step that accomplishes the same kind and the same magnitude of dimension reduction as the biological retina, and, therefore, offers the same array of advantages as the biological retina.

Note that in general, the approaches described herein differ from previous preprocessors that filter image data with, for example, a difference-of-Gaussians type filter, because they may provide a complete or near complete mimicking of the retina. Similarly, it differs from other linear-nonlinear cascade models in that it is effective on a broad range of stimuli, not just white noise stimuli or stimuli with Gaussian statistics. Thus, the filtering is much more complete, and it greatly enhances the power of current machine vision algorithms. Most importantly, it allows current machine vision algorithms to generalize, i.e., to be trained in one setting (one environment or lighting condition) and generalize to other environments, which has been a long-standing challenge (see e.g., Figs. 10, 11, and 15 as described in detail below).

Moreover, in some embodiments, because the retinal processing is accurately modeled for a broad range of stimuli (e.g., as a result of optimization using both WN- and NS-generated data), the pre-processing for the machine vision system works well over a broad range of conditions (similar to the way the retina works over a broad range of conditions). Advantageously, this allows the retinal preprocessing techniques to be used in machine vision applications that require robust performance under a variety of conditions (e.g., lighting changes, complex, changing visual scenes, many different environments, etc.).

In one aspect, a method is disclosed including: receiving raw image data corresponding to a series of raw images; processing the raw image data with an encoder to generate encoded data, where the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a vertebrate retina; and applying a first machine vision algorithm to data generated based at least in part on the encoded data.

Some embodiments include generating a series of retinal images based on the encoded data. Some embodiments include determining pixel values in the retinal images based on the encoded data. In some embodiments, determining pixel values in the retinal images based on the encoded data includes determining a pixel intensity or color based on encoded data indicative of a retinal cell response.

In some embodiments, the data indicative of a retinal cell response is indicative of at least one from the list consisting of: a retinal cell firing rate, a retinal cell output pulse train, and a generator potential.

Some embodiments include applying the first machine vision algorithm to the series of retinal images.

In some embodiments, the machine vision algorithm includes at least one select from the list consisting of: an object recognition algorithm, an image classification algorithm, a facial recognition algorithm, an optical character recognition algorithm, a content-based image retrieval algorithm, a pose estimation algorithm, a motion analysis algorithm, an egomotion determination algorithm, a movement tracking algorithm, an optical flow determination algorithm, a scene reconstruction algorithm, a 3D volume recognition algorithm, and a navigation algorithm.

In some embodiments, the machine vision algorithm exhibits better performance when applied to the series of retinal images than when applied to a corresponding set of raw images that have not been processed using the encoder.

In some embodiments, the machine vision algorithm exhibits better performance when applied to a series of retinal images including natural scenes than when applied to a corresponding series of raw images that have not been processed using the encoder.

In some embodiments, the machine vision algorithm includes an algorithm for the detection or identification of a human within a series of images; and where the machine vision algorithm exhibits better detection or identification accuracy when applied to a range of retinal images including the human than when applied to a corresponding set of raw images that have not been processed using the encoder.

In some embodiments, the series of images includes the human includes images of the human located in a natural scene.

In some embodiments, the series of images including the human includes images of the human located in a natural scene that is different from natural scenes used to train the machine vision algorithm.

In some embodiments, the machine vision algorithm includes an algorithm for navigation through a real or virtual environment, and where the machine vision algorithm exhibits better navigation performance when applied to a series of retinal images including a natural scene than when applied to a corresponding set of raw images that have not been processed using the encoder.

In some embodiments, the machine vision algorithm exhibits fewer unwanted collision events during navigation when applied to a series of retinal images including a natural scene than when applied to a corresponding set of raw images that have not been processed using the encoder.

In some embodiments, the series of retinal images correspond to an environment that was not used to train the machine vision algorithm.

Some embodiments include applying a machine imaging algorithm to the series of retinal images to identify one or more retinal images of interest; and identifying one or more raw images of interest corresponding to the retinal images of interest. Some embodiments include processing the raw images of interest. In some embodiments, processing the raw images of interest includes applying a second machine vision algorithm to the raw images of interest. In some embodiments, the first machine vision algorithm includes an algorithm that has been trained on a set of retinal images; and the second machine vision algorithm includes an algorithm that has been trained on a set of raw images.

In some embodiments, applying the first machine vision algorithm includes applying a navigation algorithm. In some embodiments, applying the navigation algorithm includes: processing the series of retinal images to determine motion information indicative of motion at a plurality of image locations in the series of images; classifying spatial regions in the series of images based on the motion information; and generating a navigation decision based on the classification of the spatial regions. In some embodiments, the motion information is indicative of an optical flow in the series of images. Some embodiments include using a convolutional neural network to classify the spatial regions.

Some embodiments include controlling the motion of a robotic apparatus based on results from navigation algorithm.

Some embodiments include controlling the motion of a virtual object in a virtual space based on results from navigation algorithm.

Some embodiments include training a machine vision algorithm based on the retinal images. In some embodiments, training the machine vision algorithm includes: (i) applying the machine vision algorithm to a set of retinal images to generate an output; (ii) determining performance information indicative of the performance of the machine vision algorithm based on the output; and (iii) modifying one or more characteristics of the machine vision algorithm based on the performance information. Some embodiments include iteratively repeating steps (i) through (iii) until a selected performance criteria is reached.

In some embodiments, the trained machine vision algorithm is characterized by a set of parameters, and where the parameters differ from the corresponding parameters that would be obtained by equivalent training of the machine vision algorithm using raw images corresponding to the retinal images.

In some embodiments, processing the raw image data with an encoder to generate encoded data includes generating encoded data that contains a reduced amount of information relative to the corresponding raw image data. In some such embodiments, the machine vision algorithm exhibits better performance when applied to the series of retinal images than when applied to a corresponding set of raw images that have not been processed using the encoder.

In some embodiments, the amount of information contained in the encoded data is compressed by a factor of at least about 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more, e.g. in the range of 1.1 – 1,000 or any subrange thereof, relative to the corresponding raw image data.

In some embodiments, the vertebrate includes at least one selected from the list consisting of: a mouse, and a monkey.

In some embodiments, the retinal cells include ganglion cells. In some embodiments, the retinal cells include at least two classes of cells. In some embodiments, the at least two classes of cells includes ON cells and OFF cells.

In some embodiments, the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a vertebrate retina over a range of input that includes natural scene images, including spatio-temporally varying images.

In some embodiments, processing the raw image data with an encoder to generate encoded data includes: processing the raw image data to generate a plurality of values, X, transforming the plurality of X values into a plurality of response values, λ_m, indicative of a corresponding response of a retinal cell in the retina, m, and generating the encoded data based on the response values. In some embodiments, the response values correspond to retinal cell firing rates. In some embodiments, the response values correspond to a function of the retinal cell firing rates. In some embodiments, the response values correspond to retinal cell output pulses. In some embodiments, the response values correspond to retinal cell generator potential, i.e., the output of the convolution of the image with the spatiotemporal filter(s).

In some embodiments, processing the raw image data with an encoder to generate encoded data includes: receiving images from the raw image data and, for each image, rescaling the luminance or contrast to generate a rescaled image stream; receiving a set of N rescaled images from the rescaled image stream and applying a spatiotemporal transformation to the set of N images to generate a set of retinal response values, each value in the set corresponding to a respective one of the retinal cells; generating the encoded data based on the retinal response values.

In some embodiments, the response values include retina cell firing rates. In some embodiments N is at least 5, at least about 20, at least about 100 or more, e.g., in the range of 1-1,000 or any subrange thereof.

In some embodiments, applying a spatiotemporal transformation includes: convolving of the N rescaled images with a spatiotemporal kernel to generate one or more spatially-temporally transformed images; and applying a nonlinear function to the spatially-temporally transformed images to generate the set of response values.

In some embodiments, applying a spatiotemporal transformation includes: convolving the N rescaled images with a spatial kernel to generate N spatially transformed images; convolving the N spatially transformed images with a temporal kernel to generate a temporal transformation output; and applying a nonlinear function to the temporal transformation output to generate the set of response values.

In some embodiments, the encoder is characterized by a set of parameters, and where the values of the parameters are determined using response data obtained experimentally from a vertebrate retina while said retina is exposed to white noise and natural scene stimuli.

In some embodiments, the encoder is configured such that the Pearson's correlation coefficient between a test input stimulus and a corresponding stimulus reconstructed from the encoded data that would be generated by the encoder in response to the test input stimulus is at least about 0.35, 0.65, at least about 0.95, or more, e.g., in the range of 0.35-1.0 or any subrange thereof. In some embodiments, the test input stimulus includes a series of natural scenes.

In another aspect, an apparatus is disclosed including: at least one memory storage device configured to store raw image data; at least one processor operably coupled with the memory and programmed to execute one or more of the methods described herein.

In some embodiments, a non-transitory computer-readable medium having computer-executable instructions for implementing the steps of one or more of the methods described herein.

In another aspect, a system is disclosed including: at least one memory storage device storing encoded data corresponding to a series of images, where the encoded data has been generated by: receiving raw image data corresponding to a series of raw images; and processing the raw image data with an encoder to generate encoded data, where the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a vertebrate retina. In some embodiments, the at least one storage device stores database information indicative of a correspondence between the encoded data and the raw image data.

Some embodiments include a processor configured to: receive query image data corresponding to a series of query images; process the query image data with an encoder to generate encoded data, where the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a vertebrate retina; compare the encoded query image data to the encoded data on the memory storage device; and based on (a) the comparison of the encoded query data to the encoded data on the memory storage device, and (b) the database information indicative of a correspondence between the encoded data and the raw image data, determine a correspondence between the query image data and the raw image data.

In another aspect, a method is disclosed including: receiving raw image data corresponding to a series of raw images; processing at least a first portion of the raw image data with an encoder to generate first encoded data, where the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a first vertebrate retina from a first vertebrate type; and processing at least a second portion of the raw image data with an encoder to generate encoded data, where the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a second vertebrate retina from a second vertebrate type different from the first vertebrate type.

Some embodiments include based on the first encoded data, selecting the second portion of the raw image data for processing.

In various embodiments, the raw image data is received in substantially real time from an image detector or from a memory that stores the raw image data, or from a combination thereof.

In another aspect, an apparatus is disclosed including: at least one memory storage device configured to store raw image data; at least one processor operably coupled with the memory and programmed to execute one or more of the methods described herein.

In another aspect, a non-transitory computer-readable medium having computer-executable instructions for implementing the steps of one or more of the methods described herein.

In another aspect, a system is disclosed including: at least one memory storage device storing encoded data corresponding to a series of images, where the encoded data has been generated by: receiving raw image data corresponding to a series of raw images; and processing the raw image data with an encoder to generate encoded data, where the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a vertebrate retina. In some embodiments, the at least one storage device stores database information indicative of a correspondence between the encoded data and the raw image data.

Various embodiments may include any of the above described elements, alone or in any suitable combination.

Brief Description of the Drawings

Fig. 1 is a block diagram showing an exemplary machine vision system.

Fig. 2 is a flow chart illustrating the operation of an encoder module.

Fig. 3A illustrates the conversion of a raw image stream (a person walking through a complex environment) into a retinal image stream. Panel A shows several frames from the raw image stream, which was acquired by a camera. Panel B shows several frames from the corresponding retinal image stream. Four different retinal image streams are shown, each using a different array of cells (OFF midget cells, ON midget cells, OFF parasol cells, and ON parasol cells, as indicated on figure).

Figs. 3B-3F show enlarged views of the raw image (Fig. 3B) and retinal images Figs. 3C-3F corresponding to the last column of Fig. 3A.

Fig. 4 is a block diagram showing a training system for training the machine vision module of the machine vision system of Fig. 1.

Fig. 5 is a flowchart illustrating the operation of the training system of Fig. 4.

Fig. 6 illustrates a machine vision system used to control the navigation of a robot through a maze. The path traveled by the robot is indicated with a dashed line.

Fig. 7 is a flow chart for one embodiment of a machine vision system used to control a navigation task.

Fig. 8 shows frames from the raw image streams (movies) used to train the navigator. These image streams were generated in a virtual environment using a rural environment as indicated in the main text. The top panel shows the first 5 frames in the image stream. The bottom panel shows selected frames from the rest of the image stream; one of every 30 frames (that is, one frame per second) is shown.

Fig. 9 shows frames from the raw image streams (movies) used to test the navigator. Three sets are shown: **A**, frames from a rural environment (one different from that used to train the navigator); **B**, a suburban environment; and **C**, a playground environment (a tire obstacle course). As in Fig. 9, the image streams were generated in a virtual environment, the top panel of each set shows the first four frames, and the bottom panel shows selected frames from the rest of the movies (in this case, one of every 15 frames (that is, one frame every half-second).

Fig. 10 illustrates trajectories showing the performance of the navigator and its ability to generalize to different environments. As described in the text and in flow chart in Fig. 7, the leading algorithm used to learn navigation tasks, the convolutional neural network (CNN), was trained two ways: 1) the standard way, i.e., using the raw visual environment (the raw image streams), and 2) using the environment after it had its dimension reduced, i.e., after it was processed through the encoder. (The training environment used was a rural environment, as shown in Fig. 8). The performance of the navigator was then tested in 3 new environments: a rural environment that was different from the one used to train the navigator, a suburban environment, and a playground environment. (Samples from each environment are shown in Fig. 9.) **A.** The navigator's performance when it learned the environment from the raw image stream. Note the disorganized trajectories and collisions. **B.** The navigator's performance when it learned the environment from the *retinal* image stream (the image stream produced by the encoder). Note the straight paths and obstacle avoidance.

Fig. 11 shows further demonstration of the navigator's high performance; specifically, it shows that the high performance generalizes not just to different environments (from rural environment to suburban environment to playground), but it also generalizes to different lighting conditions *within* an environment. A through F correspond to different positions of the sun, and therefore, different shadow conditions in the playground environment; the light conditions span sunrise to sunset, i.e., 30 degrees above the horizontal on the left side of the environment to 30 degrees above the horizontal on the right side. *Light gray*, the performance of the navigator when it was trained on raw image streams (from the rural environment using one lighting condition, as shown in Fig. 8). As shown here, the performance of the navigator is low when it is placed in a new environment, and this remains true across light conditions. The height of each bar corresponds to the fraction of trials in which the navigator successfully stayed within the playground tire course without colliding with one of the tires. Error bars indicate the standard error of the mean (SEM). *Dark grey*, the performance of the navigator when it was trained on the *retinal* image streams (same rural environment using same single lighting condition, but this time processed through the encoder). As shown, the performance of the navigator is high, and the high performance holds across light conditions. Thus, training on the retinal image streams (i.e., training on the dimension-reduced images produced by the encoder) leads to high performance that generalizes both to new environments and to multiple lighting conditions (sunrise to sunset, see above).

Fig. 12 is a flow chart for one embodiment of a machine vision system used to control a face recognition task.

Fig. 13 shows frames from a raw image stream (movie) used to train the face recognition algorithm (the Viola-Jones-Snow algorithm as mentioned in the main text). The image stream was recorded at a rate of 24 frames per second; here, every 12th frame is shown (one frame every half-second).

Fig. 14 shows frames from a raw image stream (movie) used to *test* the face recognition algorithm's performance. Note that this is the same person as shown in Fig, 13, but in a different environment with different hairstyle, etc. As indicated in the main text, the goal of the face recognition algorithm is to recognize new image streams as belonging to the target person, even though the algorithm was only trained on other images streams of this person). As in Fig. 13, the image stream was recorded at a rate of 24 frames per second; here, every 12th frame is shown (one frame every half-second).

Fig. 15 shows the performance of the face recognition algorithm when it was trained two ways: 1) using the standard approach, i.e., training it with raw image streams, and 2) using the approach described in this application (that is, using the raw image streams processed by the encoder). In both cases, the face recognition algorithm was trained on many image streams (250-800 two-frame image streams from 4-5 videos of the target face and 2000 two-frame image streams from >100 videos of others faces). Performance was then measuring using 50-800 two-frame image streams from a previously unseen video, that is, a video not used in the training set. (See Figs. 13 and 14 for sample frames from both the training and testing sets.) Performance is shown for two sets of tasks, one where the standard approach performs very weakly, and one where it performs moderately well. The height of the bars indicates the fraction of trials in which the face recognizer successfully recognized the target face. Error bars indicate the standard error of the mean (SEM). As shown, when the task was challenging (A), the approach described in this application, provides a major (4-fold) improvement, over the standard approach. When the task was less challenging, i.e., when the standard approach performs moderately well, the approach described in this application still provides improvement (by a factor of 1.5).

Fig. 16 shown a process flow for an exemplary hybrid image processing method using both a retinal encoder approach and a traditional approach to image processing

Fig. 17 is a block diagram of a system for digital fingerprinting using retinal encoded data.

Figs. 18A-18F illustrate the performance of a retinal encoder models when tested with movies of natural scenes. In each figure, the performance of a conventional linear-nonlinear (LN) model is shown on the left, and the performance of the linear-nonlinear (LN) model of the type described in this application is shown on the right. Performance is shown via raster plots and peri-stimulus time histograms (PSTHs).

Detailed Description

Fig. 1 shows an exemplary machine vision system 100 featuring a camera 102, an encoder module 104, a machine vision module 106, and a system 108 controlled by the machine vision module. The camera 102 receives visual stimulus and converts it to digital image data e.g., a stream of digital images. This digital image data may be referred to herein as a "raw" image data. It is to be understood that raw image data may include any image data prior to processing by a retinal encoder.

The encoder module 104 receives the image data and processes the data using one or more retinal encoders of the type described herein and/or in the Prosthesis Applications. The output of the encoder module, referred to as "retinal image data" is passed to the machine vision module, which processes the retinal image data, e.g., using one or more machine vision techniques know in the art and/or described herein. Based on the machine vision processing, the machine vision module 106 generates output that that may be used for any suitable purpose. As shown, the output controls one or more systems 108, e.g., a robotic system. In some embodiments the image processing and/or control may be performed in real time or near real time.

It is to be understood that the system shown in Fig. 1 is exemplary only, and various other types of machine vision systems may be used. For example, in some embodiments, the controlled system 108 may be absent, e.g., where the output of the machine vision module is stored, output for further processing, etc., rather than used for control. In some embodiments, the camera 102 may be replaced, e.g., by a source of stored image data. In some embodiments additional elements may be included, e.g., various processors or controller, user controls, input or output devices, etc.

In various embodiments, the camera 102 may be any device capable of converting visual stimulus to a digital form, e.g., a stream of digital images. Various embodiments may include devices based on charge-coupled devices (CCDs); active pixel sensors (APS) such as complimentary metal-oxide-semiconductor (CMOS) sensors, thin-film transistors (TFTs), arrays of photodiodes; and the combinations thereof.

The digital images generated by the camera 102 may each include at least 0.01 megapixels, at least 0.1 megapixels, at least 1 megapixel, at least 2 megapixels, or more, e.g., in the range of 0.01-1000 megapixels or any subrange thereof. The stream of digital images may be characterized by a frame rate (i.e., the number of image frames per second) of at least 10 Hz, at least 50 Hz, at least 100 Hz, or more, e.g., in the range of 1-1000 Hz or any subrange thereof. The digital images may be color, grayscale, black and white, or other suitable types of images.

In some embodiments, the camera is based around a charge-coupled device (CCD). In one embodiment, the camera 100 is a Point Grey Firefly MV device (capable of 752x480 pixels, 8 bits/pixel, at 60 frames per second) (Point Grey Research, Richmond, BC, Canada). In another embodiment, the camera 100 is an E-consystems e-CAM50_OMAP_GSTIX, which integrates an Omnivision OV5642 camera module, capable of 1280x720 pixels, 8 bits/pixel, at 30 frames per second).

In some embodiments, images are acquired by the camera 102 and transmitted to the encoder module 104 with sufficient speed to allow the device 100 to operate without undesirable lag times. To accomplish this, in some embodiments, a high bandwidth connection is provided between the camera 102 and the encoder module 104. For example, a data transfer of greater than 20 MB/sec can be achieved using a USB 2.0 interface between the camera and the processing device. In other embodiments, a parallel interface is used between the camera and the processing device, such as the parallel interface integrated into the Camera Image Signal Processor on the OMAP 3530 processor (Texas Instruments, Dallas, TX). In various embodiments, other suitable connections may be used, including wired or wireless connections. The camera 102 can be interfaced with the encoder module 104 using any connection capable of high speed data transfer, including, but not limited to, serial interfaces, such as IEEE 1394 or USB 2.0; parallel interfaces; analog interfaces, such as NTSC or PAL; a wireless interface. In some embodiments, the camera could be integrated onto the same board as the encoder module.

The encoder module 104 implements processing of the image stream using the techniques described herein, including, e.g., implementing encoders perform a conversion from images to codes, mimicking the operation of retinal circuitry The transformations specified by the encoders are applied to the series of input images, producing encoded output. For example, the encoded output may be in the form of values indicative of the firing rates of retinal cells that would have been generated had the images been received by a retina. The output can also be, for example, information indicative of the retinal cells "generator potential", i.e., the output of the linear component of the retinal model (the output of the convolution of the image with the linear filters). The encoded output may be indicative of the pulse train of "spikes" generated by the retinal cells.

In some embodiments, sets of different encoders may be used to better mimic the processing of a normal retina, since there are different types of retinal output cells. Differences may correspond to a particular cell type (e.g, ON cell or OFF cell) or to the cell position on the retina (e.g., ON cell in central retina versus periphery). When the encoder module 104 has more than one encoder, the encoders may operate in parallel, either independently or through at least one or more coupling mechanisms.

Fig. 2 is a flow chart illustrating the operation of an exemplary embodiment of the encoder module 104. In step 201, the encoder module 104 receives a series of images from the camera 102 (or some other suitable source). In optional step 202, these raw images undergo pre-processing, e.g., to rescale the contrast/intensity of the images, to apply a noise filter to the images, to crop the images, etc.

In step 203 the raw images are processed to determine information indicative of the retinal cell response to the images. For example, in one embodiment, for various positions in the image field, the encoders process the image stream and output a time dependent value corresponding to the firing rate that would be generated by a retinal cell (or group of cells) if the image stream were to impinge on a retina. In one embodiment, the firing rate output is formatted as follows: for a given time t, the output is a matrix of bits where the element at position (x,y) corresponds to the firing rate of the retinal cell at position (x,y).

Note that in some embodiments, the encoders may generate information indicative of the response of the retinal cell using a metric other than firing rate. For example, the output of the encoders could correspond to the activation state of the cell, the intracellular potential, the generator potential mentioned above, etc.

In step 204, the encoded information from step 203 is used to generate images (referred to herein as "retinal images" or when referring to time-varying images, the "retinal image stream" or the "retinal image data stream") suitable for processing by the machine vision module 106. For example, where the encoded information is output as a matrix of firing rates, as described above, a firing rate retinal image may be generated, where the intensity of each pixel in the "retinal image" is determined by the firing rate value of a corresponding element in the matrix (see Fig. 3 for an example). Any suitable relationship between firing rate and pixel intensity may be used, including a linear relationship, a non-linear relationship, a polynomial relationship, a logarithmic relationship, etc. The conversion between firing rate and pixel intensity may be implement using any suitable technique including the use of a look-up table. In some embodiments, the firing rate may be represented in the retinal image using an image characteristic other than intensity. For example, in embodiment where the retinal images are color images, a color space coordinate of each pixel could correspond to the firing rate.

In optional step 205 the retinal images undergo post-processing. Any suitable processing technique may be used, including, e.g., rescaling, filtering, cropping, smoothing, etc. In step 206, the retinal images are output to the machine vision module 106.

Note that in some embodiments, step 204 and step 205 may be omitted. In this case, the output of the encoder may be sent directly to a machine vision algorithm for processing. As will be apparent to one skilled in the art, in some cases this may require the modification of known machine vision algorithms to accept input data that is not formatted as traditional image data. However, in many embodiments, this can be accomplished in a straightforward fashion, without the need for modification of the core concepts of the particular algorithm.

In some embodiments, each encoder performs a preprocessing step, followed by a spatiotemporal transformation step. The preprocessing step is a rescaling step, which may be performed in a preprocessor module of the processing device, that maps the real world image, I, into quantities, X, that are in the operating range of the spatiotemporal transformation. Note that I and X are time-varying quantities, that is, $I(j,t)$ represents the intensity of the real image at each location j and time t, and $X(j,t)$ represents the corresponding output of the preprocessing step. The preprocessing step may map as follows: $I(j,t)$ is mapped to $X(j,t)$ by $X(j,t)=a+bI(j,t)$, where a and b are constants chosen to map the range of real world image intensities into the operating range of the spatiotemporal transformation.

The rescaling can also be done using a variable history to determine the quantities a and b, and a switch can be used to set the values of these quantities under different conditions (e.g., different lighting or different contrast).

For grayscale images, both $I(j,t)$ and $X(j,t)$ have one value for each location j and time t.

For color images, the same strategy is used, but it is applied separately to each color channel, red, green, and blue. In one embodiment, the intensity $I(j,t)$ has three values (I_1, I_2, I_3) for each location j and time t, where the three values I_1, I_2, I_3 represent the red, green, and blue intensities, respectively. Each intensity value is then rescaled into its corresponding X value (X_1, X_2, X_3) by the above transformation.

In one embodiment, the spatiotemporal transformation step is carried out using a linear-nonlinear cascade (reviewed in Chichilnisky EJ 2001; Simoncelli et al 2004), where the firing rate, λ_m, for each ganglion cell, m, is given by

$$\lambda_m(t;X) = N_m((X * L_m)(j,t) \qquad (1)$$

where * denotes spatiotemporal convolution, L_m is a linear filter corresponding to the mth cell's spatiotemporal kernel, and N_m is a function that describes the mth cell's nonlinearity, and, as in the previous section X is the output of the preprocessing step, j is the pixel location, and t is time. The firing rates, λ_m, may then be used to generate a firing rate retinal image as discussed above.

L_m is parameterized as a product of a spatial function and a temporal function. For example, in one embodiment, the spatial function consists of a weight at each pixel on a grid (e.g., the digitized image in a camera), but other alternatives, such as a sum of orthogonal basis functions on the grid, can be used. In one embodiment, the grid consists of a 10 by 10 array of pixels, subserving a total of 26 by 26 degrees of visual space (where each pixel is 2.6 by 2.6 degrees in visual space), but other alternatives can be used. For example, because the area of visual space that corresponds to a retinal ganglion cell varies with spatial position on the retina and from species to species, the total array size can vary (e.g., from at or around from 0.1 by 0.1 degree to 30 by 30 degrees, which corresponds to at or around 0.01 by 0.01 degree to 3 by 3 degrees in visual space for each pixel in a 10 by 10 array of pixels.) It is appreciated that the angle ranges and size of the pixel array are only provided for illustration of one particular embodiment and that other ranges of degrees or size of pixel arrays are encompassed by the present invention. For any chosen array size, the number of pixels in the array can also vary, depending on the shape of the area in visual space that the cell represents (e.g., an array of at or around from 1 by 1 to 25 by 25 pixels). Similarly, the temporal function consists of a sum of weights at several time bins and raised cosine functions in logarithmic time at other time bins (Nirenberg et al. 2010; Pillow JW et al. 2008). Other alternatives, such as a sum of orthogonal basis functions, can also be used.

In this embodiment, the time samples span 18 time bins, 67 ms each, for a total duration of 1.2 sec, but other alternatives can be used. For example, because different ganglion cells have different temporal properties, the duration spanned by the bins and the number of bins needed to represent the cell's dynamics can vary (e.g., a duration at or around from 0.5 to 2.0 sec and a number of bins at or around from 5 to 20). Temporal properties can also vary across species, but this variation will be encompassed by the above range.

Eq. 1 can also be modified to include terms that modify the encoder's output depending on its past history (i.e., the spike train already produced by cell m), and on the past history of the output of other ganglion cells (Nirenberg et al. 2010; Pillow JW et al. 2008).

In another embodiment, the linear filter L_m is parameterized as the sum of Q terms, where each of the terms is the product of a spatial function and a temporal function.

$$L_m = \sum_{k}^{Q} S_k \otimes T_k$$

where ⊗ denotes the outer product, and S_k and T_k are the *kth* spatial and temporal functions, respectively (k ranges from 1 to Q).

In this embodiment, individual spatial functions may be parameterized as described earlier, for example, as weights at each pixel on a grid, or as the sum of orthogonal basis functions on the grid. Individual temporal functions may also be parameterized as before, for example, as the sum of weights at several time bins and raised cosine functions in logarithmic time at other time bins. Other alternatives, such as a sum of orthogonal basis functions, can also be used.

In one embodiment, Q is 2, and L_m may be written as

$$L_m = S_1 \otimes T_1 + S_2 \otimes T_2$$

where ⊗ denotes the outer product, and S_1 and T_1 are the first pair of spatial and temporal functions, and S_2 and T_2 are the second pair of spatial and temporal functions.

For both sets of parameters for L (spatial and temporal), the choice of resolution (pixel size, bin size) and span (number of pixels, number of time bins) may determined by two factors: the need to obtain a reasonably close proxy for the retina's code, and the need to keep the number of parameters small enough so that they can be determined by a practical optimization procedure (e.g., as detailed in the Prosthesis Applications). For example, if the number of parameters is too small or the resolution is too low, then the proxy will not be sufficiently accurate. If the number of parameters is too large, then the optimization procedure will suffer from overfitting, and the resulting transformation (Eq. 1) will not generalize. The use of a suitable set of basis functions is a strategy to reduce the number of parameters and hence avoids overfitting, i.e., a "dimensionality reduction" strategy. For example, the temporal function (that covers 18 time bins, 67 ms each) may be parameterized by a sum of 10 weights and basis functions; see section "Example 1, Method of building the encoder" of the Prosthesis Application and (Nirenberg et al., 2010; Pillow JW et al. 2008)

The nonlinearities N_m are parameterized as cubic splines, but other parameterizations can be used, such as, piecewise linear functions, higher-order splines, Taylor series and quotients of Taylor series. In one embodiment, the nonlinearities N_m are parameterized as cubic spline functions with 7 knots. The number of knots is chosen so that the shape of the nonlinearity is accurately captured, while overfitting is avoided (see above discussion of overfitting). At least two knots are required to control the endpoints, and thus the number of knots can range from about 2 to at least about 12. Knots are spaced to cover the range of values given by the linear filter output of the models.

For the spatiotemporal transformation step, in addition to the linear-nonlinear (LN) cascade described above, alternative mappings are also within the scope of the present invention. Alternative mappings include, but are not limited to, artificial neural networks and other filter combinations, such as linear-nonlinear-linear (LNL) cascades. Additionally, the spatiotemporal transformation can incorporate feedback from the spike generator stage (see below) to provide history-dependence and include correlations among the neurons as in (Pillow JW et al. 2008; Nichols et al, 2010). For example, this can be implemented by convolving additional filter functions with the output of the spike generator and adding the results of these convolutions to the argument of the nonlinearity in Eq. 1.

Other models may also be used for the spatiotemporal transformation step. Non-limiting examples of the models include the model described in Pillow JW et al. 2008, dynamic gain controls, neural networks, models expressed as solutions of systems of integral, differential, and ordinary algebraic equations approximated in discrete time steps, whose form and coefficients are determined by experimental data, models expressed as the result of a sequence of steps consisting of linear projections (convolution of the input with a spatiotemporal kernel), and nonlinear distortions (transformations of the resulting scalar signal by a parameterized nonlinear function, whose form and coefficients are determined by experimental data, models in which the spatiotemporal kernel is a sum of a small number of terms, each of which is a product of a function of the spatial variables and a function of the spatial variables and a function of the temporal variables, determined by experimental data, models in which these spatial and/or temporal functions are expressed as a linear combination of a set of basic functions, with the size of the set of basis function smaller than the number of spatial or temporal samples, with the weights determined by experimental data, models in which the nonlinear functions are composed of one or segments, each of which is a polynomial, whose cut points and/or coefficients are determined by experimental data, and models that combine the outputs of the above models, possibly recursively, via computational steps such as addition, subtraction, multiplication, division, roots, powers, and transcendental functions (e.g., exponentiation, sines, and cosines).

As described in the Prosthesis Applications, encoders of the type descried above can very closely mimic the input/output function of real retinal cells. As detailed therein, in some cases this may be characterized by determining a standard Pearson correlation coefficient between a reconstructed retinal image's values at each pixel, and that of the corresponding raw image. Thus, a correlation coefficient of 1 indicates that all of the original image's information was perfectly retained, while a correlation coefficient of 0 indicates that the resemblance of the reconstruction to the real image was no greater than chance.

For example, in some embodiments, the encoder is configured such that the Pearson's correlation coefficient between a test input stimulus and a corresponding stimulus reconstructed from the encoded data that would be generated by the encoder in response to the test input stimulus is at least about 0.35, 0.65, at least about 0.95, or more, e.g., in the range of 0.35-1.0 or any subrange thereof. In some embodiment, the test input stimulus includes a series of natural scenes (e.g. spatiotemporally changing scenes).

In some embodiments, the retinal encoders of the type described herein mimic the input/output function of real retinal cells for a wide range of inputs, e.g., spatio-temporally varying natural scenes. In typical embodiments, this performance is substantially better that conventional encoders.

Figs. 18A-F illustrates the performance of retinal encoder models for various cells (cells 1-6, respectively) when tested with movies of natural scenes, including landscapes, people walking, etc. In each figure, the performance of a conventional linear-nonlinear (LN) model is shown on the left, and the performance of the linear-nonlinear (LN) model of the type described in this application is shown on the right. Performance is shown via raster plots and peri-stimulus time histograms (PSTHs). The conventional (LN) model was developed based only on the experimental response of retinal cells to a white noise stimulus. In contrast, the linear-nonlinear (LN) models of the type described in this application are developed based on recorded cell responses to both white noise and natural scene stimuli.

For the examples shown, the input test stimulus for both types of models is a movie of natural scenes, taken in Central Park in New York City. As shown, the standard LN model is not highly effective on natural scene stimuli: that is, this model, which is built using white noise stimuli, does not produce spike patterns that closely match those of the real cell. In contrast, the LN model described in this application, which is built using white noise and natural scene stimuli, is highly effective. The spike patterns it produces closely match those of the real cell. (Note that the natural scene movie used to test the models is different from that used to train the models, as is required for validating any model. Note also that in each figure, the same real cell is used as the basis for both types of models. Finally, note that performance of the encoder models of the type described herein has been demonstrated with a host of other stimuli, including movies, of faces, people walking, children playing, landscapes, trees, small animals, etc., as shown in the Prosthetic Application, and in Nirenberg, et al. Retinal prosthetic strategy with the capacity to restore normal vision, PNAS 2012 and the accompanying Supplementary Information section available at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1207035109/-/DCSupplemental).

The same conclusions about performance can be drawn from the PSTHs. The light gray trace shows the average firing rate of the real cell; the dark grey trace shows the average firing rate of the model cell. The standard LN model misses many features of the firing rate; each of the different Figs. 18A-18F, show examples of the different features missed by the standard model. The model described in this application, though, captures the features of the firing rates reliably and does so for an array of different cells (many other examples are shown in the Prosthetic Application).

Fig. 3A illustrates the conversion of a raw image into a retinal image. Panel A shows several frames of the raw image stream acquired by camera 102. As shown, the raw image stream includes a person walking through a complex environment. Panel B shows the corresponding retinal image frames, where the retinal image pixel intensities correspond to firing rates generated by the encoders of the encoder module 104. Four different retinal image streams are shown, each using a different array of cells (OFF midget cells, ON midget cells, OFF parasol cells, and ON parasol cells, as indicated on figure). Note that, the retinal image frames shown are produced by the encoder module 104 after a brief time delay, corresponding processing delay time in a natural retina (as show, approximately 80 ms).

Note that it is apparent that the total amount of information contained in the retinal images is less than that of the raw images. This reduction in information can advantageously reduce the processing load on the machine vision. Moreover, because the encoders mimic the behavior of the retina, for some machine vision applications, the information retained in the retinal images will include the salient features required for the machine vision task at hand, allowing for efficient and effective operation of the machine vision module 106.

Figs. 3B-3F show enlarged views of the raw image (Fig. 3B) and retinal images (Figs. 3C-3F) corresponding to the last column of Fig. 3A. In the raw image, a human figure is moving from right to left within a relatively static, but complex, environment. Note that in all of retinal images (Figs. 3C-3F), the static environment has been de-emphasized to vary degrees, while the moving human form has been emphasized. Moreover, in both images, a "motion shadow" type effect is apparent trailing the human figure that provides an indication of the direction of motion. Accordingly, although the overall amount of information contained in the image has been reduced, that which remains emphasizes features important features, i.e., the moving human form.

Note that none of these effects are the result of any intentionally designed programming. That is, the encoder was not intentionally programmed to identify moving features. Instead the emphasis of these features is a result of the fact that the encoder mimics the natural processing that occurs in the retina. Although certain kinds of emphasized features are apparent in the present example (a human form moving against a static background), it is to be understood that for other types of input images the retina may emphasize other types of features. The key concept is that, in general, the features emphasized for any given set of images will be those determined to be salient based on millions of years of evolution of the retina. Accordingly, as described in detail below, the retinal images will be particularly advantageous when used in machine vision applications where it is known that biological vision systems perform well (e.g., certain types of pattern recognition tasks such as facial recognition, identification of human or other living forms against a complicated background, navigation through a complicated environment, rapid tracking of and reaction to moving objects, etc.).

In some embodiments, the encoders encode the image data on about the same time scale as the encoding carried out by the normal or near-normal retina. In various embodiments, the encoder operates with an acceptable processing lag time. As used herein, processing lag time refers to the amount of time between the occurrence of an event in the visual stimuli received by the camera 102, and the delivery of corresponding output code (e.g., the corresponding retinal images) to the machine vision module 106. In some embodiments, encoding module has a lag time of less than about 50 ms, less than about 20 ms, less than about 10 ms, less than about 5 ms, etc., e.g., in the range of 5-50 ms or any subrange thereof.

Referring back to Fig. 1, the machine vision module 106 receives the retinal images from the encoder module 104 and process the image using any suitable machine vision technique. Although a number of such techniques are mentioned herein, it is to be understood that these examples are not limiting, and other techniques may be used. For example, in various embodiments, one or more of the techniques described in D. A. Forsyth, J. Ponce *Computer Vision: A Modern Approach*, Second edition Prentice Hall, 2011 and/or D.H. Ballard, C.M. Brown; *Computer Vision*, Prentice-Hall Inc New Jersey, 1982 (available online at http://homepages.inf.ed.ac.uk/rbf/BOOKS/BANDB/bandb.htm), R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer 2010, available online at http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf); and E.R. Davies, *Computer and Machine Vision, Fourth Edition: Theory, Algorithms, Practicalities*, Elsevier 2012, may be used.

In various embodiments, the machine vision module 106 may implement one or more available computer vision algorithms or software tools, e.g., any of those included in the OpenCV software package, available at http://opencv.willowgarage.com/wiki/ or the Gandalf computer vision software package, available at http://gandalf-library.sourceforge.net/.

The machine vision module 106 may use the retinal images to perform any suitable task including recognition tasks (e.g., object recognition, image classification, facial recognition, optical character recognition, content-based image retrieval, pose estimation, etc.), motion analysis tasks (e.g., egomotion determination, movement tracking, optical flow determination, etc.), modeling tasks (e.g., scene reconstruction, 3D volume recognition, etc.).

In some embodiments, the machine vision module 106 may divide the visual field into domains, which may be equally or unequally sized. The domains may or may not overlap. The domains may cover a band of the visual field (for instance the entire field of view on a horizontal axis and a limited span on a vertical axis) or may cover the entire field of view.

In some embodiments, the machine vision module 106 may apply boundary edge detection techniques to the retinal images, including, e.g., first order edge detection techniques such as Canny edge detection, second order edge detection techniques, or phase congruency based edge detection techniques. Edge detection may involve the application of one or more transformations to the retinal images, e.g., the Hough transformation.

In some embodiments, the machine vision module 106 may calculate an optical flow based on the stream of retinal images. An optical flow may be indicative of a pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer (an eye or a camera) and the scene. The optical flow may be used for any number of applications including motion detection, object segmentation, time-to-collision and focus of expansion calculations, etc. Method for calculating optical flow may include, phase correlation methods, block-based methods, differential methods (such as the Lucas–Kanade, Horn–Schunck, Buxton–Buxton, and Black–Jepson methods), variational methods, discrete optimization methods, etc.

In some embodiments, the machine vision module 106 may apply one or more image segmentation techniques to segment the retinal images (e.g., to identify areas of interest). Exemplary segmentation techniques include thresholding, clustering methods, compression-based methods, histogram-based methods, edge detection (e.g., using the edge detection techniques described above), region growing methods split-and-merge methods, partial differential equation based methods (e.g., level set methods), graph partitioning methods, watershed transformation based methods, model based segmentation methods, multi-scale segmentation, semi-automatic segmentation, neural network based segmentation, etc.

In various embodiments, the machine vision module 106 may be trained using any computer learning technique known in the art. Computer learning techniques include supervised learning (e.g., including statistical classification techniques), unsupervised learning, reinforcement learning, etc. In some embodiments, machine vision module 106 may include one or more artificial neural networks which may be trained to perform various tasks.

Fig. 4 illustrates an exemplary training system 400 for training the machine vision module 106 of the machine vision system 100. The training system includes a source 402 of raw training images (e.g., a database of stored images), and encoder module 404 that generates retinal images based on the raw training images using the techniques described herein, the machine vision module 108 that receives the retinal images from the encoder, and a controller 406 that monitors and modifies the operation of the machine vision module based on the monitored performance.

Fig. 5 is a flowchart illustrating the operation of the training system 400. In step 501, the encoder 404 receives the training images from the source 402. For example, the training images may be a series of medical images of tumors, where a first portion of the images are know to correspond to malignant tumors, while a second portion of the training images correspond to benign tumors.

In step 502, the encoder converts the raw training images into retinal images. In step 503, the retinal images are output to the machine vision module 106.

In step 504, the controller 406 monitors the performance of the machine vision module 106 as it processes the retinal images to perform a task. In the case of the medical images, the machine vision module 106 may apply an image recognition technique differentiate the images of malignant tumors from images of benign tumors. The controller monitors the performance of the machine vision module 106 as it performs this task (e.g., calculating the error rate in discriminating malignant tumors). If the performance is acceptable, the process ends in step 505. If the performance is unacceptable (e.g., if the error rate is above a threshold level), in step 506 the controller 406 adjusts the machine vision module 106 (e.g., by modifying one or more parameter, by changing the connections in an artificial neural network, etc.), and the process returns to step 503. Accordingly, the controller 406 iteratively adjusts the machine vision module 106 until its performance reaches an acceptable level (e.g., the error rate is below the threshold level).

Note that in various embodiments, other suitable types of training may be used. For example, in addition or alternative to comparing the performance to a fixed threshold, the training may instead implement a convergence criteria (e.g., where iterative training continues until the incremental increase in performance per iteration falls below a threshold level).

In various embodiments, the machine vision module 106 may include any suitable control techniques, including the use of complicated artificial intelligence based systems. However, for a number of applications, machine vision module 106 may implement a relatively simple control scheme. In some such embodiments, the machine vision 106 controls the some or all of the operation of one or more systems (e.g., the movement trajectory of a robot) based on a relatively simple moment to moment classification of the retinal images received from the encoder module. That is, the control does not depend on complicated planning, but only on temporally localized classifications. Advantageously, learning algorithms know in the art are known to be amenable to the performance of these types of relatively simple classification tasks.

For example, referring to Fig. 6, in one embodiment, the machine vision system 100 is used to control a robot 600 to navigate through an environment featuring obstacles, e.g., a maze, as shown. The camera 102 of the machine vision system is mounted on the robot 600, and has a field of view that captures the scene in front of the robot.

A video stream from the camera 102 is processed by the encoder module 104 to generate a stream of retinal images. In one case, the encoder module may mimic the performance of mouse retinal ganglion cells (e.g., using a encoder characterized by the encoder parameters set forth in the subsection the Prosthesis Applications entitled "Example set of encoder parameters for a mouse ganglion cell"). In another case, the encoder module may mimic the performance of monkey retinal ganglion cells (e.g., using a encoder characterized by the encoder parameters set forth in the subsection of the Prosthesis Applications entitled "Example set of encoder parameters for a monkey ganglion cell").

The stream of retinal images is processed, e.g., using optical flow techniques, to determine the speed of motion at various locations in the images. In general, locations or domains in the image with slower speeds will correspond to objects that are distant from the robot 600, while locations with faster speed will correspond to objects that are close to the robot. To avoid running into obstacles, the machine vision module 106 controls the robot to move in a direction corresponding to the slower moving locations in the image.

For example, in one embodiment (shown in Fig. 7), the visual field (i.e., the retinal image data stream) is divided into N=7 equally-sized regions by an image segmentation step, 702. In this embodiment, the regions do not overlap, and they divide up the camera's horizontal field of view (which is 40 °) from left to right, so that each region spans 5.7 ° horizontally; in the vertical direction, they are limited to the bottom half of the navigator's field of view (which is 27 °), so that these regions span 13.5 ° vertically.) At regular intervals (e.g., every 2 seconds), two consecutive retinal images from the retinal image sequence are taken and sent to the machine vision module 106 for classification. Since each retinal image has been divided into N regions, the machine vision module receives N pairs of regions. Each pair is passed through a convolutional neural network (CNN) 704, which classifies the optical flow speed in that region. The output of this classification may be a speed label L_i for each region i, where L_i is a number between 1 and M, 1 representing a very slow average speed in the region, and M representing a very fast average speed. For example, M can be 8, so that there are 8 different speed classes.

The result is an array of N classifications 706; based on these, a turn decision is made by a turn decision module 708. The "target region" (the region to head towards) is chosen to be the region with the slowest speed classification, that is, the smallest number L_i. If there are multiple regions that are tied for having the slowest speed classification, the turn decision module 708 may select the region that is closest to center (so as to minimize the amount of turning) or some other region based on the desired use of the system. Once a target region is chosen, the machine vision module 106 (specifically, the turn decision module 708 in machine vision module 106) initiates a turn so that the navigator comes to face the center of the target region.

The example above refers to navigation of a robot. It is to be understood that in various embodiments, the techniques above may be used for other types of navigation, including navigation through a virtual world, as described in the example below.

For example, the machine vision module 106 may identify and avoid obstacles by dividing the image field of the retinal image stream into several regions or domains, and classifying the regions, into speed categories, and controlling the robot 600 to move in the direction corresponding to the image region in the lowest speed category. The machine vision module 106 may be trained to perform this classification task using a relatively simple training algorithm, such as the CNN described above and in the example below or a boosting algorithm (e.g., the *AdaBoost* algorithm, see Yoav Freund, Robert E. Schapire. "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting", 1995).

In general, the devices and techniques may be used for any suitable application including, medical image processing (e.g., automated or computer aided medical diagnosis), robotic control or navigation, industrial process monitoring and control, automated sorting applications, motion tracking based interfaces (e.g., as used with computer gaming systems), etc. The devices and techniques described herein may operate in real time or near real time, e.g., allowing for practical automation of the applications mentioned above.

Example – Virtual World Navigation

In one example assessing the effectiveness of one approach to machine vision, a navigation task was used, as this is particularly challenging (requiring processing in both space and time). This approach applied aspects of several learning algorithms commonly used for navigation, e.g., as described in LeCun, Y. et al. (2010) Convolutional Networks and Applications in Vision. Proc. International Symposium on Circuits and Systems (ISCAS'10), pp. 253-256. IEEE; Szarvas, M. et al. (2005) Pedestrian detection with convolutional neural networks. Proc. Intelligent Vehicles Symposium, pp. 224- 229. IEEE; Jackel, L. D. et al. (2006) The DARPA LAGR program: Goals, challenges, methodology, and phase I results. Journal of Field Robotics, 23, 945–973, each incorporated herein in its entirety by reference. Using these techniques a navigator was constructed that learns its environment using a Convolutional Neural Network (CNN) - a learning algorithm. The CNN was constructed using an open-source numerical processing and automatic differentiation package called Theano (available to the public at http://deeplearning.net/software/theano/).

The navigator was designed to learn the speed of things in its training environment. The navigator was given a training environment, and was used it to divide the training environment at each moment in time into n domains. The navigator then learns the speeds in the domains. The speeds provide useful information for navigating. If something is moving very quickly, it means it's very close to the virtual object navigating the environment (it's moving rapidly across your retina). If it is close, the virtual object is likely going to hit it. So the navigator assesses the domains in the environment and then moves toward the domain with the slowest speed (the one with the slowest speed is the furthest away and the safest). In this example, the navigator is not directed to head to a particular end point, but to move forward and not collide with anything.

More specifically in this example, using the method show in Fig. 7, as the navigator traverses an environment, its visual field is divided into 7 equally-sized regions by an image segmentation step, 702. In this embodiment, the regions do not overlap, and they divide up the camera's horizontal field of view (which is 40 °) from left to right, so that each region spans 5.7 ° horizontally; in the vertical direction, they are limited to the bottom half of the navigator's field of view (which is 27 °), so that these regions span 13.5 ° vertically.).

At each decision time point, an algorithm based on convolutional neural networks (CNNs) classifies the optical flow speeds in each of the domains (step 704). The output of this classification is a speed label L_i for each domain i (step 706), where L_i is a number between 1 and 8, 1 representing a very slow average speed in the domain, and 8 representing a very fast average speed.

As described earlier, based on these classifications, one for each of the 7 domains, a navigation decision is made by the turn decision module (708). The "target domain" (the domain to head towards) is chosen to be the domain with the slowest speed classification. If there are multiple domains that are tied for having the slowest speed classification, the navigator selects the one that is closest to center (so as to minimize the amount of turning); if there is still a tie, the navigator breaks it by choosing the domain to the left. Once a target region is chosen, the machine vision module (106) initiates a turn so that the navigator comes to face the center of the chosen region.

Virtual environments were created for training and testing using an open-source 3D rendering framework called Panda3D (available to the public at http://www.panda3d.org/). Streams of frames from the training set are shown in Fig. 8; streams of frames from the three testing sets are shown in Fig. 9A, B, C. As shown, the training set was a rural environment. The three testing sets were as follows: a rural environment that is different from the one used in the training set, a suburban environment, and a playground.

The performance of the navigator was compared under two conditions: 1) when it was trained the standard way, i.e., using the raw image stream as the input, and 2) when it was trained using the "retinal image stream" as the input - that is, when it used images that were processed through our encoder. In this case, the encoder used was generated using monkey midget and parasol cells as per the methods described in Nirenberg, S. and Pandarinath, C. (2012) A retinal prosthetic with the capacity to restore normal vision. Proc. Natl. Acad., in press; and Nirenberg, S. et al. (2011) Retina prosthesis and the Prosthesis Applications; each incorporated herein in its entirety by reference.

As shown in Fig 10A, when the navigator learned its environment from the raw image stream, its performance is low, many collisions occur; what is learned with the training set does not generalize to the new environments. As shown in Fig 10B, when the navigator learned the environment from the *retinal* image stream, performance was dramatically better: note the straight paths and the lack of collisions. There is clear generalization to new environments (rural, suburban, playground) - issues that have been highly problematic for artificial navigation systems, and machine learning algorithms in general.

Fig. 11 shows further demonstration of the navigator's high performance when it uses the retinal image streams as input. Specifically, it shows that the high performance generalizes not just to different environments (from rural to suburban to playground), but it also generalizes to different lighting conditions *within* an environment. A through F correspond to different positions of the sun, and therefore, different shadow conditions, in the playground environment; the light conditions span sunrise to sunset, i.e., 30 degrees above the horizontal on the left side of the environment to 30 degrees above the horizontal on the right side. As shown in the figure, when the navigator was trained on raw image streams (from the rural environment using one lighting condition), its performance does not generalize: its performance in the playground is low and this is true across light conditions. The height of each bar in the figure corresponds to the fraction of trials in which the navigator successfully stayed within the playground tire course without colliding with one of the tires. The error bars indicate the standard error of the mean (SEM). In contrast, when the navigator when it was trained on the *retinal* image streams (same rural environment using same single lighting condition, but this time processed through the encoder), its performance is high, and the high performance holds across light conditions. Thus, training on the retinal image streams (i.e., training on the images processed through the encoder) leads to high performance that generalizes both to new environments and to multiple lighting conditions (sunrise to sunset, see above).

Note that the encoders operate in real time, indicating that the processing techniques can be readily applied to non-virtual environments as well, e.g., to control the motion of a robot in a real world environment.

Example – Face Recognition

This example assesses the effectiveness of the approach described in this application to another long-standing problem in machine vision, the recognition of faces in video. Using a learning algorithm commonly used for face recognition and pedestrian detection [see Viola and Jones 2001; Viola, Jones, and Snow 2005], a system was constructed to recognize an individual's face in video, i.e., one that can classify a previously unseen image stream as a "target face" versus another or "non-target" face. The same approach can be used for many other purposes, such as, but not limited to, pedestrian detection, object recognition, object tracking, whole-person recognition, iris detection, etc. The system was implemented using the Python programming language and the NumPy numerical computing package.

An embodiment of the approach is described in Fig. 12. An input video (raw image stream) is passed through the retinal encoder 104, producing the retinal image stream. Since the task focuses on faces, the retinal image stream is then cropped to locate a face-containing region 1202. (The cropping is done after the encoder processes the raw stream, so as to avoid edge effects when the encoding is carried out.) In this example, face-containing regions were selected manually, so as to construct a training and testing set of known face examples. In other embodiments, face-containing regions could be detected in the raw image stream or in the processed image stream using the Viola-Jones algorithm [Viola and Jones, 2001]. The cropped video is then fed through a classifier 1206 (e.g., one based on a boosted cascade of Haar filters, such as in Viola Jones and Snow, 2005). The classifier 1206 designates it either as a "target face" (meaning that it is the face of the target individual) or "non-target face" (meaning that it is the face of a different individual).

Fig. 15 shows an example of the effectiveness of our approach. For this analysis a data set of faces in video was used from http://www.cs.tau.ac.il/~wolf/ytfaces/. The reference is Lior Wolf, Tal Hassner and Itay Maoz. Face Recognition in Unconstrained Videos with Matched Background Similarity. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR),* 2011.

Using this data set, several face recognition tasks were performed. The general procedure was to train the face recognition algorithm on a "target face". The algoritm was presented with an array of videos showing a person's face, the target face. The algorithm's ability to recognize the face was tested by presenting it with previously unseen videos of the same person's face along with videos of other faces, "non-target faces". The job of the algorithm was to correctly classify the test videos as either target faces or a non-target faces.

Figs. 13 and 14 show images from example videos. Fig. 13 shows frames from a video that was used to train the face recognition algorithm, and Fig. 14 shows frames from a video that was used to test the algorithm. As shown, the person in the test video (Fig, 14) is the same as that in the training video (Fig. 13), but shown in a different environment with a different hairstyle, etc.

The performance of the algorithm was tested under two conditions: when we trained it in the standard way, i.e., using the raw image streams of the faces, and when we trained it using the retinal image streams of the faces (that is, the raw image streams after they were processed by our encoder). In both cases, the training was performed using short (two-frame) movies. The number of two-frame movies used in the training ranged from 250-800 for the target face (taken from 4-5 different videos), and 2000 for the non-target faces (taken from >100 videos). Performance was then measuring using 50-800 two-frame movies taken from previously unseen video, that is, videos not used for the training.

As shown in Fig. 15, the use of the encoder had a clear impact on performance. The results for two kinds of tasks are shown: the first consists of very challenging tasks, defined as ones where the standard approach performs very weakly; the second consists of easier tasks, where the standard approach performs moderately well. As shown, when the task was difficult (Fig. 15A), the approach that incorporates the encoder provides a major (4-fold) improvement, over the standard approach. When the task was less challenging, i.e., when the standard approach performs moderately well, the approach that incorporates the encoder still provides substantial improvement (a factor of 1.5 over the standard method).

In an alternate embodiment, the task is slightly modified, so that the face detection step is bypassed, and instead, cropped videos of the appropriate size for the classifier 1206 are generated in an automated fashion from the input video, whether or not faces are present in a particular part of the video. Then, classification is applied to these new cropped videos as before, or a modified classification is performed, where the output classes are "target face" and "non-target face," or "non-face."

In an alterative embodiment, the analysis could be performed using N frames, where N could be 1, 3 or more frames, as many as the processor can handle, as opposed to the 2-frame videos used for the analysis in Fig. 15.

In addition, these classifications may be used by themselves, for instance to alert a user to the presence of the individual in the video, or they may be combined in some way, for instance by waiting for several positive detections ("target face" classifications) to occur within a specified time window before issuing a signal.

Note that, although a number of exemplary applications of retinal processing to machine vision have been described, embodiments directed to numerous other applications may be used.

In general, the encoder approach is likely to be advantageous for visual tasks that animals (vertebrates) perform well, especially those where animal visual systems are known to perform better than existing machine techniques. As noted above, the encoder approach may be particularly effective in cases where it would be advantageous to reduce the total amount of information from the raw image stream (e.g., to allow or faster processing), while maintaining salient features in the data. For example, as noted above, in some embodiments, the encoder approach will typically be particularly advantageous when used in, e.g., certain types of pattern recognition tasks such as facial recognition, identification of human or other living forms against a complicated background, navigation through a complicated environment, rapid tracking of and reaction to moving objects, etc.

Note that for certain applications where biological systems do not typically perform well, the encoder approach may have limitations. This may particularly be the case in applications that require a high level of detailed information or precision measurement. For example, referring back to retinal images shown Figs. 3B-F, note that while these images advantageously emphasize the presence and motion of the human figure, the retinal images do not provide a sharp outline of the human figure that would be useful, e.g., in determining precise biometric information such as the human's absolute height or other absolute bodily dimensions. To determine this type of information, it may be better to apply machine vision algorithms to the raw image.

In some embodiments, a hybrid approach may be used to provide the advantages of both the encoder based approach to machine vision and a traditional approach applied to the raw image data.

For example, in some embodiments, a raw image stream may be processed using any of the retinal encoder based techniques described herein. The resulting retinal image data may be processed (e.g., using a machine vision algorithm, such as machine vision algorithm trained using retina images), and the results used to inform subsequent analysis of the corresponding raw images (e.g., using a machine vision algorithm, such as machine vision algorithm trained using raw images).

Fig. 16 illustrates an exemplary process of this type. In steps 1701 and 1702, raw images are obtained and used to generate a stream of retinal images, using any of the techniques described herein. In step 1703, the retinal images are analyzed, e.g., using a machine vision algorithm.

In step 1704, the results of the analysis of the retinal images are used to identify retinal images (or segments thereof) that are of interest. For example, in a person-recognition task, the encoder approach, which performs dimension reduction on the image in the way that the normal retina does to generate retinal images, can allow rapid identification of body types - by gait, signature gestures, etc. One of its strengths is that it rapidly pulls out motion information, which is particularly useful for this purpose. The encoder approach can thus serve as a prescreening approach to reduce the space of possible matches to the target individual (by excluding candidates with the wrong body type, gait, gestures, etc.)

In step 1705, the raw images (or segments thereof) that correspond to the identified retinal images may are analyzed. For example, in the case of a person recognition-task, an algorithm that uses the raw image (where little or no dimension reduction is used) may be applied to a subset of images to more positively identify the person using more detailed feature analysis (e.g., by extracting detailed biometric information such as an accurate height or other bodily dimensions of the person).

In various embodiments, the method described above may be reversed, with prescreening done on raw images, followed by subsequent analysis using a retinal encoder approach. In some embodiments, an iterative technique may be applied, with multiple rounds of alternative raw and encoder based analysis. In other embodiments, the different types of processing may occur in parallel, and the results synthesized. In general any suitable combination of traditional and encoder based approaches may be used.

As noted above, in various embodiments, the retinal processing operates to reduce the total amount of information from the raw image data (to achieve efficiency, in a way analogous to the way the retina does) while retaining salient features for a given application. For example, in some embodiments, even though the total amount of information in the retinal encoded data is reduced, the machine vision algorithm may exhibit better performance when applied to the encoded data than when applied to corresponding raw image data. This result was seen in both of the examples provided above, where navigation and facial recognition algorithms applied to "compressed" retinal images substantially outperformed the same algorithm applied to raw images.

In various embodiments, the retinal encoded data may be compressed by a factor of at least 1.5, at least 2, at least 3, at least 4, at least 5, or more, e.g., in the range of 1-100 or any subrange thereof. In some embodiments, this compression corresponds to a dimension reduction produced by the encoders. For example, in some embodiments, the bit rates of the retinal encoders may be quantified and can be compared to the entropy of the raw image data used as stimulus by the encoder (also measured in bits per unit time), and the ratio taken to determine a compression ratio. For example, in some cases described in the Prosthesis applications an encoder is described with a bit rate of 2.13 bits/s compared to an input raw data bit rate of 4.9 bits/s. Thus, the data compression produced by the encoders was in this example nearly 7-fold.

In some embodiments, the processing techniques described herein may be applied in an information storage and retrieval context. Referring to Fig. 17, a system 1800 includes a memory storage device 1801 (e.g., a hard drive or other compute memory) operatively coupled to a processor 1802. The storage device 1801 stores retinal image data that has been generated from raw image data using the techniques described herein. As detailed above, in some embodiments, the retinal image data may be compressed relative to the raw data, while maintaining certain salient features. Accordingly, the stored retinal data may, in some embodiments, be used as a representation, or "fingerprint" of corresponding raw data. In some embodiments, storage device stores database information indicative of a correspondence between the encoded data and the raw image data. For example, a particular video clip could be used to generate a corresponding retinal image stream, and the retinal image stream stored on the device 1801 with a tag identifying it with the raw video clip.

In some embodiments, the processor 1802 can be used to match incoming data with data stored on the storage device 1801. In some embodiments, the processor 1802 may receive query image data (e.g., a raw video clip) corresponding to a series of query images. The processor 1802 may then process the query image data with a retinal encoder to generate retinal encoded query data. The processor can then compare the retinal encoded query data with retinal encoded data stored on the storage device 1801. If a match is found, the processor can then read the tag on the stored data, and output information associating the query data video clip with the video clip used to generate the matching stored retinal image. In some embodiments, because the retinal encoded data is compressed and/or has had salient features enhanced, the matching of the encoded stored and query data may be faster and/or more accurate than trying to directly match the corresponding raw image clips.

The examples shown in this application and the Prosthetic Application used encoders built from data obtained from the mouse and monkey retina. However, it is to be understood that various embodiments may use encoders built from other species as well, such as, but not limited to birds, cats, snakes, and rabbits, which can be constructed using the procedure described in complete detail in the Prosthetic Applications.

In various embodiments, the overall function of the techniques described here is to utilize the preprocessing (particularly the dimension reduction) performed by the visual system (particularly the retina) to advance machine vision. For some applications, the preprocessing performed by retinas of other species may apply; e.g., encoders constructed from bird retinas may be particularly effective for flying navigators; similarly, encoders constructed from fast moving animals, such as tigers, may be particularly effective for navigators that need to operate at high speeds. In some embodiments, encoders based on multiple species may be used, and the results combined to provide advantageous synergies (e.g., using bird based encoders for basic flight navigation tasks, while using monkey based encoders for object recognition tasks when an object of interest is encountered during the flight).

Similarly, the approach generalizes to encoders built from higher visual areas, such as the lateral geniculate nucleus, superior colliculus, or visual cortex. The Prosthetic Applications describe the construction of encoders for retinal cells; the same method, again described in complete detail, including the mathematical formalism, can be also used to obtain encoders for higher visual areas, which can similarly serve as a preprocessing step for machine vision algorithms.

The invention techniques described herein can be used as front end processing (or filtering) for essentially any machine vision algorithm, as it works in an analogous way to the retina. Just as the retina preprocesses visual information for use by the brain – to allow it to perform a host of visually-guided activities, such as navigation, object and face recognition, figure-ground discrimination, predator detection, food versus non-food detection, among many others - the encoder(s), which together form a "virtual retina", can preprocess visual information for a host of machine algorithms.

What the retina does essentially is take the staggering amount of information in the visual world and reduces it to the essentials, the essentials needed by the brain for the survival of living beings. Because the encoders very accurately mimic the input/output relations of the retina (and do this for essentially any visual input, as shown in the prosthetic application), this means that the encoders reduce the information in the visual world in the same way. Thus, in various embodiments, the techniques described herein may provide front end processing for machine vision algorithms that is the same, or close to the same, as what the retina offers the brain, that is, it has the same speed, efficiency, and qualitative and quantitative filtering.

A corollary of this is that the encoders also impact the way machine vision algorithms are, or can be, constructed. Current algorithms are constructed to use raw images as their input, or images preprocessed in other ways (e.g. using difference of Gaussians filters). When images are processed through retinal encoders as described herein, the result is a new type of input for machine vision algorithms, i.e., input that has never previously been available. In some embodiments, this new input may allow for particular classes of algorithms to be adapted or optimized in a new way. For example, various machine vision algorithms are classified by a set of parameters which may be determined at least partially by on a training set of images, and/or images processed by the algorithm while completing a given task. When retinal image data are used in place of raw images, the resulting parameters of the algorithm will differ from those that would have been obtained using corresponding raw image data. In some cases, this will cause the algorithm to exhibit improved performance for a given task.

In some cases, because the machine vision algorithm is being trained using images that mimic the visual system of a vertebrate, the algorithm may advantageously adapt to acquire some of the performance qualities of the system. For example, because the retinal processing highlights the salience of certain aspects of images, a machine vision algorithm trained on retinal encoded data may "learn" to become more sensitive to these image aspects.

The examples above show two instances of machine vision algorithms – a navigator and a face recognizer – and in both cases, the algorithms changed their structure when applied to retinal processed input. Both algorithms were learning algorithms characterized by a set of weight parameters, and it was found that these parameters were different when the algorithm was applied to retinal image data versus when the images were applied to raw image data. The improved performance of the algorithms in the retinal processed case (relative to the raw image case) was due largely or completely to the change in the weight parameters. Note that this improved performance generalized to navigation and recognition tasks in environments and conditions that differed from the environment and conditions used in the training. This is evidence that, in some embodiments, the structure of a machine vision algorithm trained using retinal image data may fundamentally changes in a way that is beneficial and generalizes beyond the training environment and conditions. Similarly, new algorithm constructions may be developed to utilize this new input data; that is, not just new weights or parameters on current algorithms but new algorithms that more directly match or utilize the new input data described here.

The present methods and devices may process any type of image data. For example, the image data may be generated in response to visible light, but may also be generated by other types of electromagnetic radiation such as infrared, ultraviolet or other wavelengths across the electromagnetic spectrum. In some embodiments, the image data may be artificial or virtual image data (e.g., generated based on a model of a virtual environment). In some embodiments, the artificial image data may be related to the visualization of any kind of suitable data, including for example, medical imaging data (magnetic resonance imaging data, computer aided tomography data, seismic imaging data, etc.).

The image data may be a single image or a plurality of images; additionally, the images may be static or may vary in a spatiotemporal fashion. Simple shapes such as diagrams or comparatively complex stimuli such as natural scenes may be used. Additionally, the images may be grayscale or in color or combinations of grey and color.

In one embodiment, the stimuli may comprise white noise ("WN") and/or natural stimuli ("NS") such as a movie of natural scenes or combinations of both.

The scope of the present invention is not limited by what has been specifically shown and described hereinabove. Those skilled in the art will recognize that there are suitable alternatives to the depicted examples of materials, configurations, constructions and dimensions. Numerous references, including patents and various publications, are cited and discussed in the description of this invention and attached reference list. The citation and discussion of such references is provided merely to clarify the description of the present invention and is not an admission that any reference is prior art to the invention described herein. All references cited and discussed in this specification are incorporated herein by reference in their entirety.

While various inventive embodiments have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the function and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the inventive embodiments described herein. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the inventive teachings is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific inventive embodiments described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, inventive embodiments may be practiced otherwise than as specifically described and claimed. Inventive embodiments of the present disclosure are directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the inventive scope of the present disclosure. The above-described embodiments can be implemented in any of numerous ways. For example, the embodiments may be implemented using hardware, software or a combination thereof. When implemented in software, the software code can be executed on any suitable processor or collection of processors, whether provided in a single computer or distributed among multiple computers.

Further, it should be appreciated that a computer may be embodied in any of a number of forms, such as a rack-mounted computer, a desktop computer, a laptop computer, or a tablet computer. Additionally, a computer may be embedded in a device not generally regarded as a computer but with suitable processing capabilities, including a Personal Digital Assistant (PDA), a smart phone or any other suitable portable or fixed electronic device.

Also, a computer may have one or more input and output devices. These devices can be used, among other things, to present a user interface. Examples of output devices that can be used to provide a user interface include printers or display screens for visual presentation of output and speakers or other sound generating devices for audible presentation of output. Examples of input devices that can be used for a user interface include keyboards, and pointing devices, such as mice, touch pads, and digitizing tablets. As another example, a computer may receive input information through speech recognition or in other audible format.

Such computers may be interconnected by one or more networks in any suitable form, including a local area network or a wide area network, such as an enterprise network, and intelligent network (IN) or the Internet. Such networks may be based on any suitable technology and may operate according to any suitable protocol and may include wireless networks, wired networks or fiber optic networks.

A computer employed to implement at least a portion of the functionality described herein may include a memory, one or more processing units (also referred to herein simply as "processors"), one or more communication interfaces, one or more display units, and one or more user input devices. The memory may include any computer-readable media, and may store computer instructions (also referred to herein as "processor-executable instructions") for implementing the various functionalities described herein. The processing unit(s) may be used to execute the instructions. The communication interface(s) may be coupled to a wired or wireless network, bus, or other communication means and may therefore allow the computer to transmit communications to and/or receive communications from other devices. The display unit(s) may be provided, for example, to allow a user to view various information in connection with execution of the instructions. The user input device(s) may be provided, for example, to allow the user to make manual adjustments, make selections, enter data or various other information, and/or interact in any of a variety of manners with the processor during execution of the instructions.

The various methods or processes outlined herein may be coded as software that is executable on one or more processors that employ any one of a variety of operating systems or platforms. Additionally, such software may be written using any of a number of suitable programming languages and/or programming or scripting tools, and also may be compiled as executable machine language code or intermediate code that is executed on a framework or virtual machine.

In this respect, various inventive concepts may be embodied as a computer readable storage medium (or multiple computer readable storage media) (e.g., a computer memory, one or more floppy discs, compact discs, optical discs, magnetic tapes, flash memories, circuit configurations in Field Programmable Gate Arrays or other semiconductor devices, or other non-transitory medium or tangible computer storage medium) encoded with one or more programs that, when executed on one or more computers or other processors, perform methods that implement the various embodiments of the invention discussed above. The computer readable medium or media can be transportable, such that the program or programs stored thereon can be loaded onto one or more different computers or other processors to implement various aspects of the present invention as discussed above.

The terms "program" or "software" are used herein in a generic sense to refer to any type of computer code or set of computer-executable instructions that can be employed to program a computer or other processor to implement various aspects of embodiments as discussed above. Additionally, it should be appreciated that according to one aspect, one or more computer programs that when executed perform methods of the present invention need not reside on a single computer or processor, but may be distributed in a modular fashion amongst a number of different computers or processors to implement various aspects of the present invention.

Computer-executable instructions may be in many forms, such as program modules, executed by one or more computers or other devices. Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. Typically the functionality of the program modules may be combined or distributed as desired in various embodiments.

Also, data structures may be stored in computer-readable media in any suitable form. For simplicity of illustration, data structures may be shown to have fields that are related through location in the data structure. Such relationships may likewise be achieved by assigning storage for the fields with locations in a computer-readable medium that convey relationship between the fields. However, any suitable mechanism may be used to establish a relationship between information in fields of a data structure, including through the use of pointers, tags or other mechanisms that establish relationship between data elements.

Also, various inventive concepts may be embodied as one or more methods, of which an example has been provided. The acts performed as part of the method may be ordered in any suitable way. Accordingly, embodiments may be constructed in which acts are performed in an order different than illustrated, which may include performing some acts simultaneously, even though shown as sequential acts in illustrative embodiments.

As used herein, natural scene is to be understood to refer to an image of a natural environment, e.g., as described in Geisler WS Visual perception and the statistical of properties of natural scenes. Annu. Rev. Psychol. 59:167-92 (2008). In some embodiments, natural scenes may be replaced with any suitable complex image, e.g., an image characterized by a spatial and/or temporal frequency power spectrum that generally conforms to a inverse frequency squared law. In some embodiments, e.g., where a short video clip is used, the spectrum of the complex image may deviate somewhat from the inverse square law. For example, in some embodiments, the complex image may have a spatial or temporal a power spectrum of the form $1/f^x$, where f is the frequency and x is in the range of, e.g., 1-3, or any subrange thereof (e.g. 1.5-2.5, 1.75-2.25, 1.9-2.1, etc.)

A white noise image refers to a noise image having a spatial frequency power spectrum that is essentially flat.

As used herein the term "light" and related terms (e.g. "optical", "visual") are to be understood to include electromagnetic radiation both within and outside of the visible spectrum, including, for example, ultraviolet and infrared radiation.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, *i.e.*, elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "or" should be construed in the same fashion, *i.e.*, "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A or B", when used in conjunction with open-ended language such as "including" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "or" as defined above. For example, when separating items in a list, "or" or "or" shall be interpreted as being inclusive, *i.e.*, the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (*i.e.* "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

In the claims, as well as in the specification above, all transitional phrases such as "including," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, *i.e.*, to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

Variations, modifications and other implementations of what is described herein will occur to those of ordinary skill in the art without departing from the spirit and scope of the invention. While certain embodiments of the present invention have been shown and described, it will be obvious to those skilled in the art that changes and modifications may be made without departing from the spirit and scope of the invention. The matter set forth in the foregoing description and accompanying drawings is offered by way of illustration only and not as a limitation.

References

Ahuja A, Dorn J, Caspi A, McMahon M, Dagnelie G, Dacruz L, Stanga P, Humayun M, Greenberg R (2010) Blind subjects implanted with the Argus II retinal prosthesis are able to improve performance in a spatial-motor task. Br J Ophthalmol.

Arenkiel et al., In vivo light-induced activation of neural circuitry in transgenic mice expressing channelrhodopsin-2. Neuron (2007) 54(2): 205-18.

Ausubel et al., Current Protocols in Molecular Biology , John Wiley & Sons, New York, 1989

Bach, M et al (2008) Visual evoked potential-based acuity assessment in normal vision, artificially degraded vision, and in patients. Br J Ophthalmol 92:396-403

Barnstable et al., Thy-1 antigen: a ganglion cell specific marker in rodent retina. Neuroscience (1984) 11(4): 847-55.

Bi A, Cui J, Ma Y-P, Olshevskaya E, Pu M, Dizhoor AM, Pan Z-H (2006) Ectopic expression of a microbial-type rhodopsin restores visual responses in mice with photoreceptor degeneration. Neuron 50:23-33.

Bomash I, Roudi Y, Nirenberg S. (2010) A virtual retina that works on a broad array of stimuli including natural scenes: A tool to simplify the problem of population coding. *Society for Neuroscience.* Program No. 891.5.

Bookstein R et al (1990) Promoter deletion and loss of retinoblastoma gene expression in human prostate carcinoma. Proc. Nati. Acad. Sci. USA 87 (19):7762-7766

Busskamp V, et al (2010) Genetic reactivation of cone photoreceptors restores visual responses in retinitis pigmentosa. Science 329:413-417.

Cai et al. (2010) Gene delivery to mitotic and postmitotic photoreceptors via compacted DNA nanoparticles results in improved phenotype in a mouse model of retinitis pigmentosa. FASEB J. 24: 1178-1191.

Campagnola L, Wang H, Zylka MJ. (2008) Fiber-coupled light-emitting diode for localized photostimulation of neurons expressing channelrhodopsin-2. Journal of Neuroscience Methods. 169:27-33.

Cardin JA, et al (2010) Targeted optogenetic stimulation and recording of neurons in vivo using cell-type-specific expression of Channelrhodopsin-2. Nat Protoc 5(2): 247-54.

Chader GJ, Weiland J, Humayun MS (2009) Artificial vision: needs, functioning, and testing of a retinal electronic prosthesis. Prog Brain Res 175:317-332.

Chiappa, K. (1997) Evoked Responses in Clinical Medicine, Third Edition, Lippincott-Raven

Chichilnisky EJ. (2001) A simple white noise analysis of neuronal light responses. Network 12(2): 199-213

Chopdar A, Chakravarthy U, Verma D (2003) Age related macular degeneration. BMJ 326:485-488.

Cover T and Thomas J. (2006) Elements of Information Theory, 2nd Edition. Hoboken, NJ: Wiley

Dann JF, Buhl EH. (1987) Retinal ganglion cells projecting to the accessory optic system in the rat. J Comp Neurol 262(1): 141-58.

Dedek K, et al (2008) Ganglion cell adaptability: does the coupling of horizontal cells play a role? PLoS One. 3(3):e1714.

Douglas RM et al (2005) Independent visual threshold measurements in the two eyes of freely moving rats and mice using a virtual-reality optokinetic system. Vis Neurosci. 22(5):677-84.

Duda RO, Hart PE (2001) Pattern Classification (2nd Edition) Wiley, NY,

Enroth-Cugell et al., (1966) The contrast sensitivity of retinal ganglion cells of the cat. J Physiol 187(3): 517--52.

European Patent Application No. 19891976

Famulare M, Fairhall A. (2010) Feature selection in simple neurons: how coding depends on spiking dynamics. Neural Comput 22(3): 581-98

Field et al., (2007) Information processing in the primate retina: circuitry and coding. Annu Rev Neurosci 30: 1--30.

Fitzgerald et al. (1994) Retinal signal transmission in Duchenne muscular dystrophy. J Clin Invest 93: 2425-30.

Foley JM, Legge GE (1981) Contrast detection and near-threshold discrimination in human vision. Vision Res. 21(7):1041-53.

Fried S, Werblin F, McMahon MJ (2006) US Pat. 2006/0129207 Mimicking neural coding in retinal ganglion cells with short pulse electrical stimulation. In: (US, ed).

Friedman DS, O'Colmain BJ, Muñoz B, Tomany SC, McCarty C, de Jong PTVM, Nemesure B, Mitchell P, Kempen J, Eye Diseases Prevalence Research Group (2004) Prevalence of age-related macular degeneration in the United States. Arch Ophthalmol 122:564-572.

Geisler WS (200). Visual perception and the statistical properties of natural scenes. Annu. Rev. Psychol. 59:167-92 (2008)

Gerding H, Benner FP, Taneri S (2007) Experimental implantation of epiretinal retina implants (EPI-RET) with an IOL-type receiver unit. J Neural Eng 4:S38-49.

Giolli RA, Blanks RHI, Lui F. (2005) The accessory optic system: basic organization with an update on connectivity, neurochemistry, and function. Prog Brain Res 151: 407-40.

Golan L, Reutsky I, Farah N & Shoham S. (2009) Design and characteristics of holographic neural photo-stimulation systems, Journal of Neural Engineering 6 066004, (2009)

Graham-Rowe D (2009) A Brighter Future for Retinal Implants. In: Technology Review, http://www.technologyreview.com/biomedicine/23539/. Boston, MA: MIT.

Greenberg KP, et al. (2011). Differential Targeting of Optical Neuromodulators to Ganglion Cell Soma and Dendrites Allows Dynamic Control of Center-Surround Antagonism. Neuron 69, 713–720.

Grinstead CM and Snell JL (1997) Introduction to Probability. American Mathematical Society; 2nd Revised edition

Grossman N, Poher V, Grubb MS, Kennedy GT, Nikolic K, McGovern B, Palmini RB, Gong Z, Drakakis EM, Neil, MAA, Dawson MD, Burrone J, Degenaar P. (2010) Multi-site optical excitation using ChR2 and micro-LED array. *J. Neural Eng*, 7(1):1-13.

Han et al, (2009), Millisecond-Timescale Optical Control of Neural Dynamics in the Nonhuman Primate Brain, Neuron 62, 191-198.

Hand DJ. (1981) Discrimination and classification. Wiley Series in Probability and Mathematical Statistics.

Huberman AD, Manu M, Koch SM, Susman MW, Lutz AB, Ullian EM, Baccus SA, Barres BA (2008) Architecture and activity-mediated refinement of axonal projections from a mosaic of genetically identified retinal ganglion cells. Neuron. 2008 Aug 14;59(3):425-38.

Huberman AD, Wei W, Elstrott J, Stafford BK, Feller MB, Barres BA (2009) Genetic Identification of an On-Off Direction-Selective Retinal Ganglion Cell Subtype Reveals a Layer-Specific Subcortical Map of Posterior Motion. *Neuron*. 62(3):327-334.

Ivanova E, Pan Z-H (2009) Evaluation of the adeno-associated virus mediated long-term expression of channelrhodopsin-2 in the mouse retina. *Molecular Vision* 15:1680-1689

Izhikevich EM (2007) Dynamical systems in neuroscience: the geometry of excitability and bursting. Cambridge, MA: MIT Press

Izhikevich EM (2010) Hybrid spiking models. Review. Phil. Trans. R. Soc. A (2010) 368, 5061–5070

Jacobs AL et al (2009) ,Ruling out and ruling in neural codes. Proc Natl Acad Sci U S A. 106(14):5936-41.

Jeffreys, Harold (1961). The Theory of Probability. The Oxford University Press.

Kass RE, Ventura V, Brown EN. (2005) Statistical issues in the analysis of neuronal data. J Neurophysiol 94(1): 8-25.

Kawasaki et al., Variability of the relative afferent pupillary defect. Am J Ophthalmol (1995). 120: 622–633.

Kay MA, Glorioso JC, Naldini L. (2001) Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. Nat Med. 7(1):33-40. Review.

Kelly S, Shire D, Chen J, Doyle P, Gingerich M, Drohan W, Theogarajan L, Cogan S, Wyatt J, Rizzo JI (2009) Realization of a 15-channel, hermetically-encased wireless subretinal prosthesis for the blind. In, pp 200--203.

Kibbel S, Harscher A, Wrobel W-G, Zrenner E, Rothermel A (2009) Design and Performance of an improved active subretinal chip. In: World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany (Kim SI, Suh TS, Dössel O, Schlegel WC, eds), pp 192-195: Springer Berlin Heidelberg.

Koilkonda RD, Hauswirth WW, Guy J. (2009) Efficient expression of self-complementary AAV in ganglion cells of the ex vivo primate retina. Mol Vis. 15:2796-802.

Kuffler SW. (1953) Discharge patterns and functional organization of mammalian retina. J Neurophysiol 16(1): 37-68.

Lagali PS, Balya D, Awatramani GB, Münch TA, Kim DS, Busskamp V, Cepko CL, Roska B (2008) Light-activated channels targeted to ON bipolar cells restore visual function in retinal degeneration. Nat Neurosci 11:667-675.

Lesica NA et al. (2007) Adaptation to stimulus contrast and correlations during natural visual stimulation. Neuron 55(3): 479-491.

Lettvin et al., (1959) What the frog's eye tells the frog's brain. Proceedings of the Institute of Radio Engineers 47(11): 1940-51.

Liao et al. (2007) In vivo gene delivery in the retina using polyethylenimine. BioTechniques 2007, 42:285-288.

Loewenstein JI, Montezuma SR, Rizzo JF, III (2004) Outer Retinal Degeneration: An Electronic Retinal Prosthesis as a Treatment Strategy. Arch Ophthalmol 122:587-596.

Maguire et al. Safety and efficacy of gene transfer for Leber's congenital amaurosis. N Engl J Med (2008) 358: 2240–2248.

Mancuso et al., (2009) Gene therapy for red-green colour blindness in adult primates. Nature 461(7265): 784-7.

Martin et al. 2002. Gene delivery to the eye using adeno-associated viral vectors. Methods 28:267-275.

McGowan MH et al (1998) Characterization of the Mouse Aldose Reductase Gene and Promoter in a Lens Epithelial Cell Line. Mol Vis 1998; 4:2

McLaughlin SK, Collis P, Hermonat PL, Muzyczka N. (1988) Adeno-associated virus general transduction vectors: analysis of proviral structures. J Virol. 62(6):1963-73.

Meytlis M, Bomash I, Pillow JW, Nirenberg S. (2009) Assessing the importance of correlated firing using large populations of neurons. *Society for Neuroscience.* Program No. 165.3.

Morgans CW, et al (2009) TRPM1 is required for the depolarizing light response in retinal ON-bipolar cells. Proc Natl Acad Sci U S A 106(45): 19174-8.

Nanduri D, Humayun M, Greenberg R, McMahon M, Weiland J (2008) Retinal prosthesis phosphene shape analysis. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp 1785--1788. Vancouver, BC.

Nichols Z, Meytlis M, Nirenberg S. (2010) Correlations play a negligible role in coding white noise and natural scene stimuli in complete retinal populations. *Submitted.*

Nirenberg S (2000) Photoablation of cells expressing beta-galactosidase. *Methods Mol Biol.* **135**:475-80

Nirenberg S and Cepko, C (1993). Targeted ablation of diverse cell classes in the nervous system in vivo. *J Neurosci.* 13(8):3238-51.

Nirenberg S and Latham PE. (1998) Population coding in the retina. *Curr. Opin. Neurobiol.* **8**(4):488-493

Nirenberg S and Meister M. (1997). The light response of retinal ganglion cells is truncated by a displaced amacrine circuit. *Neuron* **18**:637-650

Nirenberg S et al (2001) Retinal ganglion cells act largely as independent encoders. Nature 411(6838):698-701.

Nirenberg S et al. (2010) Heterogeneous response dynamics in retinal ganglion cells: the interplay of predictive coding and adaptation. J Neurophysiol 103(6): 3184-94

Norcia, AM, and Tyler, CW (1985) Spatial frequency sweep VEP: visual acuity during the first year of life. Vision Res. 25(10):1399-408

Norcia, AM, et al (1989). Measurement of spatial contrast sensitivity with the swept contrast VEP. Vision Res. 1989;29(5):627-37.

Okuyama et al. (1993). Binocular infrared optometer for measuring accommodation in both eyes simultaneously in natural-viewing conditions Applied Optics, Vol. 32. No 22, p. 4147

Pandarinath et al (2010a) A novel mechanism for switching a neural system from one state to another. Front Comput Neurosci. 31;4:2.

Pandarinath et al (2010b) Symmetry breakdown in the ON and OFF pathways of the retina at night: functional implications. J Neurosci 30(30): 10006-14.

Paninski L, Pillow J, Lewi J. (2007) Statistical models for neural encoding, decoding, and optimal stimulus design. Prog Brain Res. 165:493-507.

Paninski L. (2004) Maximum likelihood estimation of cascade point-process neural encoding models. Network 15(4): 243-62

Panzeri S, et al (2007) Correcting for the sampling bias problem in spike train information measures. J Neurophysiol. 98(3):1064-72. Review.

Pelli DG, Robson JG, & Wilkins AJ (1988) The design of a new letter chart for measuring contrast sensitivity. Clinical Vision Sciences 2, 187-199

Perry VH, Silveira LC. (1988) Functional lamination in the ganglion cell layer of the macaque's retina. Neuroscience. 25(1):217-23.

Petrs-Silva et al., (2009) High-efficiency transduction of the mouse retina by tyrosine-mutant AAV serotype vectors. Mol Ther 17(3): 463-71.

Petersen-Jones et al., (2009) AAV retinal transduction in a large animal model species: comparison of a self-complementary AAV2/5 with a single-stranded AAV2/5 vector. Mol Vis 15: 1835-42.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP. (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454(7207): 995-9

Prusky GT, et al (2004) Rapid quantification of adult and developing mouse spatial vision using a virtual optomotor system. Invest Ophthalmol Vis Sci. 45(12):4611-6.

Published PCT Application WO1996013598

Published PCT application WO1998048027

Published PCT Application WO2000015822

Published PCT application WO2001094605
Published PCT application WO2002082904
Published PCT application WO2003047525
Published PCT application WO2003080648
Published PCT application WO2003093479
Published PCT application WO2003104413
Published PCT application WO2005080573
Published PCT application WO2007127428
Published PCT application WO2010011404
Pun L (1969), Introduction to Optimization Practice, ISBN 471-70233-1
Purpura K, Tranchina D, Kaplan E, Shapley RM. (1990) Light adaptation in the primate retina: analysis of changes in gain and dynamics of monkey retinal ganglion cells. Vis Neurosci 4(1): 75--93.
Rolls ET, Baylis GC, Leonard CM. Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey. Vision Res. 1985;25(8):1021-35.
Sambrook et al, "Molecular Cloning. A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory, New York (1989)
Sauer B. (1987) Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. Mol Cell Biol. 7(6):2087-96.
Shapley RM, Victor JD. (1981) How the contrast gain control modifies the frequency responses of cat retinal ganglion cells. J Physiol. 318:161-79.
Sharpee TO et al. (2008) On the Importance of Static Nonlinearity in Estimating Spatiotemporal Neural Filters With Natural Stimuli. J Neurophysiol 99(5): 2496-509
Sheridan C (2011) Gene Therapy finds its niche Nature Biotechnology 29(2):121-128
Siegert S, Scherf BG, Punta KD, Didkovsky N, Heintz N, Roska B (2009). Genetic address book for retinal cell types. *Nature Neuroscience*. 12:1197-1204.
Simoncelli et al. (2004) Characterization of neural responses with stochastic stimuli. The cognitive neurosciences: 327—38

Simonelli et al. (2010) Gene Therapy for Leber's Congenital Amaurosis is Safe and Effective Through 1.5 Years After Vector Administration, Molecular Therapy 18 3, 643–650.

Sinclair JR,et al (2004) .Selective ablation of a class of amacrine cells alters spatial processing in the retina. J Neurosci. 24(6):1459-67.

Sjostrand et al. (1999). Morphometric study of the displacement of retinal ganglion cells subserving cones within the human fovea. Graefe's Arch Clin Exp Ophthalmol 237:1014–1023.

Soucy ER et al. (1998) A novel signaling pathway from rod photoreceptors to ganglion cells in mammalian retina. *Neuron* **21**:481-493

Stone et al., (1993) Response properties of ganglion cells in the isolated mouse retina. Vis Neurosci 10(1): 31-9.

Strong SP, et al (1998) On the application of information theory to neural spike trains. Pac Symp Biocomput. 621-32.

Thyagarajan S, van Wyk M, Lehmann K, Löwel S, Feng G, Wässle H (2010) Visual function in mice with photoreceptor degeneration and transgenic expression of channelrhodopsin 2 in ganglion cells. J Neurosci 30:8745-8758.

Tomita H, Sugano E, Isago H, Hiroi T, Wang Z, Ohta E, Tamai M (2010) Channelrhodopsin-2 gene transduced into retinal ganglion cells restores functional vision in genetically blind rats. Exp Eye Res 90:429-436.

Troy JB, Bohnsack DL, Chen J, Guo X, Passaglia CL. (2005) Spatiotemporal integration of light by the cat X-cell center under photopic and scotopic conditions. Vis Neurosci 22(4): 493--500.

Troy JB, Bohnsack DL, Diller LC. (1999) Spatial properties of the cat X-cell receptive field as a function of mean light level. Vis Neurosci 16(6): 1089--104.

Turchinovich et al. (2010) Non-viral siRNA delivery into the mouse retina in vivo. BMC Ophthalmology 10:25.

U.S. Patent 7149586

U.S. Patent 5,856,152

U.S. Patent 5,871,982

U.S. Patent 6533798

U.S. Patent Publication 20080221653

U.S. Patent Publication 20080249588

U.S. Patent Publication 20090088399

U.S. Patent Publication 20090326623,

U.S. Patent Publication 20100272688

U.S. Patent Publication Nos. 20070261127

Ueda et al, (1997) The mGluR6 5' upstream transgene sequence directs a cell-specific and developmentally regulated expression in retinal rod and ON-type cone bipolar cells. J Neurosci. 17(9):3014-23.

van Adel et al. (2003) Delivery of ciliary neurotrophic factor via lentiviral-mediated transfer protects axotomized retinal ganglion cells for an extended period of time. Hum. Gene Ther. 14:103-115.

Victor JD, Shapley RM. (1979) The nonlinear pathway of Y ganglion cells in the cat retina. J Gen Physiol. 74(6):671-89.

Victor JD. (1987) The dynamics of the cat retinal X cell centre. The Journal of Physiology 386(1): 219.

Volgyi B, Deans MR, Paul DL, Bloomfield SA (2004) Convergence and Segregation of the Multiple Rod Pathways in Mammalian Retina. J Neurosci 24(49):11182–11192.

Walther W, Stein U. (2000) Viral vectors for gene transfer: a review of their use in the treatment of human diseases. Drugs. 60(2):249-71. Review.

Wassle H. (2004) Parallel processing in the mammalian retina. Nat Rev Neurosci 5(10): 747-57.

Wells et al. (2005) Optical stimulation of neural tissue in vivo. Optics Letters 30(5):504-506,

Winter JO, Cogan SF, Rizzo JFI (2007) Retinal prostheses: current challenges and future outlook. J Biomater Sci Polym Ed 18:1031-1055.

Wright AF. (1997) Gene therapy for the eye. Br J Ophthalmol 81(8): 620-623 Review.

Yonehara K, Ishikane H, Sakuta H, Shintani T, Nakamura-Yonehara K, et al. (2009) Identification of Retinal Ganglion Cells and Their Projections Involved in Central

Transmission of Information about Upward and Downward Image Motion. PLoS ONE 4(1): e4320.

Yonehara K, Shintani T, Suzuki R, Sakuta H, Takeuchi Y, et al. (2008) Expression of SPIG1 Reveals Development of a Retinal Ganglion Cell Subtype Projecting to the Medial Terminal Nucleus in the Mouse. PLoS ONE 3(2): e1533.

Zhang Y, Ivanova E, Bi A, Pan Z-H (2009) Ectopic expression of multiple microbial rhodopsins restores ON and OFF light responses in retinas with photoreceptor degeneration. J Neurosci 29:9186-9196.

Zrenner E, et al (2009) Subretinal Microelectrode Arrays Allow Blind Retinitis Pigmentosa Patients to Recognize Letters and Combine them to Words. BMEI '09. 2nd International Conference on Biomedical Engineering and Informatics. Issue Date: 17-19 Oct. 2009. ISBN: 978-1-4244-4132-7. Pages 1 - 4.

Ballard DH et al. (1982) *Computer Vision*, Prentice-Hall Inc New Jersey, 1982

Chichilnisky EJ. (2001) A simple white noise analysis of neuronal light responses. Network 12(2): 199-213

Forsyth DA et al. (2003) *Computer Vision: A Modern Approach* , Prentice Hall, Nirenberg S et al. (2010) Heterogeneous response dynamics in retinal ganglion cells: the interplay of predictive coding and adaptation. J Neurophysiol 103(6): 3184-94.

Freund Y et al. (1995) A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting *Journal of Computer and System Sciences* 55, 119-139

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP. (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454(7207): 995-9

Simonelli et al. (2010) Gene Therapy for Leber's Congenital Amaurosis is Safe and Effective Through 1.5 Years After Vector Administration, Molecular Therapy 18 3, 643–650

Nichols Z, Meytlis M, Nirenberg S. (2010) Correlations play a negligible role in coding white noise and natural scene stimuli in complete retinal populations. *Submitted.*

Nirenberg S et al. (2010) Heterogeneous response dynamics in retinal ganglion cells: the interplay of predictive coding and adaptation. J Neurophysiol 103(6): 3184-94

CLAIMS

What is claimed is:

1. A method including:

receiving raw image data corresponding to a series of raw images;

processing the raw image data with an encoder to generate encoded data, where the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a vertebrate retina; and

applying a first machine vision algorithm to data generated based at least in part on the encoded data.

2. The method of any preceding claim, further including generating a series of retinal images based on the encoded data.

3. The method of claim 2, including determining pixel values in the retinal images based on the encoded data.

4. The method of claim 3, where determining pixel values in the retinal images based on the encoded data includes determining a pixel intensity or color based on encoded data indicative of a retinal cell response.

5. The method of claim 4, where the data indicative of a retinal cell response is indicative of at least one from the list consisting of: a retinal cell firing rate, a retinal cell output pulse train, and a generator potential.

6. The method of any one of claims 2-6, further including:

applying the first machine vision algorithm to the series of retinal images.

7. The method of claim 6, where the machine vision algorithm includes at least one select from the list consisting of: an object recognition algorithm, an image classification

algorithm, a facial recognition algorithm, an optical character recognition algorithm, a content-based image retrieval algorithm, a pose estimation algorithm, a motion analysis algorithm, an egomotion determination algorithm, a movement tracking algorithm, an optical flow determination algorithm, a scene reconstruction algorithm, a 3D volume recognition algorithm, and a navigation algorithm.

8. The method of any preceding claim, where the machine vision algorithm exhibits better performance when applied to the series of retinal images than when applied to a corresponding set of raw images that have not been processed using the encoder.

9. The method of claim 8, where the machine vision algorithm exhibits better performance when applied to a series of retinal images including natural scenes than when applied to a corresponding series of raw images that have not been processed using the encoder.

10. The method of claim 8 or 9, where the machine vision algorithm includes an algorithm for the detection or identification of a human within a series of images; and where the machine vision algorithm exhibits better detection or identification accuracy when applied to a range of retinal images including the human than when applied to a corresponding set of raw images that have not been processed using the encoder.

11. The method of claim 10, where the series of images including the human includes images of the human located in a natural scene.

12. The method of claim 11, where the series of images including the human includes images of the human located in a natural scene that is different from ~~any~~ natural scenes used to train the machine vision algorithm

13. The method of claim 8 or 9, where the machine vision algorithm includes an algorithm for navigation through a real or virtual environment, and where the machine vision algorithm exhibits better navigation performance when applied to a series of

retinal images including a natural scene than when applied to a corresponding set of raw images that have not been processed using the encoder.

14. The method of claim 13, where the machine vision algorithm exhibits fewer unwanted collision events during navigation when applied to a series of retinal images including a natural scene than when applied to a corresponding set of raw images that have not been processed using the encoder.

15. The method of claim 14, where the series of retinal images correspond to an environment that was not used to train the machine vision algorithm.

16. The method of any preceding claim, further including:

applying a machine imaging algorithm to the series of retinal images to identify one or more retinal images of interest; and

identifying one or more raw images of interest corresponding to the retinal images of interest.

17. The method of claim 16, further including processing the raw images of interest.

18. The method of claim 17, where the processing the raw images of interest includes applying a second machine vision algorithm to the raw images of interest.

19. The method of claim 18, where:

the first machine vision algorithm includes an algorithm that has been trained on a set of retinal images; and

the second machine vision algorithm includes an algorithm that has been trained on a set of raw images.

20. The method of any preceding claim, where applying the first machine vision algorithm includes applying a navigation algorithm,

21. The method of claim 20, where applying the navigation algorithm includes:

processing the series of retinal images to determine motion information indicative of motion at a plurality of image locations in the series of images;

classifying spatial regions in the series of images based on the motion information; and

generating a navigation decision based on the classification of the spatial regions.

22. The method of claim 21, where motion information is indicative of an optical flow in the series of images.

23. The method of claim 21 or 22, including:

using a convolutional neural network to classify the spatial regions.

24. The method of any one of claims 21-23 further including controlling the motion of a robotic apparatus based on results from navigation algorithm.

25. The method of any one of claims 18-24, further including controlling the motion of a virtual object in a virtual space based on results from navigation algorithm.

26. The method of claim 24 or 25, wherein the navigation algorithm was trained based on image data representative of a virtual space.

27. The method of any preceding claim, further including: training a machine vision algorithm based on the retinal images.

28. The method of claim 27, where training the machine vision algorithm includes:

(i) applying the machine vision algorithm to a set of retinal images to generate an output;

(ii) determining performance information indicative of the performance of the machine vision algorithm based on the output;

(iii) modifying one or more characteristics of the machine vision algorithm based on the performance information.

29. The method of claim 28, further including:

iteratively repeating steps (i) through (iii) until a selected performance criteria is reached.

30. The method of any one of claims 27-29, where the trained machine vision algorithm is characterized by a set of parameters, and where the parameters differ from the corresponding parameters that would be obtained by equivalent training of the machine vision algorithm using raw images corresponding to the retinal images.

31. The method of any one of claims 6-30, where:

processing the raw image data with an encoder to generate encoded data includes generating encoded data that contains a reduced amount of information relative to the corresponding raw image data; and

where the machine vision algorithm exhibits better performance when applied to the series of retinal images than when applied to a corresponding set of raw images that have not been processed using the encoder.

32. The method claim 31, where the amount of information contained in the encoded data is compressed by a factor of at least about 2 relative to the corresponding raw image data.

33. The method claim 31, where the amount of information contained in the encoded data is compressed by a factor of at least about 5 relative to the corresponding raw image data.

34. The method claim 31, where the amount of information contained in the encoded data is compressed by a factor of at least about 10 relative to the corresponding raw image data.

35. The method of any preceding claim, where the vertebrate includes at least one selected from the list consisting of: a mouse, and a monkey.

36. The method of any preceding claim, where the retinal cells include ganglion cells.

37. The method of any preceding claim, where the retinal cells include at least two classes of cells.

38. The method of any preceding claim, where the at least two classes of cells includes ON cells and OFF cells.

39. The method of any preceding claim, where the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a vertebrate retina over a range of input that includes natural scene images, including spatio-temporally varying images.

40. An apparatus including:

at least one memory storage device configured to store raw image data; and

at least one processor operably coupled with the memory and programmed to execute the method of any one of claims 1-38.

41. A non-transitory computer-readable medium having computer-executable instructions for implementing the steps of the method of any one of claims 1-38.

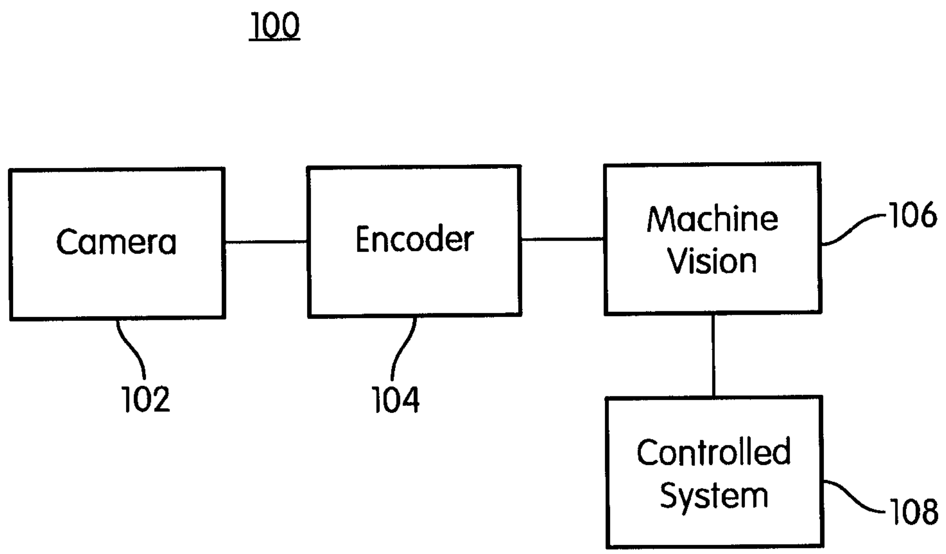

100
Camera
Encoder
Machine Vision
Controlled System
102
104
106
108

FIG. 1

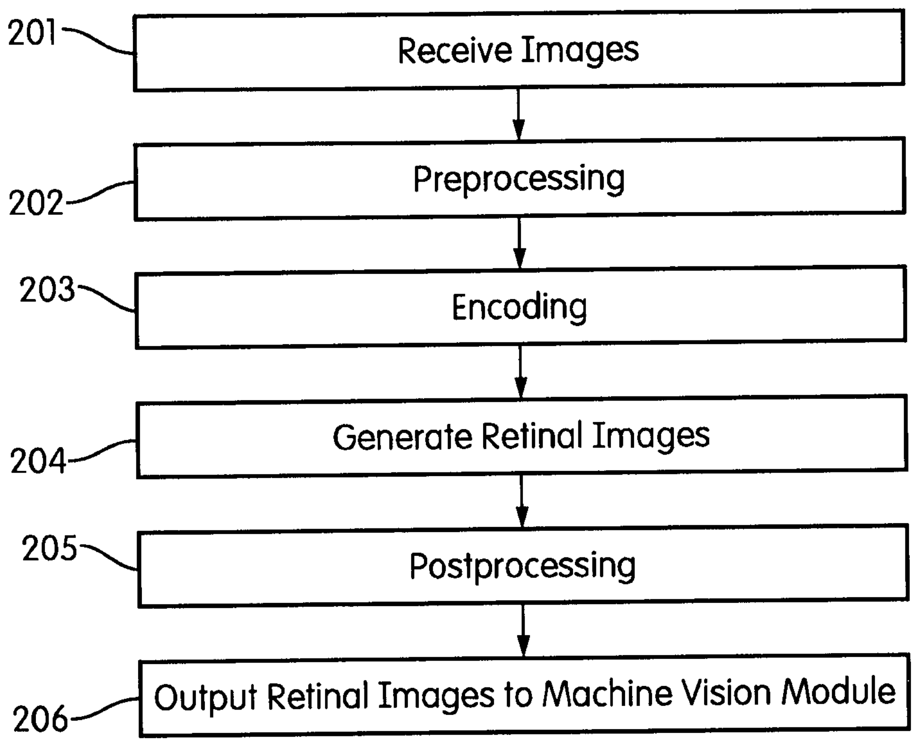

201
Receive Images
202
Preprocessing
203
Encoding
204
Generate Retinal Images
205
Postprocessing
206
Output Retinal Images to Machine Vision Module

FIG. 2


A Raw image stream
B Retinal image stream (Raw image processed through the encoders)
OFF midget
ON midget
OFF parasol
ON parasol

FIG. 3A

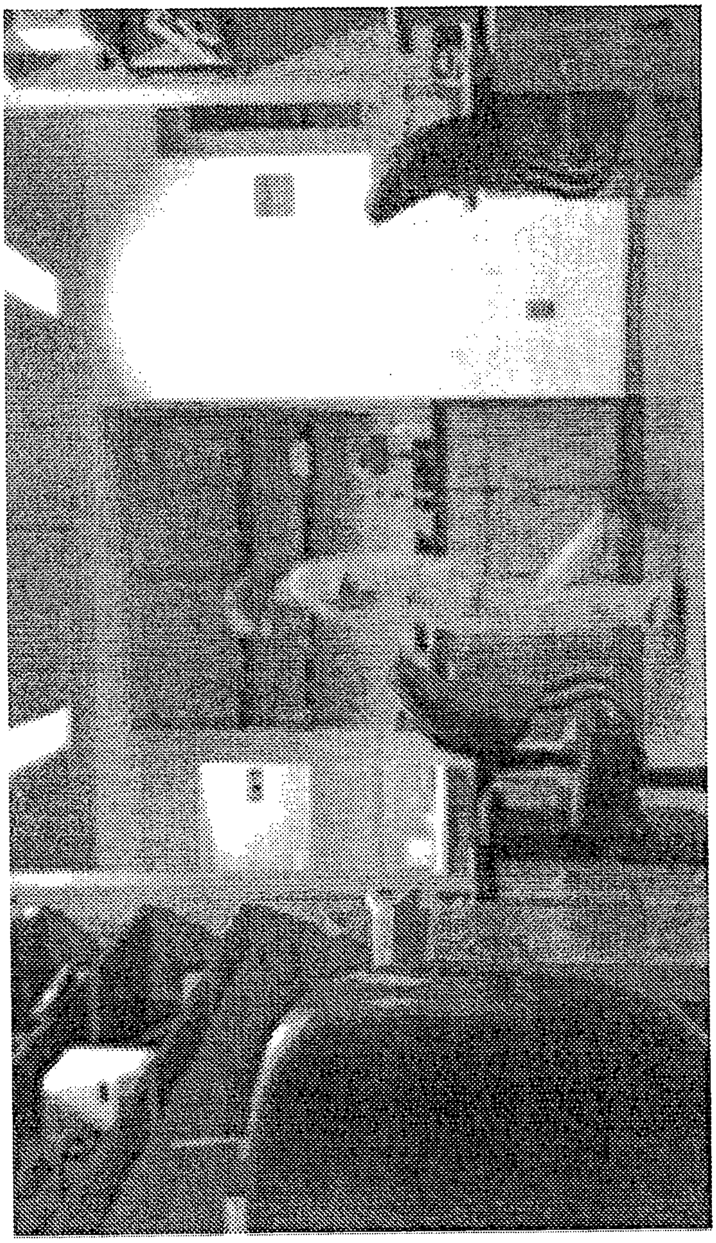


FIG. 3B

FIG. 3C

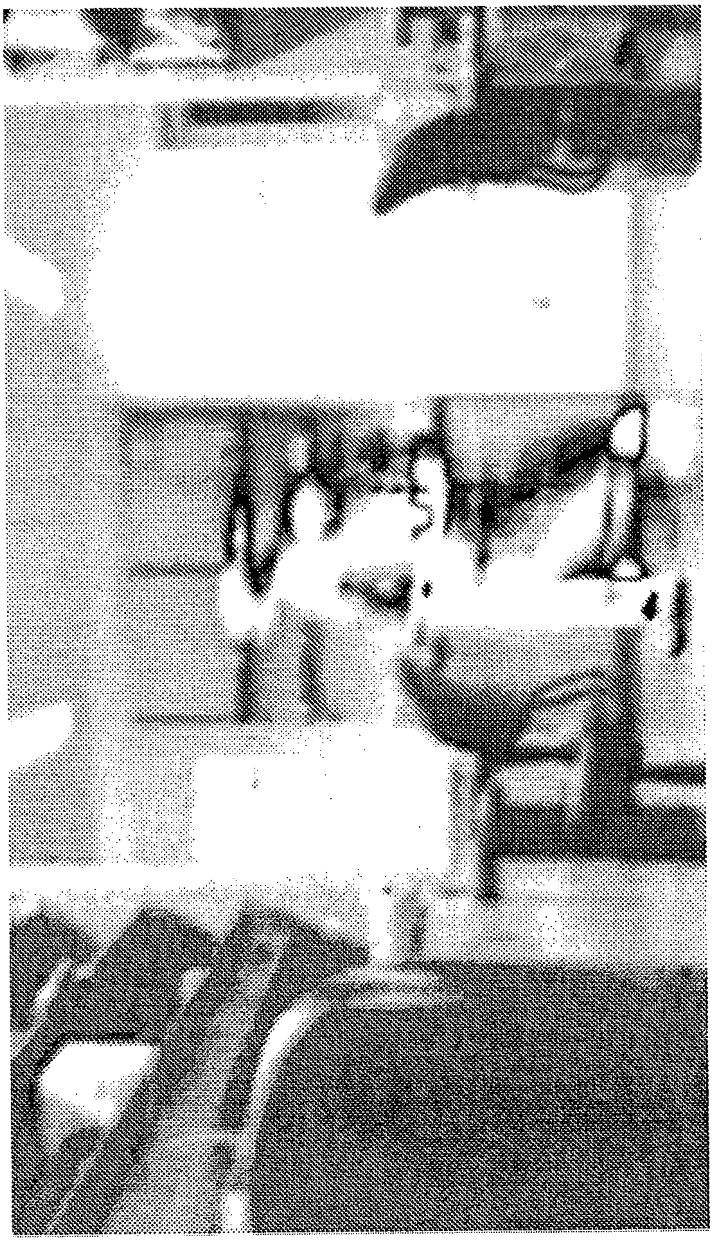


FIG. 3D

FIG. 3E

FIG. 3F

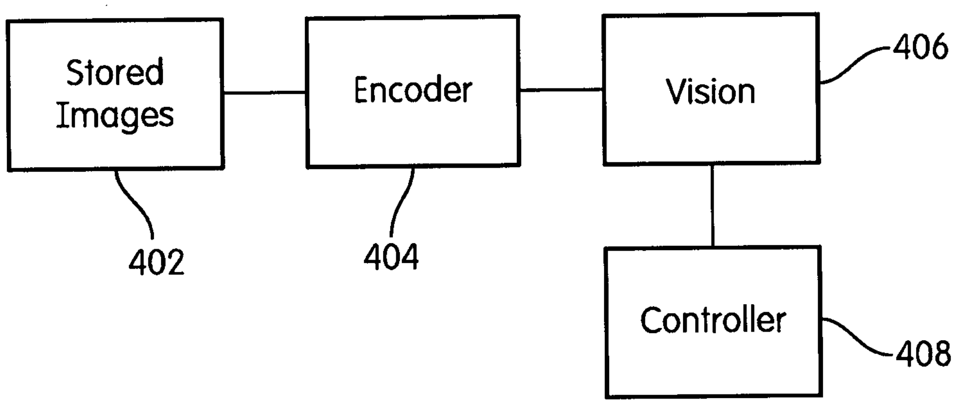


FIG. 4

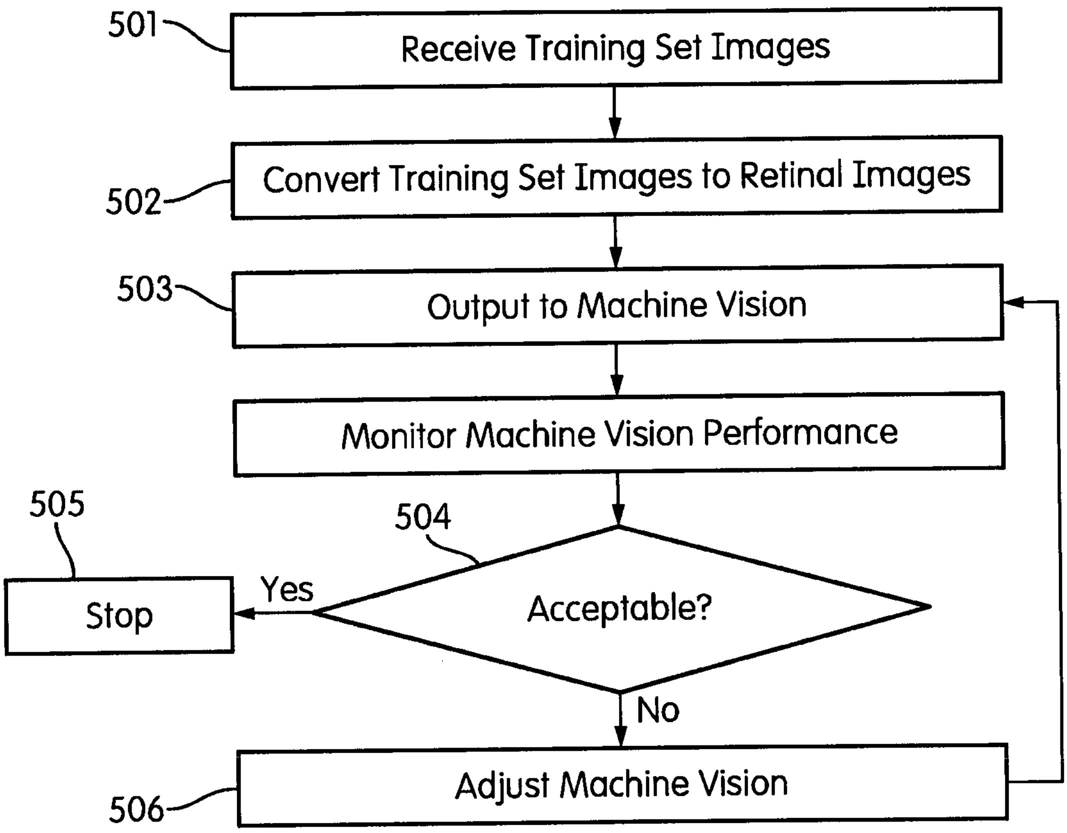

501
Receive Training Set Images
502
Convert Training Set Images to Retinal Images
503
Output to Machine Vision
Monitor Machine Vision Performance
505
504
Stop
Yes
Acceptable?
No
506
Adjust Machine Vision

FIG. 5

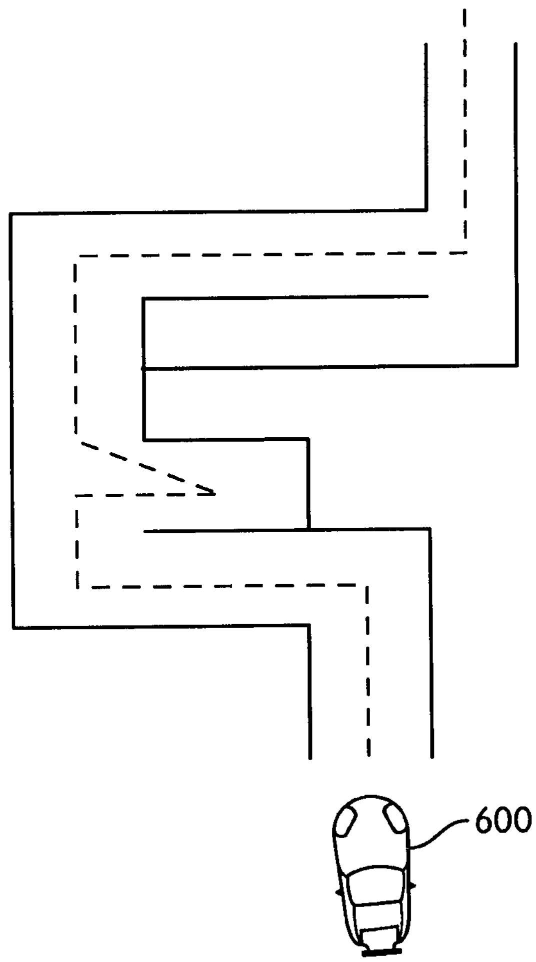


FIG. 6

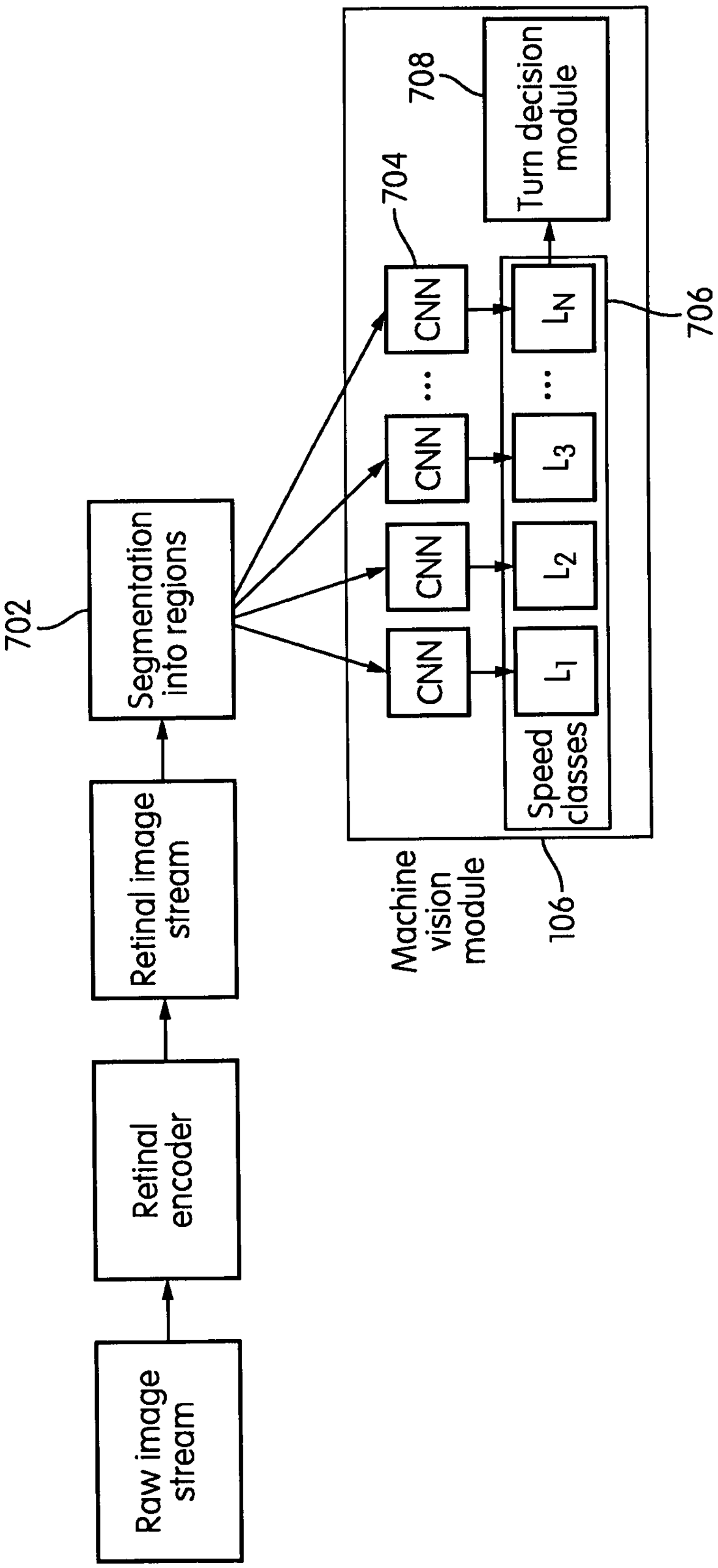

Raw image stream
Retinal encoder
Retinal image stream
Segmentation into regions
702
CNN
CNN
CNN
...
CNN
704
Speed classes
L_1
L_2
L_3
...
L_N
706
Turn decision module
708
Machine vision module
106

FIG. 7

First five frames or training sequence

Subsequent frames sampled more sparsely to cover the rest of the movie

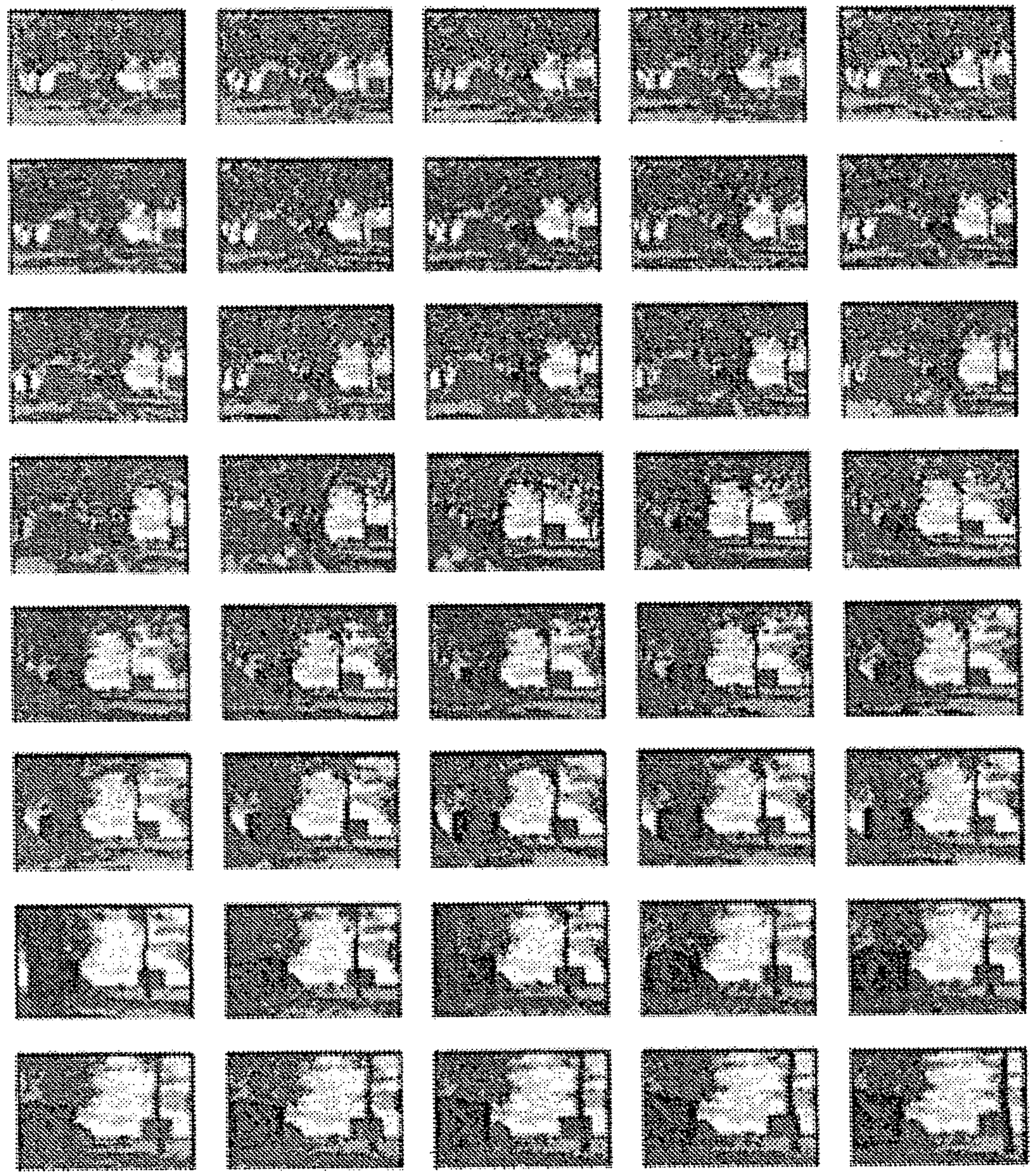
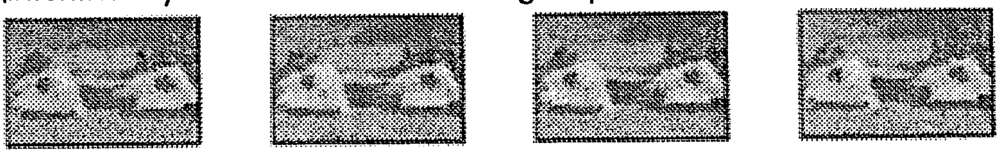


FIG. 8

Testing set 1: First four frames of rural sequence (intentionally different from the training sequence)



Subsequent frames sampled more sparsely to cover the rest of the movie

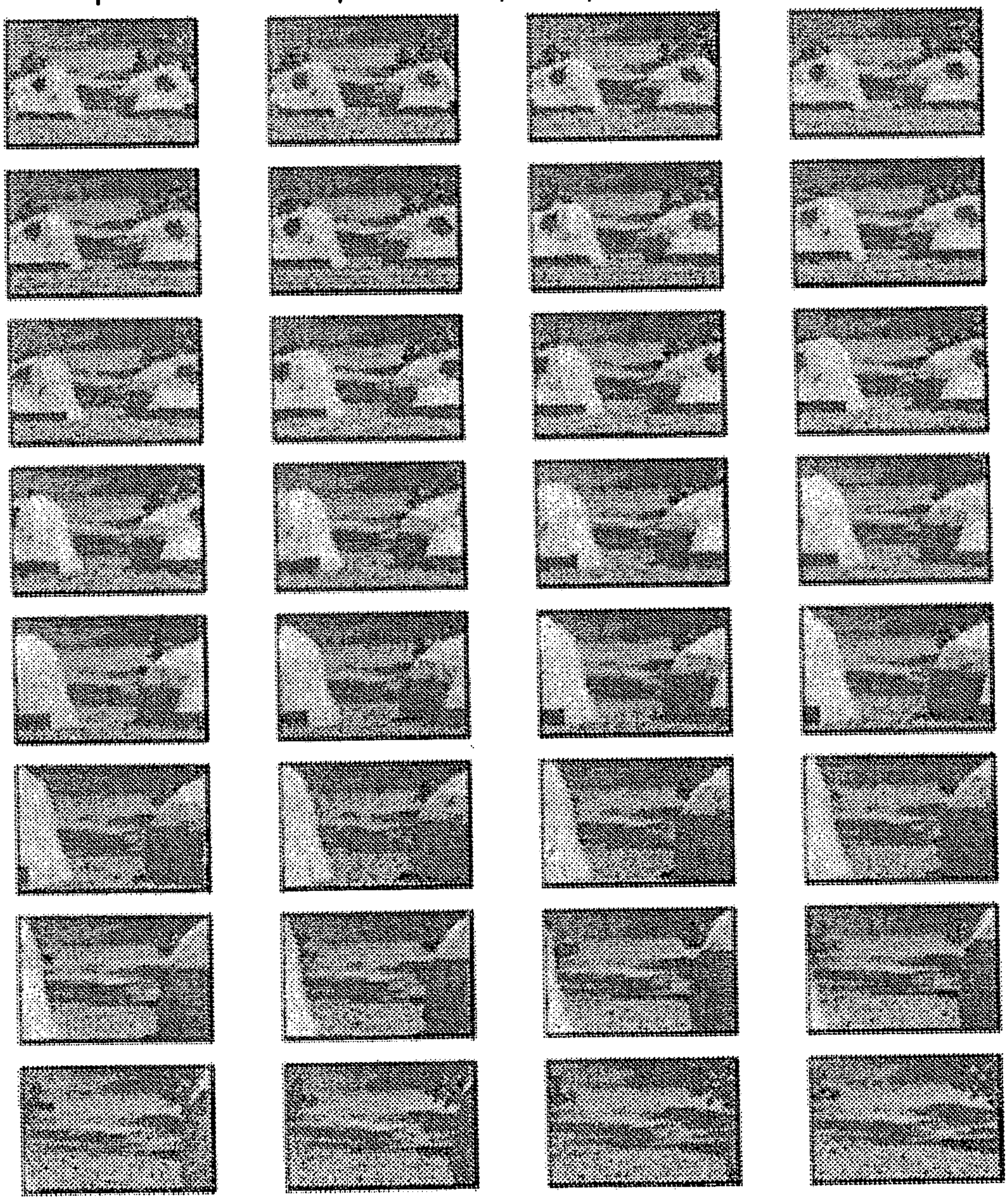
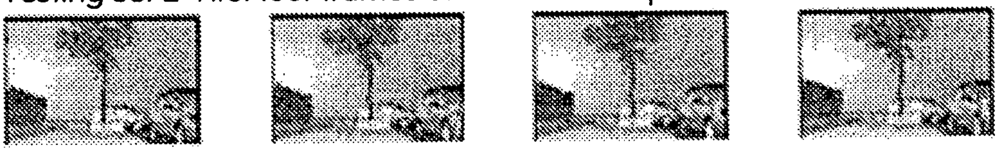


FIG. 9A

Testing set 2: First four frames of suburban sequence



Subsequent frames sampled more sparsely to cover the rest of the movie

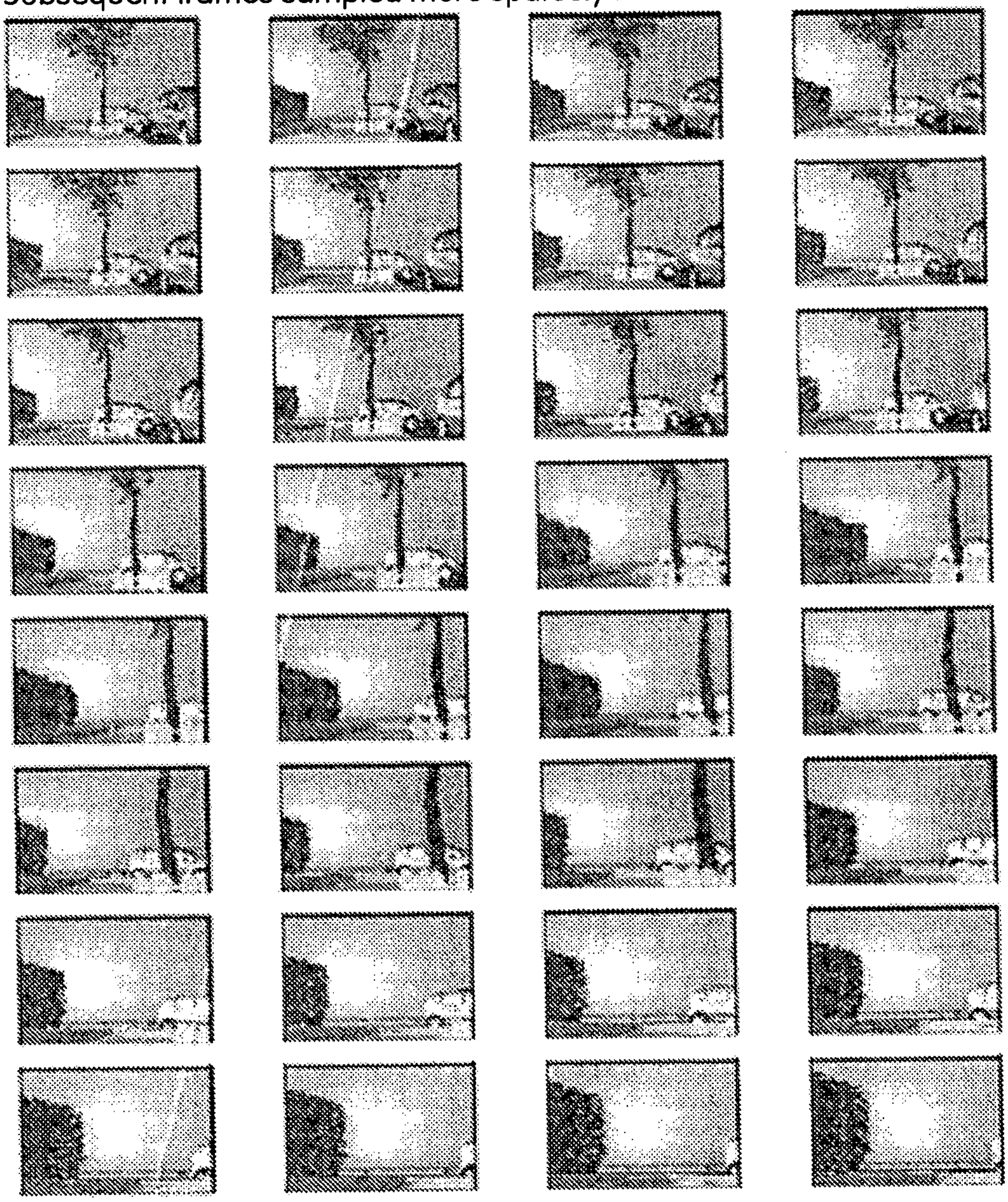
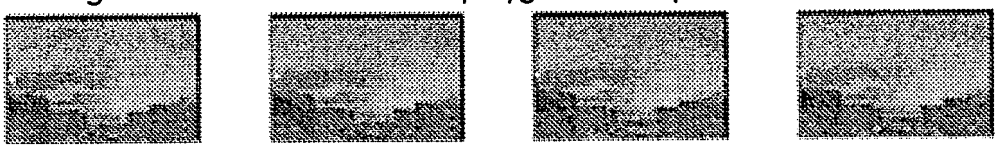


FIG. 9B

Testing set 3: First four frames of playground sequence



Subsequent frames sampled more sparsely to cover the rest of the movie

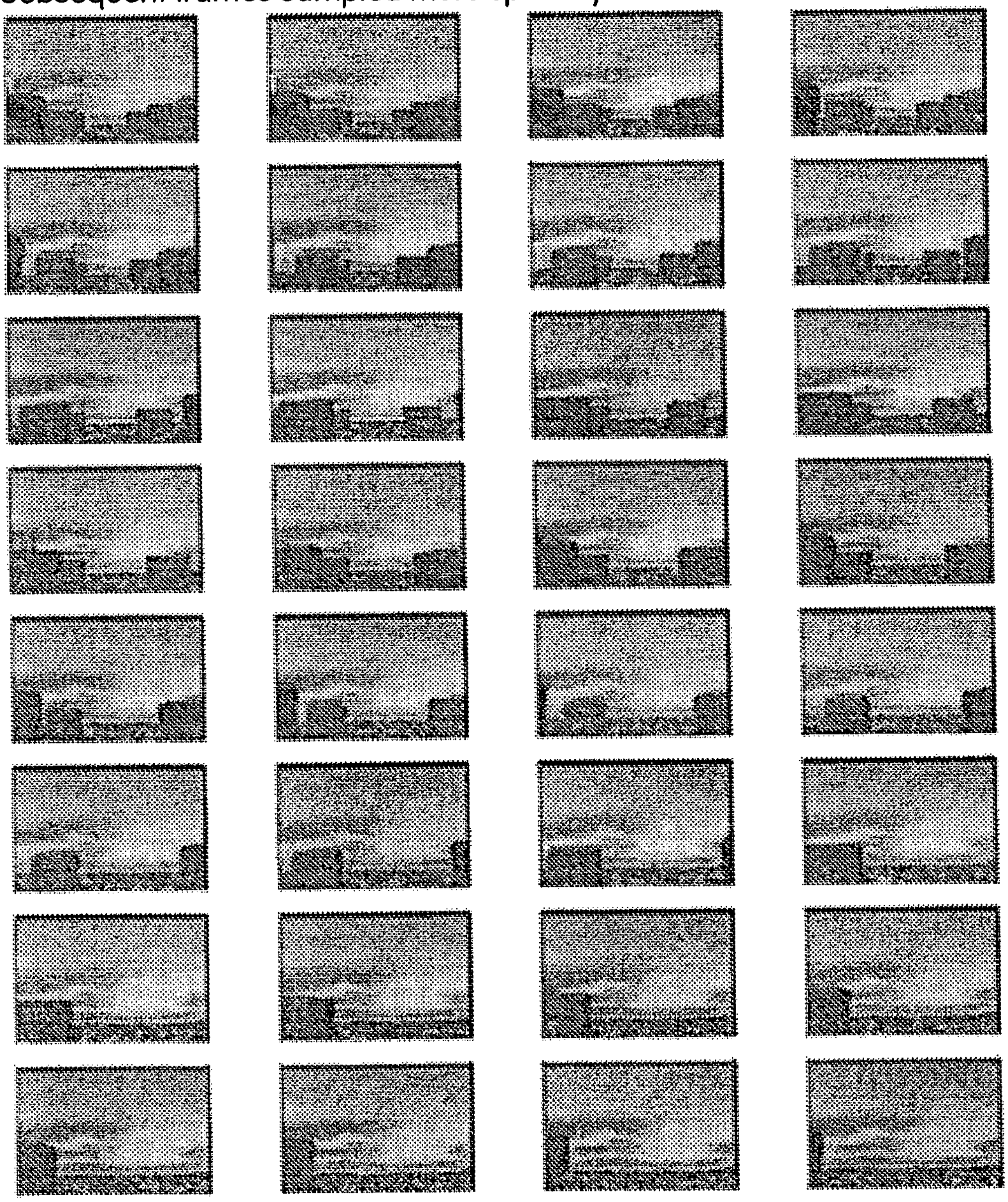


FIG. 9C

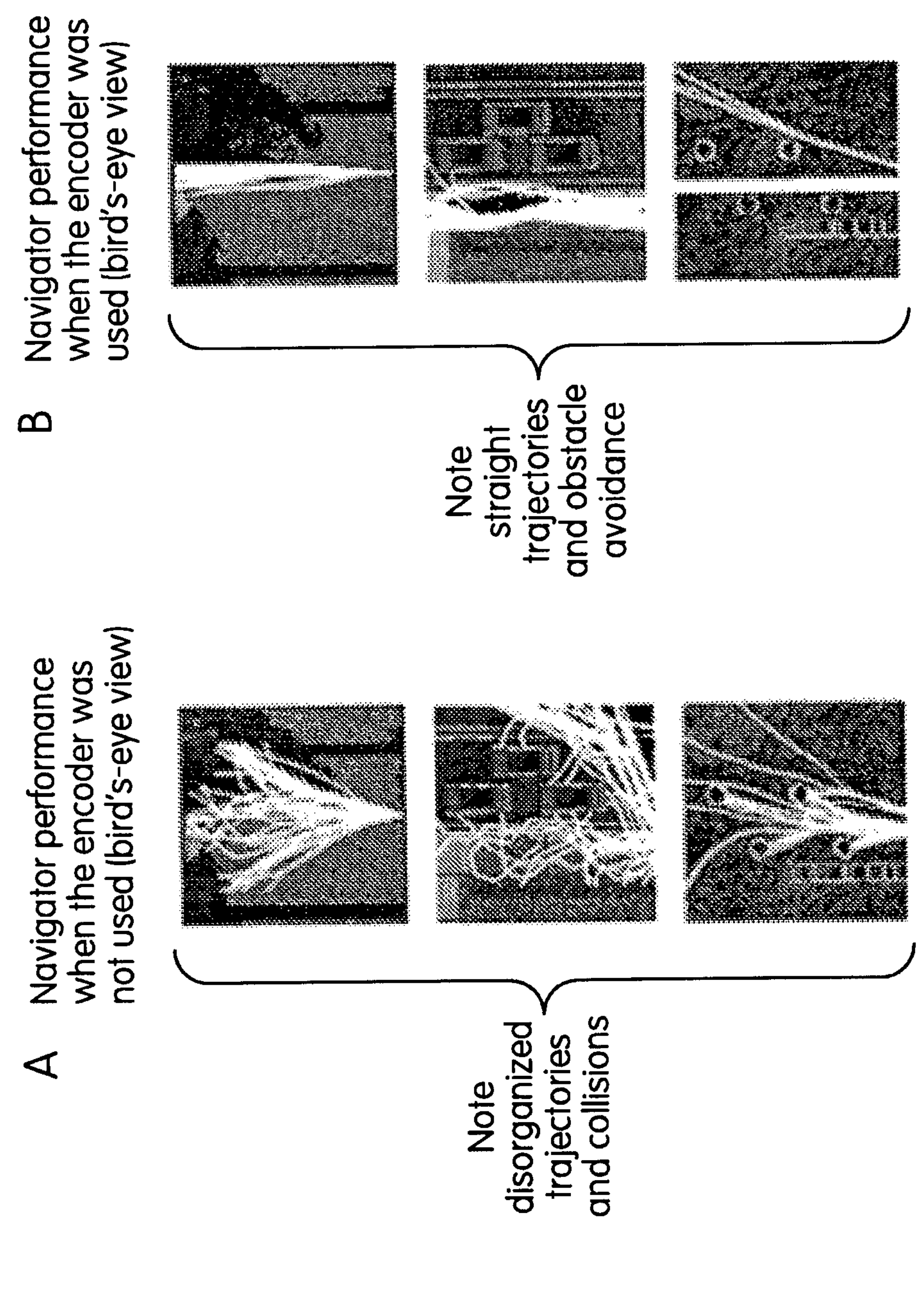

A
Navigator performance when the encoder was not used (bird's-eye view)
Note disorganized trajectories and collisions
B
Navigator performance when the encoder was used (bird's-eye view)
Note straight trajectories and obstacle avoidance

FIG. 10

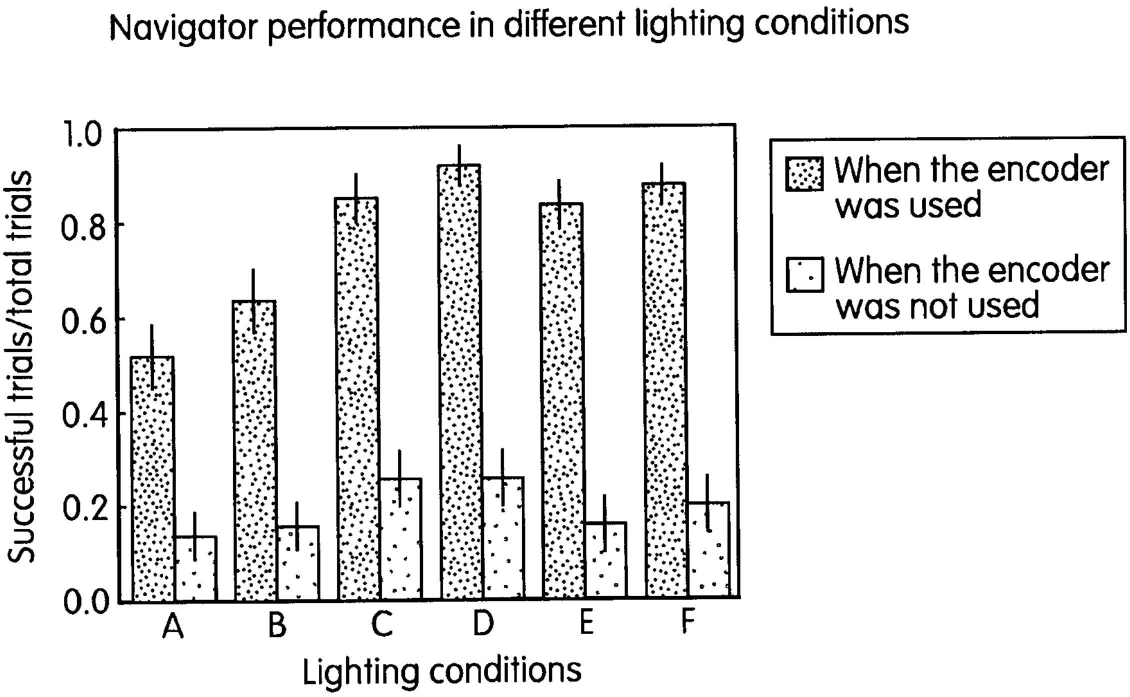

Navigator performance in different lighting conditions
1.0
0.8
0.6
0.4
0.2
0.0
Successful trials/total trials
A
B
C
D
E
F
Lighting conditions
When the encoder was used
When the encoder was not used

FIG. 11

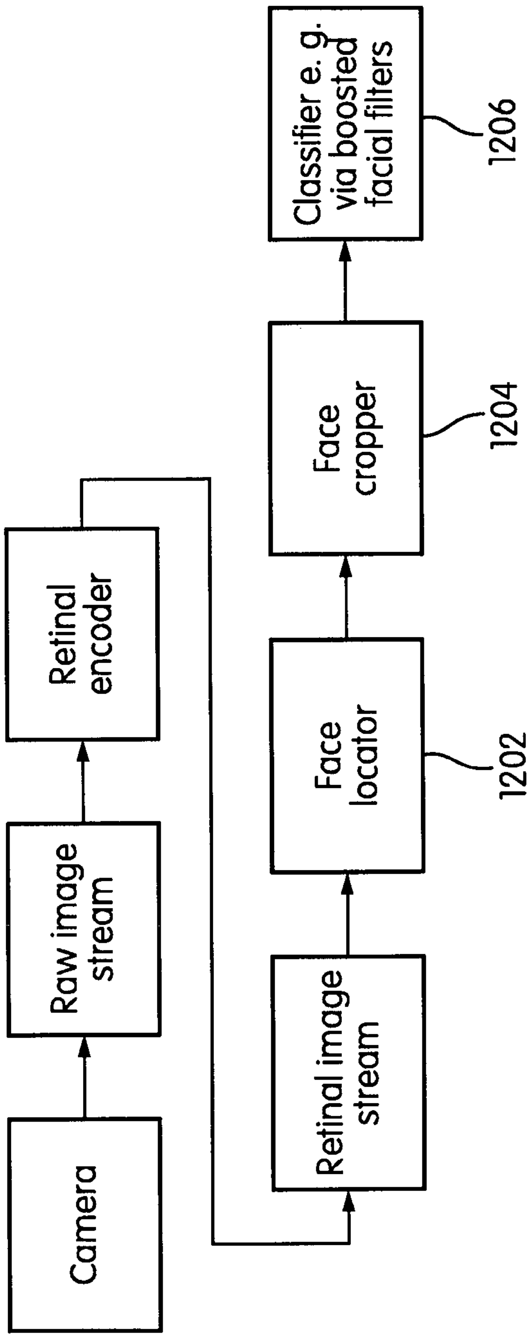

Camera
Raw image stream
Retinal encoder
Retinal image stream
Face locator
1202
Face cropper
1204
Classifier e. g. via boosted facial filters
1206

FIG. 12

Frames from an image stream used to train face recognizer

FIG. 13

Frames from an image stream used to test face recognizer

FIG. 14

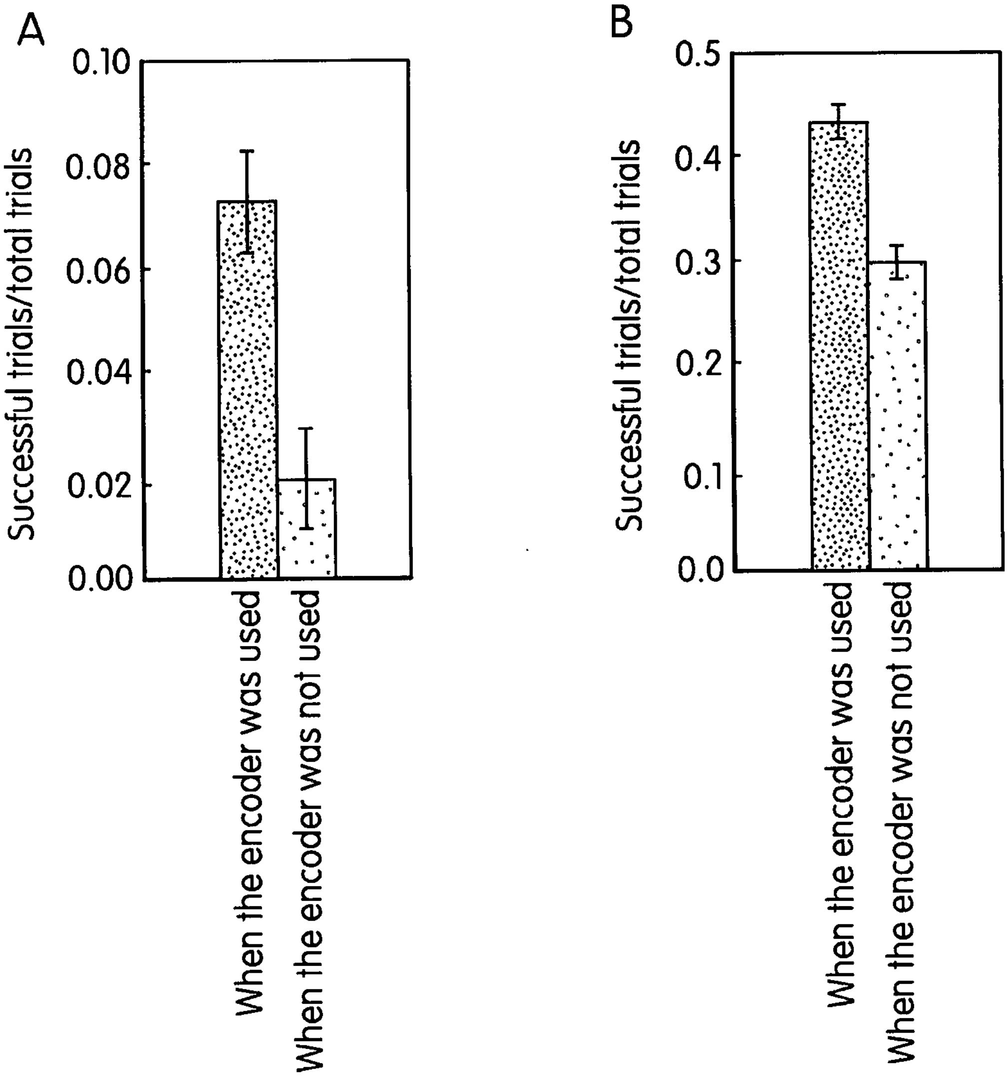

A
B
Successful trials/total trials
0.10
0.08
0.06
0.04
0.02
0.00
0.5
0.4
0.3
0.2
0.1
0.0
When the encoder was used
When the encoder was not used

FIG. 15

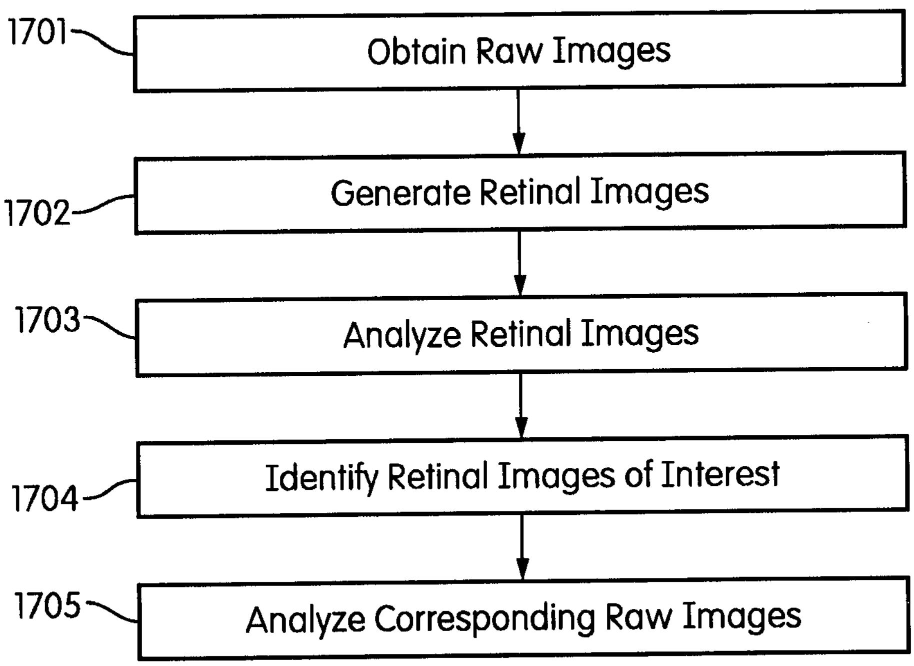

1701
Obtain Raw Images
1702
Generate Retinal Images
1703
Analyze Retinal Images
1704
Identify Retinal Images of Interest
1705
Analyze Corresponding Raw Images

FIG. 16

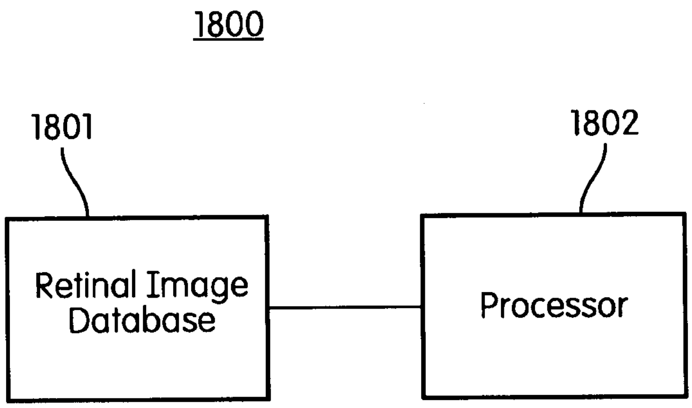


FIG. 17

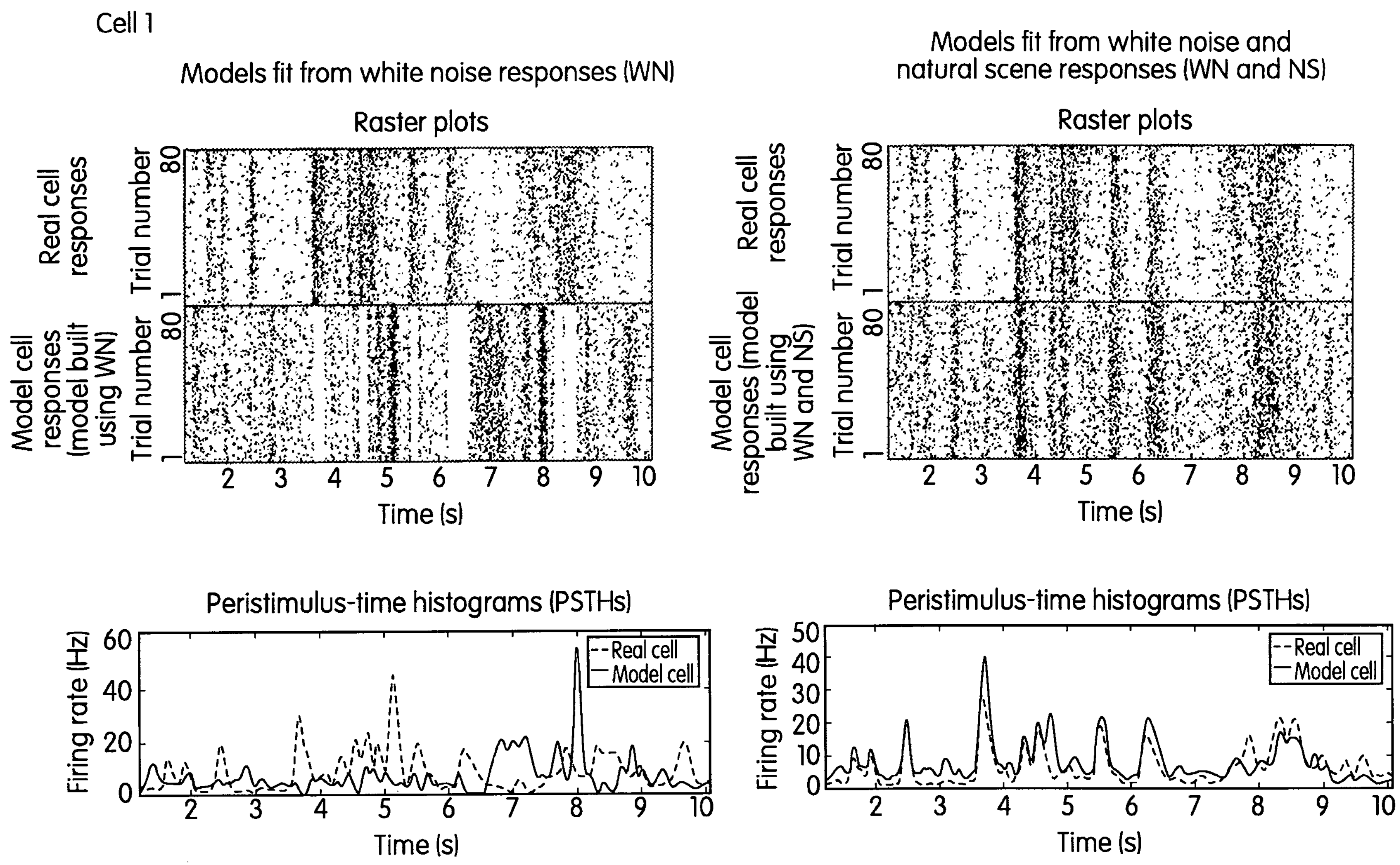

Cell 1
Models fit from white noise responses (WN)
Raster plots
Real cell responses
Model cell responses (model built using WN)
Trial number
80
1
2
3
4
5
6
7
8
9
10
Time (s)
Models fit from white noise and natural scene responses (WN and NS)
Model cell responses (model built using WN and NS)
Peristimulus-time histograms (PSTHs)
Firing rate (Hz)
60
40
20
0
50
30
10
Real cell
Model cell

FIG. 18A

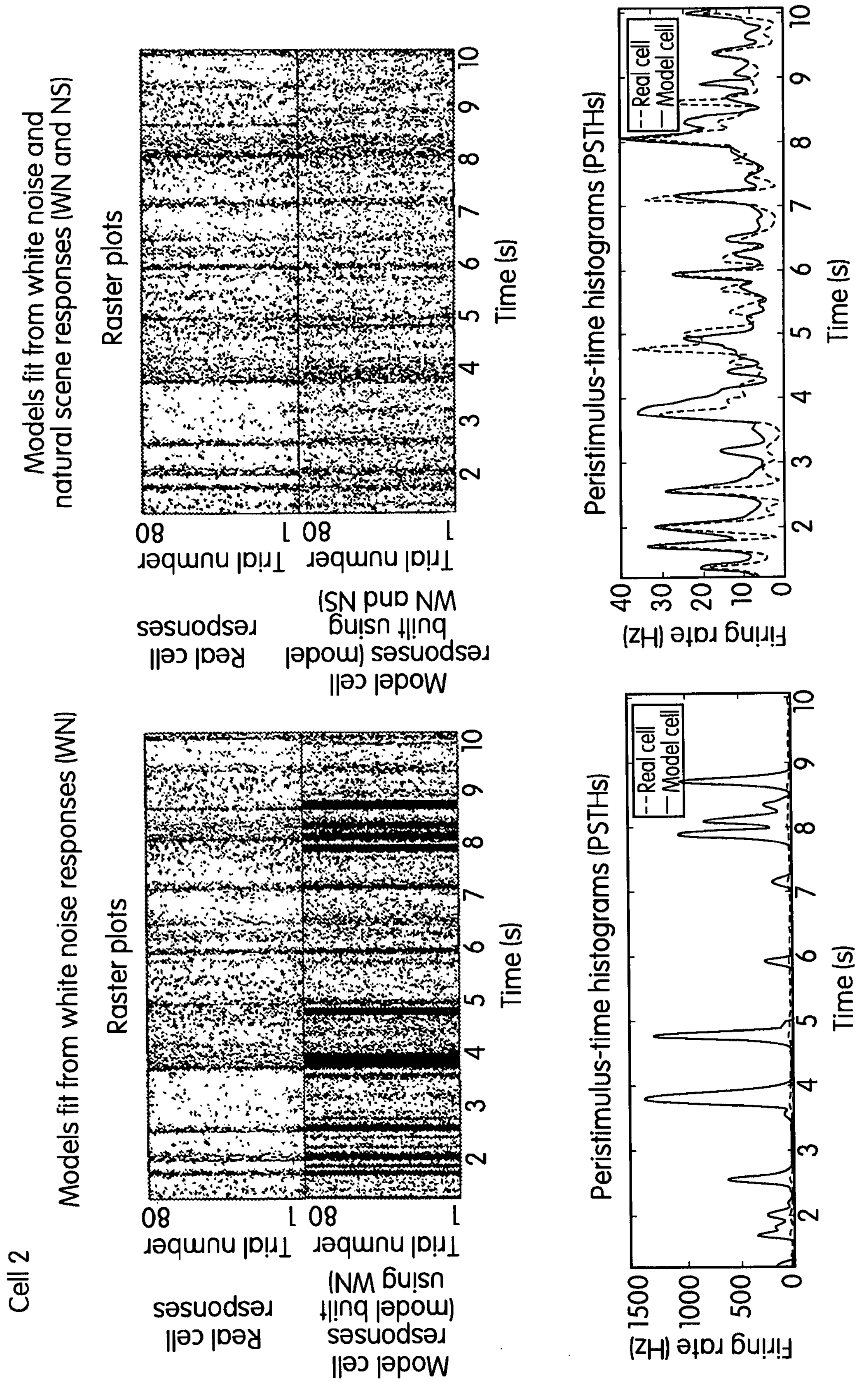

Cell 2
Models fit from white noise responses (WN)
Raster plots
Real cell responses
Model cell responses (model built using WN)
Trial number
80
1
Time (s)
2
3
4
5
6
7
8
9
10
Peristimulus-time histograms (PSTHs)
Firing rate (Hz)
1500
1000
500
0
-- Real cell
— Model cell
Models fit from white noise and natural scene responses (WN and NS)
Model cell responses (model built using WN and NS)
40
30
20
10

FIG. 18B

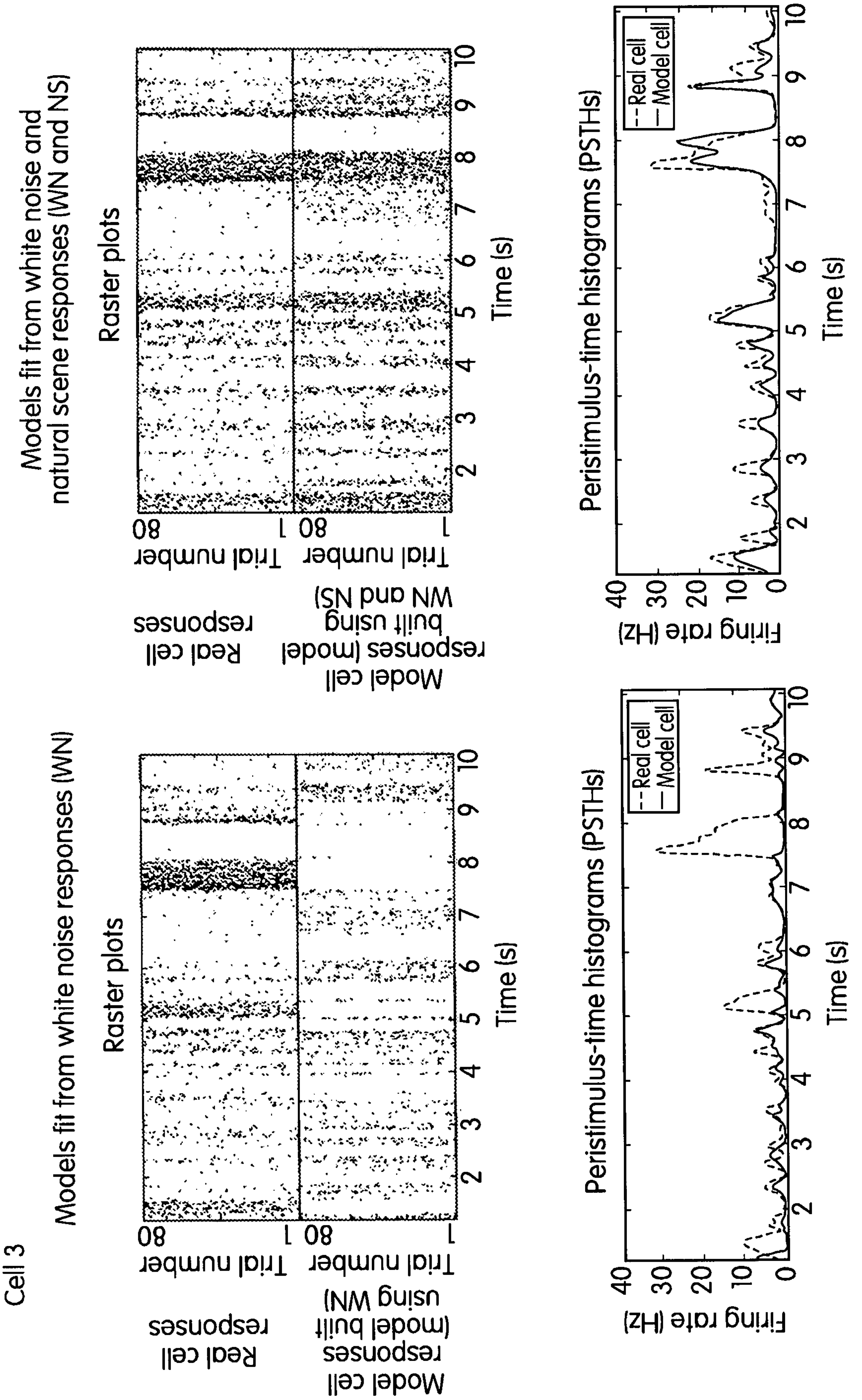

Cell 3
Models fit from white noise responses (WN)
Raster plots
Real cell responses
Model cell responses (model built using WN)
Trial number
80
1
Time (s)
2
3
4
5
6
7
8
9
10
Peristimulus-time histograms (PSTHs)
Firing rate (Hz)
40
30
20
10
0
Real cell
Model cell
Models fit from white noise and natural scene responses (WN and NS)
Model cell responses (model built using WN and NS)

FIG. 18C

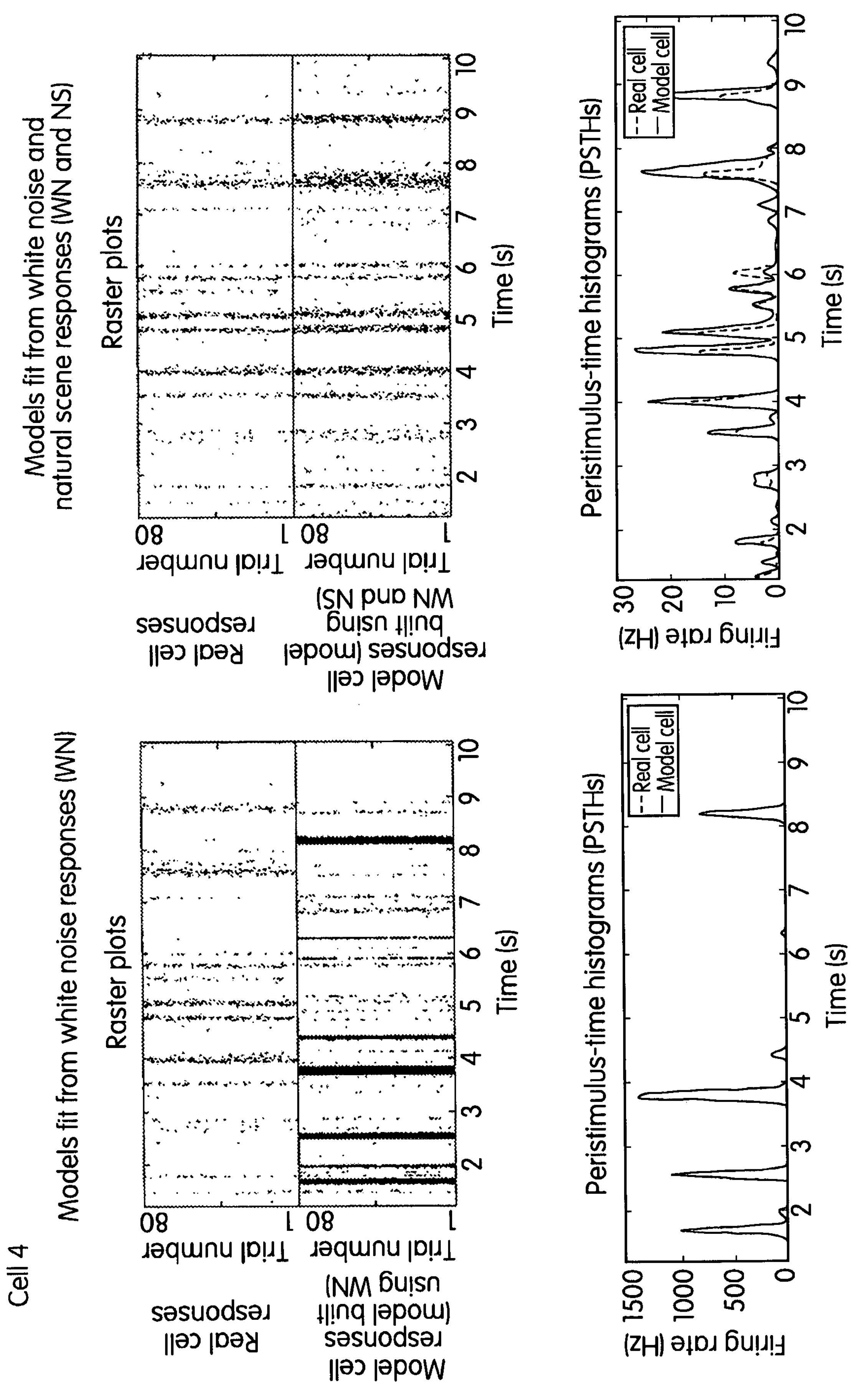

Cell 4
Models fit from white noise responses (WN)
Raster plots
Real cell responses
Model cell responses (model built using WN)
Trial number
80
1
Time (s)
Peristimulus-time histograms (PSTHs)
Firing rate (Hz)
- - Real cell
— Model cell
Models fit from white noise and natural scene responses (WN and NS)
Model cell responses (model built using WN and NS)

FIG. 18D

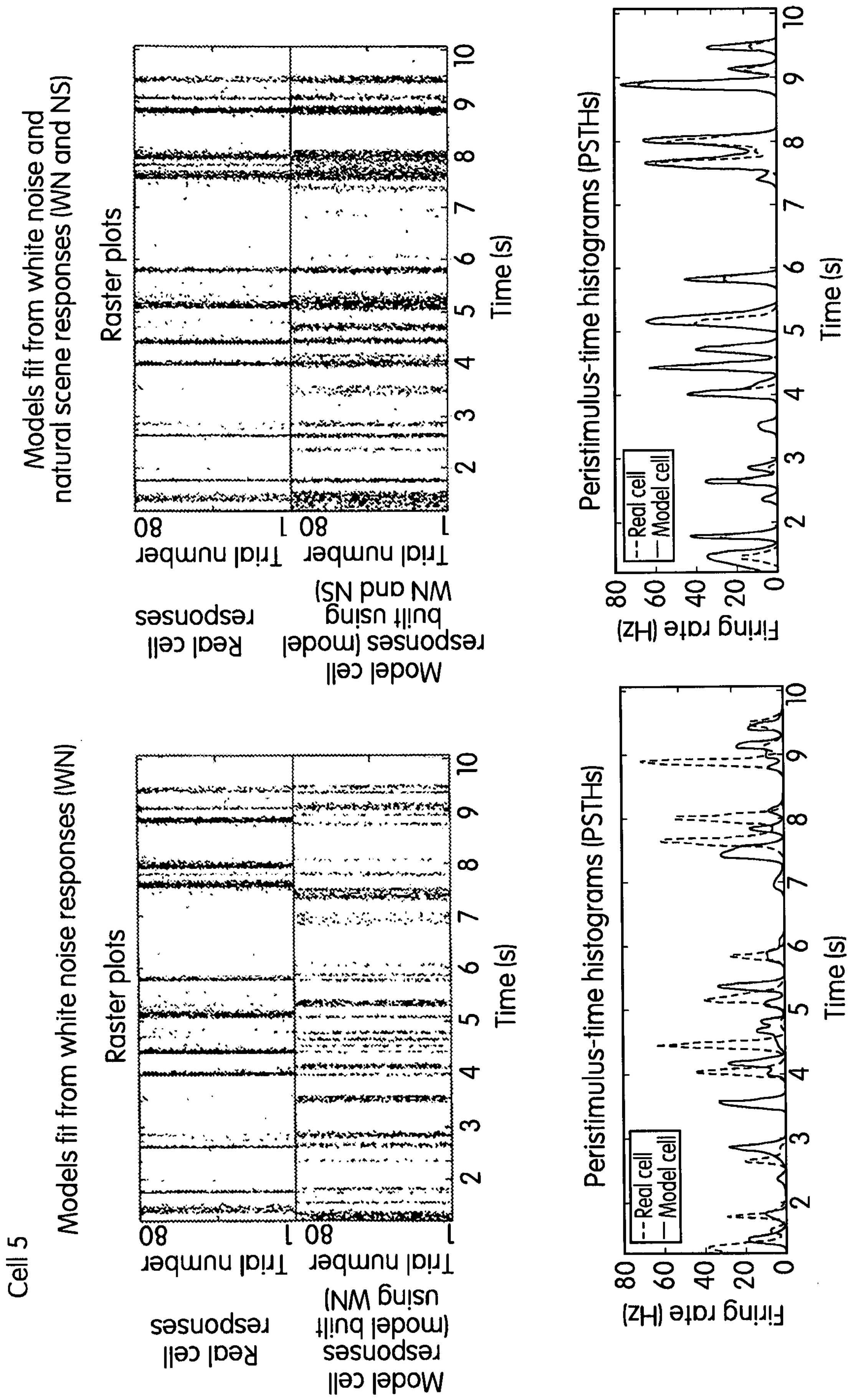

Cell 5
Models fit from white noise responses (WN)
Raster plots
Real cell responses
Model cell responses (model built using WN)
Trial number
80
1
Time (s)
2 3 4 5 6 7 8 9 10
Peristimulus-time histograms (PSTHs)
Firing rate (Hz)
0 20 40 60 80
Real cell
Model cell
Models fit from white noise and natural scene responses (WN and NS)
Model cell responses (model built using WN and NS)

FIG. 18E

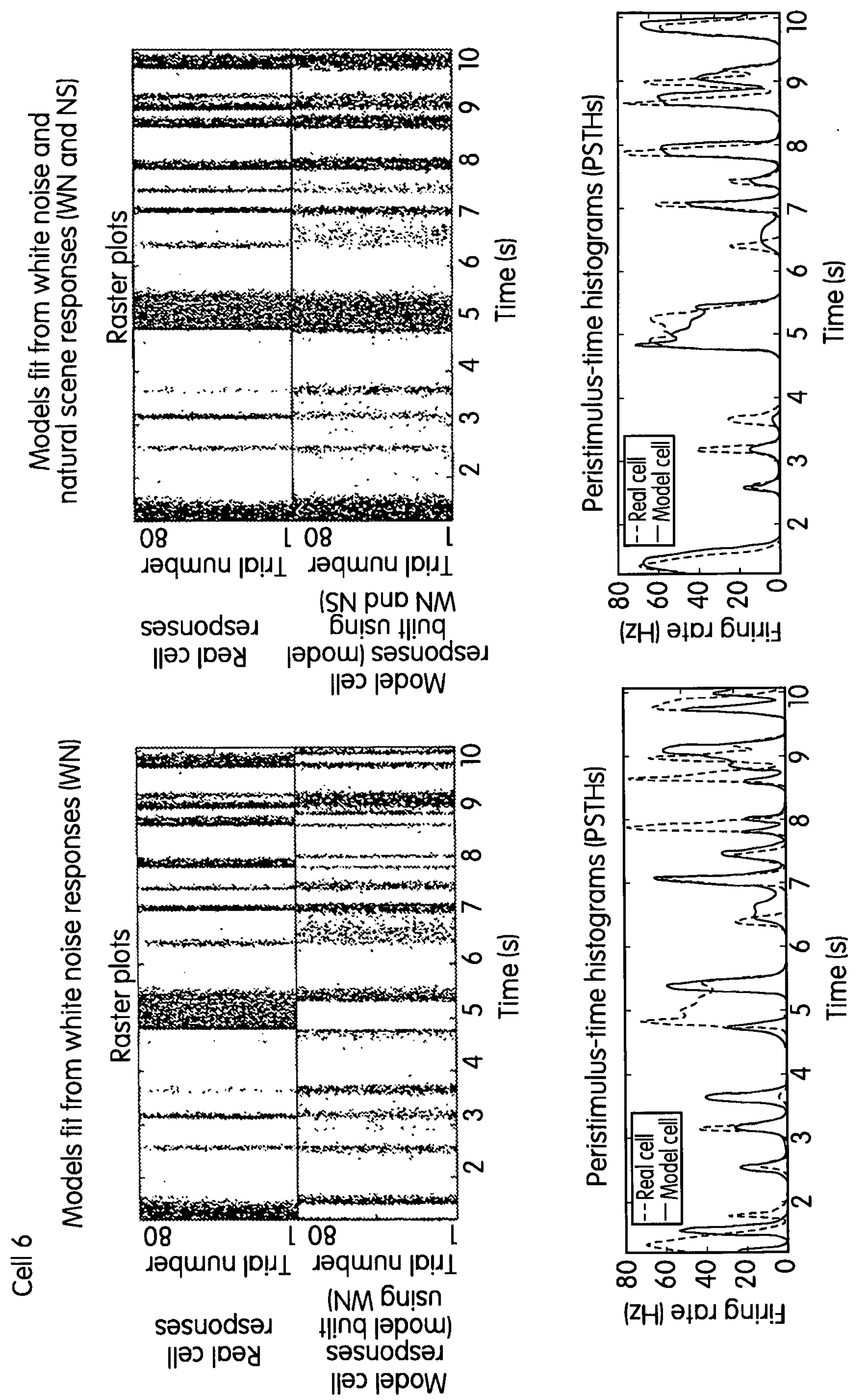

Cell 6
Models fit from white noise responses (WN)
Raster plots
Real cell responses
Model cell responses (model built using WN)
Trial number
Time (s)
Peristimulus-time histograms (PSTHs)
Real cell
Model cell
Firing rate (Hz)
Models fit from white noise and natural scene responses (WN and NS)
Model cell responses (model built using WN and NS)

FIG. 18F

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/52348

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06K 9/20 (2012.01)
USPC - 382/324
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)
USPC: 382/324

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 382/324; 382/117, 167; 348/370, 576 (keyword search - view terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Patbase; Google Scholar; Google Patents
Search Terms Used: Imagine retina cell pixel value intensity color machine vision encode mimic emulate simulate series sequence

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2011/0091102 A1 (CIMBALISTA JR) 21 April 2011 (21.04.2011) entire document, especially Abstract; para [0063], [0065]	1-5
Y	US 2006/0005160 A1 (SCHULTZ et al.) 05 January 2006 (05.01.2006) entire document, especially Abstract; para [0072]-[0075]	1-5
Y	US 6,801,655 B2 (WOODALL) 05 October 2004 (05.10.2004) entire document, especially Abstract; col 18, ln 56 to col 19, ln 15	5

☐ Further documents are listed in the continuation of Box C. ☐

* 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	Date of mailing of the international search report
11 October 2012 (11.10.2012)	02 NOV 2012

Name and mailing address of the ISA/US	Authorized officer:
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/52348

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 6-41
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

(19) 中华人民共和国国家知识产权局

(12) 发明专利申请

(10) 申请公布号 CN 103890781 A

(43) 申请公布日 2014.06.25

(21) 申请号 201280052177.5

(22) 申请日 2012.08.24

(30) 优先权数据

61/527,493 2011.08.25 US

61/657,406 2012.06.08 US

(85) PCT国际申请进入国家阶段日

2014.04.24

(86) PCT国际申请的申请数据

PCT/US2012/052348 2012.08.24

(87) PCT国际申请的公布数据

WO2013/029008 EN 2013.02.28

(51) Int.Cl.

G06K 9/20 (2006.01)

(71) 申请人 康奈尔大学

地址 美国纽约州

(72) 发明人 塞拉·尼伦伯格 伊利亚·鲍马斯

(74) 专利代理机构 北京万慧达知识产权代理有限公司 11111

代理人 戈晓美 杨颖

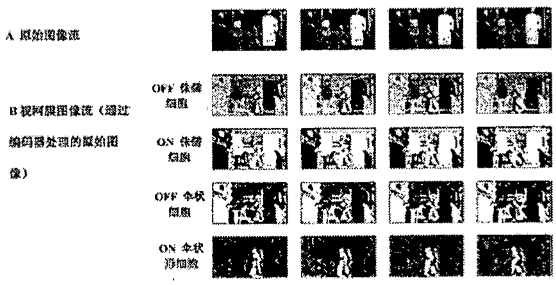
权利要求书3页 说明书35页 附图27页

(54) 发明名称

用于机器视觉的视网膜编码器

(57) 摘要

公开了一种方法，该方法包括：接收与一系列原始图像对应的原始图像数据；用编码器处理原始图像数据以生成编码数据，其中所述编码器的特征在于输入/输出转换，所述输入/输出转换本质上模拟脊椎动物视网膜的一个或多个视网膜细胞的输入/输出转换；以及将第一机器视觉算法用于至少一部分基于所述编码数据而产生的数据。



CN 103890781 A

1. 一种方法，该方法包括：

接收原始图像数据，所述原始图像数据与一系列原始图像对应；

用编码器处理所述原始图像数据，以生成编码数据，其中所述编码器的特征在于输入/输出转换，该输入/输出转换本质上模拟脊椎动物视网膜的一个或多个视网膜细胞的输入/输出转换；以及

将第一机器视觉算法用于至少一部分基于所述编码数据而产生的数据。

2. 根据前述任意一项权利要求所述的方法，其特征在于，所述方法还包括根据所述编码数据产生一系列视网膜图像。

3. 根据权利要求 2 所述的方法，其特征在于，该方法包括根据所述编码数据确定所述视网膜图像中的像素值。

4. 根据权利要求 3 所述的方法，其特征在于，其中，根据所述编码数据确定所述视网膜图像中的像素值包括：根据指示视网膜细胞响应的编码数据来确定像素强度或颜色。

5. 根据权利要求 4 所述的方法，其特征在于，其中，所述指示视网膜细胞响应的数据指示下述各项中的至少一项：视网膜细胞的放电频率、视网膜细胞输出脉冲串，以及发生器电势。

6. 根据权利要求 2-6 中任意一项所述的方法，其特征在于，该方法还包括：

将所述第一机器视觉算法用于所述一系列视网膜图像。

7. 根据权利要求 6 所述的方法，其特征在于，其中所述机器视觉算法包括下述各项中的至少一项：对象识别算法、图像分类算法、面部识别算法、光学字符识别算法、基于内容的图像检索算法、姿态估计算法、运动分析算法、自我运动判定算法、运动追踪算法、光流确定算法、场景重建算法、三维体积识别算法，以及导航算法。

8. 根据前述任意一项权利要求所述的方法，其特征在于，其中当用于所述一系列视网膜图像时，所述机器视觉算法表现出比用于相应的一组未经所述编码器处理的原始图像时更好的性能。

9. 根据权利要求 8 所述的方法，其特征在于，其中当用于一系列包括自然场景的视网膜图像时，所述机器视觉算法表现出比用于相应的一系列未经所述编码器处理的原始图像时更好的性能。

10. 根据权利要求 8 或 9 所述的方法，其特征在于，其中所述机器视觉算法包括在一系列图像中检测或识别人的算法；当用于各种包括人的一系列视网膜图像时，所述机器视觉算法表现出比用于相应的一组未经所述编码器处理的原始图像时更好的检测或识别准确度。

11. 根据权利要求 10 所述的方法，其特征在于，其中所述包括人的一系列图像包括位于自然场景中的人的图像。

12. 根据权利要求 11 所述的方法，其特征在于，其中所述包括人的一系列图像包括位于自然场景中的人的图像，所述位于自然场景中的人的图像不同于用于训练所述机器视觉算法的自然场景的图像。

13. 根据权利要求 8 或 9 所述的方法，其特征在于，其中所述机器视觉算法包括用于通过真实或虚拟的环境导航的算法，当用于一系列包括自然场景的视网膜图像时，该机器视觉算法表现出比用于对应的一组未经所述编码器处理的原始图像时更好的导航性能。

14. 根据权利要求 13 所述的方法，其特征在于，其中当用于一系列包括自然场景的视网膜图像时，所述机器视觉算法在导航中表现出比用于对应的一组未经所述编码器处理的原始图像时更少的不希望的碰撞事件。

15. 根据权利要求 14 所述的方法，其特征在于，其中所述一系列视网膜图像与未被用于训练所述机器视觉算法的环境对应。

16. 根据前述任意一项权利要求所述的方法，其特征在于，该方法还包括：

将机器成像算法用于所述一系列视网膜图像，以识别一个或多个目标视网膜图像；以及

识别与所述目标视网膜图像对应的一个或多个目标原始图像。

17. 根据权利要求 16 所述的方法，其特征在于，该方法还包括处理所述目标原始图像。

18. 根据权利要求 17 所述的方法，其特征在于，其中处理所述目标原始图像包括将第二机器视觉算法用于所述目标原始图像。

19. 根据权利要求 18 所述的方法，其特征在于，其中：

所述第一机器视觉算法包括已经在一组视网膜图像上训练过的算法；

所述第二机器视觉算法包括已经在一组原始图像上训练过的算法。

20. 根据前述任意一项权利要求所述的方法，其特征在于，其中应用所述第一机器视觉算法包括应用导航算法。

21. 根据权利要求 20 所述的方法，其特征在于，其中应用导航算法包括：

处理所述一系列视网膜图像，以确定运动信息，所述运动信息指示一系列图像中多个图像位置的运动；

根据所述运动信息对所述一系列图像的空间区域进行分类；以及

根据所述空间区域的分类生成导航决定。

22. 根据权利要求 21 所述的方法，其特征在于，其中运动信息指示所述一系列图像中的光流。

23. 根据权利要求 21 或 22 所述的方法，其特征在于，该方法包括：

使用卷积神经网络对所述空间区域进行分类。

24. 根据权利要求 21-23 中任意一项所述的方法，其特征在于，该方法还包括根据导航算法得到的结果来控制机器人装置的运动。

25. 根据权利要求 18-24 中任意一项所述的方法，其特征在于，该方法还包括根据导航算法得到的结果来控制虚拟空间中虚拟对象的运动。

26. 根据权利要求 24 或 25 所述的方法，其特征在于，其中所述导航算法是用根据代表虚拟空间的图像数据训练过的算法。

27. 根据前述任意一项权利要求所述的方法，其特征在于，该方法还包括：根据所述视网膜图像训练机器视觉算法。

28. 根据权利要求 27 所述的方法，其特征在于，其中训练机器视觉算法包括：

(i) 将所述机器视觉算法用于一组视网膜图像，以产生输出；

(ii) 根据所述输出，确定指示所述机器视觉算法性能的性能信息；

(iii) 根据所述性能信息，调整所述机器视觉算法的一个或多个特性。

29. 根据权利要求 28 所述的方法，其特征在于，该方法还包括：

迭代地重复步骤 (i) 至 (iii)，直至达到选定的性能标准为止。

30. 根据权利要求 27-29 中任意一项所述的方法，其特征在于，其中所述经过训练的机器视觉算法的特征在于一组参数，并且其中所述参数不同于通过用与所述视网膜图像对应的原始图像对机器视觉算法进行同等训练所获得的对应参数。

31. 根据权利要求 6-30 中任意一项所述的方法，其特征在于，其中：

用编码器处理所述原始图像数据以生成编码数据包括生成编码数据，相对于相应的原始图像数据，该编码数据包含的信息量减少；

其中当用于一系列视网膜图像时，所述机器视觉算法表现出比用于相应的一组未经所述编码器处理的原始图像时更好的性能。

32. 根据权利要求 31 所述的方法，其特征在于，其中所述编码数据所包含的信息量相对于相应的原始图像数据被压缩至少约 2 倍。

33. 根据权利要求 31 所述的方法，其特征在于，其中所述编码数据所包含的信息量相对于相应的原始图像数据被压缩至少约 5 倍。

34. 根据权利要求 31 所述的方法，其特征在于，其中所述编码数据所包含的信息量相对于相应的原始图像数据被压缩至少约 10 倍。

35. 根据前述任意一项权利要求所述的方法，其特征在于，其中所述脊椎动物包括下述各项中的至少一项：鼠和猴。

36. 根据前述任意一项权利要求所述的方法，其特征在于，其中所述视网膜细胞包括神经节细胞。

37. 根据前述任意一项权利要求所述的方法，其特征在于，其中所述视网膜细胞包括至少两种类型的细胞。

38. 根据前述任意一项权利要求所述的方法，其特征在于，其中所述至少两种类型的细胞包括 ON 细胞和 OFF 细胞。

39. 根据前述任意一项权利要求所述的方法，其中所述编码器的特征在于输入／输出转换，该输入／输出转换本质上模拟脊椎动物视网膜的一种或多种视网膜细胞的输入／输出转换，所述脊椎动物视网膜覆盖一定范围的输入，该输入包括自然场景图像，该自然场景图像包括随时空变化的图像。

40. 一种装置，其特征在于，该装置包括：

至少一个存储装置，该存储装置被配置为存储原始图像数据；以及

至少一个处理器，该处理器可操作地与所述存储器耦合，并且该处理器被编程以执行如权利要求 1-38 中任意一项所述的方法。

41. 非临时性计算机可读介质，其特征在于，该介质具有计算机可执行的指令，用于执行如权利要求 1-38 中任意一项所述方法的步骤。

用于机器视觉的视网膜编码器

[0001] 相关申请的交叉引用

[0002] 本申请要求序列号为61/527,493(2011年4月25日申请)和61/657,406(2012年6月8日申请)的美国临时申请的优先权。前述各申请的内容通过引用而整体并入本文。

[0003] 本申请还与序列号为61/308,681(2010年2月26日申请)、61/359,188(2010年6月28日申请)、61/378,793(2010年8月31日申请)、61/382,280(2010年9月13日申请)的美国临时申请;序列号为13/230,488(2011年9月12日申请)的美国专利申请;以及序列号为PCT/US2011/026526(2011年2月28日申请)和PCT/US2011/049188(2011年8月25日申请)的国际专利申请有关。前述各申请的内容通过引用而整体并入本文。

[0004] 关于联邦政府资助研究或开发的声明

[0005] 本发明在美国政府的支持下进行,其获得了由美国国立卫生研究院(NIH)的美国国家眼睛研究所授予的R01EY12978号基金的资助。美国政府对本发明享有一定的权利。

技术领域

[0006] 本发明涉及用于机器视觉的方法和设备。本发明特别涉及通过模拟动物视网膜性能的编码器处理图像,以及将所述处理的图像应用于机器视觉中的方法和设备。

背景技术

[0007] 机器视觉(或计算机视觉)是指允许计算机使用可视化信息的技术,例如,从图像中提取信息,解决某些任务,或者是在广义或狭义上"理解"场景。通常,机器视觉关注的是从图像数据中提取信息。所述图像数据可以是多种形式,例如单个图像、视频序列、来自多个摄像机的视图,或更高维度的数据(例如,由医学扫描仪得到的三维图像)。

[0008] 机器视觉有许多应用,可以是相对简单的任务,例如用于对生产线上经过的物体进行计数的工业系统,也可以是更复杂的任务,例如面部识别,以及感知任务(例如,允许机器人在复杂的环境中导航)。机器视觉的应用的非限制性例子包括用于控制过程(例如,工业机器人或无人驾驶车辆)、检测事件(例如,用于视觉监控或人数统计)、组织信息(例如,索引图像和图像序列数据库)、为物体或环境建模(例如,工业检查、医学图像分析或地形建模),以及交互(例如,作为用于人机交互装置的输入)的系统。

[0009] 在许多应用中,机器视觉涉及高运算量的昂贵任务。单色数字图像可以由数百万或更多的像素组成,每个像素具有准值,例如定义颜色空间(例如,熟悉的RGB颜色空间、YCbCr空间、HSV空间,等)中像素的坐标的多个(例如,8或24)比特值。视频流可包括这些图像的序列,所述图像序列的帧频是,例如,每秒数十帧,相当于每秒数百兆比特的比特率或更多。许多机器视觉应用需要对这类图像或视频流进行快速处理(例如,追踪和应对物体的运动、识别或分类沿装配流水线移动的物体,以使机器人能够对它的环境作出实时反应,等)。

[0010] 在这样的时间限制内处理这么大的数据量是非常具有挑战性的。因此,人们希望找到用于处理图像数据的技术,所述技术可以减少信息的原始数量,同时保持(甚至提高)

对于手头的机器视觉任务而言最突出的图像数据的性质。所述预处理的图像数据，而不是原始数据，随后可被输入到机器视觉系统，同时能降低系统的处理负担，并允许足够迅速的反应和潜在的改进性能。

[0011] 已认识到脊椎动物眼睛的视网膜提供的就是这种性质的图像处理，接受视觉刺激，并将所述刺激转换成大脑可以理解的形式。所述系统（进化了数百万年的进化过程）是非常高效和有效的，正如哺乳动物高水平的复合视觉感知（尤其是猴子和人类）所证明的。

[0012] 已提出了多种方法来开发图像数据预处理方案，所述方案用于基于视网膜运作的抽象模型的机器视觉。然而，这些模型都基于大致近似于视网膜的实际表现。

[0013] 背景技术部分的一部分改编自维基百科上提供的关于机器视觉的文章，该文章可从 http://en.wikipedia.org/wiki/Computer_vision 获得，并根据知识共享署名－相同方式共享许可来使用。

发明内容

[0014] 本发明描述的实施方式使用了能近乎完整地复制视网膜执行的操作的编码器。如上文通过引用并入的国际专利申请（文下简称“假体应用”）中所详细描述的，所述编码器可用于开发高效视网膜假体。在本发明中，所述编码器被应用于机器视觉。

[0015] 当被用作预处理步骤（具体而言，降维步骤），所述编码器本质上提高了机器视觉算法的性能。在某些实施方式中，编码器允许机器视觉算法在范围广阔的环境和照明条件中非常有效地提取信息，包括无法通过其他方法来提取的信息。在现有的机器视觉算法部分有效的情况下，这种降维可具有强有力的增强作用。所述编码器能够更有效地，以及更快、更高效地进行提取（性能更高）。

[0016] 如在假体应用中所详细描述的，申请人已开发了一种假体装置，所述假体装置能接收刺激，并通过一组编码器将所述刺激转换成一组代码，通过接口将所述代码转换为信号，随后通过高分辨率传感器激活多个视网膜细胞，所述高分辨率传感器由来自所述接口的信号驱动。激活多个视网膜细胞使得视网膜神经节细胞对宽范围的刺激产生应答，其与来自正常视网膜的视网膜神经节细胞对相同刺激产生的时间依赖性应答本质上类似。申请人已认识到，在这类装置中使用的编码器可以适于处理图像数据，以用于机器视觉的应用。

[0017] 在假体应用中所描述的视网膜假体像正常视网膜那样也是图像处理器－它从所接收到的刺激中提取必要的信息，并将信息重新格式化为大脑可以理解的动作电位的形式。由正常视网膜产生的动作电位的模式被称为视网膜代码或神经节细胞代码。视网膜假体将视觉刺激转换成相同的代码，或相近的代替模式，从而使被损坏或退化的视网膜可以产生正常或接近正常的输出。因为视网膜假体使用与正常视网膜相同的代码或其相近的代替模式，被损伤或退化的视网膜的神经节细胞的放电模式，即它们的动作电位与那些通过正常神经节细胞所产生的模式是相同的或本质上类似的。因此，所述假体允许视网膜将与正常视网膜相同的关于视觉世界的信号传送到大脑。

[0018] 如在假体应用中所详细描述，编码器使用用于视网膜细胞的输入／输出模式，该模式通过来自真实视网膜细胞对各种刺激，例如，白噪声（WN）和自然场景（NS）影片的输入／输出的响应数据而产生。在某些实施方式中，编码器基于线性非线性级联模式，该模式包

括特征在于具有若干参数的时空转换。这些参数根据从真实视网膜实验获得的数据进行优化，从而产生能近似模拟真实细胞对较宽范围的刺激进行响应的转换。结果得到能捕捉自然图像（静态或随时空改变）的输入／输出关系的模式，如人脸、风景、正在走路的人、孩子在玩耍等，而不只是针对白噪声刺激或具有高斯统计的刺激。在假体应用中，以及在下文将具体讨论的图 18A-18F 中显示了对较宽范围刺激的有效性。

[0019] 由于这种方法利用通过实验获得的数据，生成的编码器可以准确地模拟视网膜的处理，而不需要详细的对视网膜潜在处理方案的抽象理解。例如，可以认为，灵长类和人类视网膜的处理突出的是视觉刺激中对模式识别的任务（例如，面部识别）有用的特征，而不强调或消除其它特征（例如，冗余信息或噪声），以便大脑可以高效地进行处理。然而，尚未对作为亿万年自然选择过程结果的所述处理方案细节有完整的抽象理解。然而，尽管缺乏所述抽象理解，通过准确地模拟视网膜响应，本文所述的装置和技术可以获得所述处理的优点。

[0020] 换句话说，在本文所述不同的实施方式中，方法是数据驱动的 - 也就是说，它采用视网膜输入／输出关系的数据驱动模式，从而能提供逼真的图像预处理。这使下游机器视觉算法具有能完成与生物视网膜相同种类和相同数量级降维的预处理步骤，并因此提供与生物视网膜相同的一系列优势。

[0021] 需要注意的是，在一般情况下，先前的预处理器使用，例如高斯差型过滤器，来过滤图像数据，本文所描述的方法与先前的预处理器不同，因为本文的预处理器可以对视网膜进行完整或接近完整的模拟。同样的，它不同于其他线性非线性级联模式，差别在于它对较宽范围的刺激有效，而不只是对白噪声刺激或具有高斯统计的刺激有效。因此，过滤更彻底，而且大大提高了现有机器视觉算法的能力。最重要的是，它可以让现有的机器视觉算法来概括，即，在一项设置中（一种环境或照明条件）被培训，并能推广到其他环境中，而这一直以来是长期存在的挑战（参见例如，下文中具体描述的图 10、图 11 和图 15）。

[0022] 此外，在一些实施方式中，由于视网膜的处理为较宽范围的刺激（例如，同时使用 WN 产生的数据和 NS 产生的数据优化得到的结果）准确地建模，机器视觉系统的预处理能在很宽范围的条件（类似于视网膜在很宽范围的条件下工作的方式）下良好运行。有利的是，这使得视网膜预处理技术能被用于根据各种条件（例如，光照变化、复杂的、变化的视觉场景、许多不同的环境等）需要强健性能的机器视觉应用中。

[0023] 一方面，公开了一种方法，所述方法包括：接收原始图像数据，所述原始图像数据与一系列原始图像对应；用编码器处理所述原始图像数据，以生成编码数据，其中所述编码器的特征在于输入／输出转换，该输入／输出转换本质上模拟脊椎动物视网膜的一个或多个视网膜细胞的输入／输出转换；以及将第一机器视觉算法用于至少部分基于所述编码数据而产生的数据。

[0024] 一些实施方式包括根据编码数据产生一系列视网膜图像。一些实施方式包括根据编码数据确定所述视网膜图像中的像素值。在一些实施方式中，根据编码数据确定所述视网膜图像中的像素值的步骤包括，根据指示视网膜细胞响应的编码数据来确定像素强度或颜色。

[0025] 在一些实施方式中，指示视网膜细胞响应的数据指示下述各项中的至少一项：视网膜细胞的放电频率、视网膜细胞输出脉冲串，以及发生器电势。

[0026]　一些实施方式包括将第一机器视觉算法用于所述一系列视网膜图像。

[0027]　在一些实施方式中，机器视觉算法包括下述各项中的至少一项：对象识别算法、图像分类算法、面部识别算法、光学字符识别算法、基于内容的图像检索算法、姿态估计算法、运动分析算法、自我运动判定算法、运动追踪算法、光流确定算法、场景重建算法、三维体积识别算法，以及导航算法。

[0028]　在一些实施方式中，当用于一系列视网膜图像时，所述机器视觉算法表现出比用于相应的一组未经编码器处理的原始图像时更好的性能。

[0029]　在一些实施方式中，当用于一系列包括自然场景的视网膜图像时，所述机器视觉算法表现出比用于相应的一系列未经所述编码器处理的原始图像时更好的性能。

[0030]　在一些实施方式中，所述机器视觉算法包括在一系列图像中检测或识别人的算法；并且其中当用于各种包括人在内的一系列视网膜图像时，所述机器视觉算法表现出比用于相应的一组未经所述编码器处理的原始图像时更好的检测或识别准确度。

[0031]　在一些实施方式中，所述包括人的一系列图像包括位于自然场景中的人的图像。

[0032]　在一些实施方式中，所述包括人的一系列图像包括位于自然场景中的人的图像，所述位于自然场景中的人的图像不同于用于训练机器视觉算法的自然场景的图像。

[0033]　在一些实施方式中，所述机器视觉算法包括通过真实或虚拟的环境导航的算法，并且其中当用于一系列包括自然场景的视网膜图像时，该机器视觉算法表现出比用于对应的一组未经所述编码器处理的原始图像时更好的导航性能。

[0034]　在一些实施方式中，当用于一系列包括自然场景的视网膜图像时，所述机器视觉算法在导航中表现出比用于对应的一组未经所述编码器处理的原始图像时更少的不希望的碰撞事件。

[0035]　在一些实施方式中，所述一系列视网膜图像与未被用于训练所述机器视觉算法的环境对应。

[0036]　一些实施方式包括将机器成像算法用于一系列视网膜图像，以识别一个或多个目标视网膜图像；并且识别与所述目标视网膜图像对应的一个或多个目标原始图像。一些实施方式包括处理所述目标原始图像。在一些实施方式中，处理所述目标原始图像包括将第二机器视觉算法用于所述目标原始图像。在一些实施方式中，所述第一机器视觉算法包括已经在一组视网膜图像上训练过的算法；并且所述第二机器视觉算法包括已经在一组原始图像上训练过的算法。

[0037]　在一些实施方式中，应用所述第一机器视觉算法包括应用导航算法。在一些实施方式中，应用所述导航算法包括：处理所述一系列视网膜图像，以确定指示所述一系列图像中多个图像位置的运动的运动信息；根据所述运动信息对所述一系列图像的空间区域进行分类；以及根据所述空间区域的分类生成导航决定。在一些实施方式中，所述运动信息指示所述一系列图像中的光流。一些实施方式包括使用卷积神经网络对所述空间区域进行分类。

[0038]　一些实施方式包括根据导航算法得到的结果来控制机器人装置的运动。

[0039]　一些实施方式包括根据导航算法得到的结果来控制虚拟空间中虚拟对象的运动。

[0040]　一些实施方式包括根据所述视网膜图像训练机器视觉算法。在一些实施方式中，训练所述机器视觉算法包括：(i) 将所述机器视觉算法用于一组视网膜图像，以产生输出；

(ii) 根据所述输出，确定指示所述机器视觉算法性能的性能信息；以及 (iii) 根据所述性能信息，调整所述机器视觉算法的一个或多个特性。一些实施方式包括迭代地重复步骤 (i) 至 (iii)，直至达到选定的性能标准为止。

[0041] 在一些实施方式中，所述经过训练的机器视觉算法的特征在于一组参数，并且其中所述参数不同于通过用与所述视网膜图像对应的原始图像对机器视觉算法进行同等训练所获得的对应参数。

[0042] 在一些实施方式中，用编码器处理所述原始图像数据以生成编码数据，包括生成编码数据，相对于相应的原始图像数据，该编码数据包含的信息量减少；在一些实施方式中，当用于一系列视网膜图像时，所述机器视觉算法表现出比用于相应的一组未经所述编码器处理的原始图像时更好的性能。

[0043] 在一些实施方式中，所述编码数据所包含的信息量相对于相应的原始图像数据被压缩至少约 1.5、2、3、4、5、6、7、8、9、10，或更多倍，例如在 1.1-1,000 或其任何子范围的范围内。

[0044] 在一些实施方式中，脊椎动物包括下述各项中的至少一项：鼠和猴。

[0045] 在一些实施方式中，所述视网膜细胞包括神经节细胞。在一些实施方式中，所述视网膜细胞包括至少两种类型的细胞。在一些实施方式中，所述至少两种类型的细胞包括 ON 细胞和 OFF 细胞。

[0046] 在一些实施方式中，所述编码器的特征在于输入 / 输出转换，该输入 / 输出转换本质上模拟脊椎动物视网膜的一种或多种视网膜细胞的输入 / 输出转换，所述脊椎动物视网膜覆盖一定范围的输入，该输入包括自然场景图像，该自然场景图像包括随时空变化的图像。

[0047] 在一些实施方式，用编码器处理所述原始图像数据以生成编码数据的步骤包括：处理所述原始图像数据以生成多个值 X，将所述多个 X 值转换成多个响应值 λ_m，λ_m 指示视网膜中视网膜细胞 m 的相应响应，并根据所述响应值生成编码数据。在一些实施方式中，所述响应值与视网膜细胞放电频率对应。在一些实施方式中，所述响应值与所述视网膜细胞放电频率的功能对应。在一些实施方式中，所述响应值与视网膜细胞输出脉冲对应。在一些实施方式中，所述响应值与视网膜细胞发生器电势对应，即，用时空滤波器卷积所述图像的输出。

[0048] 在一些实施方式中，用编码器处理原始图像数据以生成编码数据包括：从所述原始图像数据接收图像，对于每个图像，重新调节亮度或对比度以生成重缩放的图像流；从所述重缩放的图像流接收一组 N 个重缩放图像，并将时空转换用于所述一组 N 个图像，以生成一组视网膜响应值，所述一组值中的每个值与所述视网膜细胞中的每一个对应；根据视网膜响应值生成所述编码数据。

[0049] 在一些实施方式中，所述响应值包括视网膜细胞放电频率。在一些实施方式中，N 是至少 5、至少约 20、至少约 100 或更多，例如，在 1-1,000 的范围或其任何子范围内。

[0050] 在一些实施方式中，应用时空转换包括：用时空核函数卷积所述 N 个重缩放图像，以生成一个或多个空间 - 时间上转换的图像；并且将非线性函数用于所述空间 - 时间上转换的图像，以生成一组响应值。

[0051] 在一些实施方式，应用时空变换包括：用空间核函数卷积所述 N 个重缩放图像，以

生成N个空间转换的图像；用时间核函数卷积N个空间转换的图像以生成时间转换的输出；并且将非线性函数用于所述时间转换的输出以生成所述一组响应值。

[0052] 在一些实施方式中，所述编码器的特征在于一组参数，并且其中当脊椎动物视网膜暴露于白噪声和自然场景刺激时，使用从所述视网膜通过实验获得的响应数据来确定所述参数的值。

[0053] 在一些实施方式中，设置所述编码器，从而使得测试输入刺激和从所述编码数据重构的相应刺激之间的皮尔森相关系数为至少约 0.35、0.65，至少约 0.95，或更多，例如，在 0.35-1.0 的范围或其任何子范围内，所述编码数据在响应所述测试输入刺激时由编码器产生。在一些实施方式中，所述测试输入刺激包括一系列自然场景。

[0054] 另一方面，公开了一种装置，所述装置包括：至少一个存储装置，该存储装置被配置为存储原始图像数据；至少一个处理器，该处理器可操作地与所述存储器耦合，并且该处理器被编程以执行如本文所述的一种或多种方法。

[0055] 在一些实施方式中，公开了具有计算机可执行的指令、用于执行如本文所述的一种或多种方法的步骤的非临时性计算机可读介质。

[0056] 另一方面，公开了一种系统，所述系统包括：至少一个存储装置，存储与一系列图像对应的编码数据，其中所述编码数据的生成是通过：接收与一系列原始图像对应的原始图像数据，并且用编码器处理所述原始图像数据以生成编码数据，其中所述编码器的特征在于输入 / 输出转换，该输入 / 输出转换本质上模拟脊椎动物视网膜的一个或多个视网膜细胞的输入 / 输出转换。在一些实施方式中，所述至少一个存储装置存储指示所述编码数据和所述原始图像数据之间对应关系的数据库信息。

[0057] 一些实施方式包括处理器，所述处理器被配置成：接收与一系列查询图像对应的查询图像数据；用编码器处理所述查询图像数据，以生成编码数据，其中所述编码器的特征在于输入 / 输出转换，所述输入 / 输出转换本质上模拟脊椎动物视网膜的一个或多个视网膜细胞的输入 / 输出转换；比较所述编码查询图像数据与存储装置上的编码数据；并且基于 (a) 所述编码查询数据与所述存储装置上的编码数据的比较，以及 (b) 指示所述编码数据和所述原始图像数据之间对应关系的数据库信息，确定所述查询图像数据和所述原始图像数据之间的对应关系。

[0058] 另一方面，公开了一种方法，所述方法包括：接收与一系列原始图像对应的原始图像数据；用编码器处理所述原始图像数据的至少第一部分，以生成第一编码数据，其中所述编码器的特征在于输入 / 输出转换，该输入 / 输出转换本质上模拟第一种脊椎动物类型的第一脊椎动物视网膜的一个或多个视网膜细胞的输入 / 输出转换；以及用编码器处理所述原始图像数据的至少第二部分，以生成编码数据，其中所述编码器的特征在于输入 / 输出转换，该输入 / 输出转换本质上模拟与第一种脊椎动物类型不同的第二种脊椎动物类型的第二脊椎动物视网膜的一个或多个视网膜细胞的输入 / 输出转换。

[0059] 一些实施方式包括，根据所述第一编码数据，选择所述原始图像数据的第二部分用于处理。

[0060] 在各种实施方式中，所述原始图像数据本质上从图像检测器，或从存储所述原始图像数据的存储器，或从所述图像检测器和存储器的组合中实时接收。

[0061] 另一方面，公开了一种装置，所述装置包括：至少一个存储装置，该存储装置被配

置为存储原始图像数据；至少一个处理器，该处理器可操作地与所述存储器耦合，并且该处理器被编程以执行如本文所述的一种或多种方法。

[0062] 另一方面，公开了具有计算机可执行的指令、用于执行如本文所述的一种或多种方法的步骤的非临时性计算机可读介质。

[0063] 另一方面，公开了一种系统，所述系统包括：至少一个存储装置，存储与一系列图像对应的编码数据，其中所述编码数据的生成是通过：接收与一系列原始图像对应的原始图像数据；并且用编码器处理所述原始图像数据，以生成编码数据，其中所述编码器的特征在于输入／输出转换，该输入／输出转换本质上模拟脊椎动物视网膜的一个或多个视网膜细胞的输入／输出转换。在一些实施方式中，所述至少一个存储装置存储指示所述编码数据和所述原始图像数据之间对应关系的数据库信息。

[0064] 各种实施方式可以单独地或以任何合适的组合方式包括任何上文所述的元素。

附图说明

[0065] 图 1 是显示了示例性机器视觉系统的框图。

[0066] 图 2 是举例说明了编码器模块运行的流程图。

[0067] 图 3A 举例说明了原始图像流（人穿过复杂的环境）转换成视网膜图像流。板 A 显示了由摄像头获得的原始图像流的若干帧。板 B 显示了相应的视网膜图像流的若干帧。显示了四个不同的视网膜图像流，每个视网膜图像流使用不同的阵列单元（OFF 侏儒细胞、ON 侏儒细胞、OFF 伞状细胞（parasol cells），和 ON 伞状细胞，如图中所示）。

[0068] 图 3B-3F 显示了原始图像（图 3B）和视网膜图像的放大图。图 3C-3F 与图 3A 最后一栏对应。

[0069] 图 4 是显示了用于训练图 1 所示机器视觉系统的机器视觉模块的训练系统的框图。

[0070] 图 5 是举例说明图 4 的训练系统的操作的流程图。

[0071] 图 6 举例说明了用于控制机器人通过迷宫的导航的机器视觉系统。机器人行进的路径以虚线表示。

[0072] 图 7 是用于控制导航任务的机器视觉系统的一个实施方式的流程图。

[0073] 图 8 显示了用于训练导航仪的原始图像流（影片）的若干帧。所述图像流是在使用如正文中所述的田园环境的虚拟环境中生成。顶部面板显示了图像流中的前 5 帧。底部面板显示了从图像流剩余部分选择的帧；显示了每 30 帧中的一个（即，每秒 1 帧）。

[0074] 图 9 显示了用于测试所述导航仪的原始图像流（影片）的若干帧。显示了下述三组：A，田园环境（与用于训练导航仪的环境不同）的若干帧；B，郊区环境；以及 C，广场环境（轮胎障碍训练场）。如图 9 所示，在虚拟环境中产生的图像流，各组的顶部面板显示了前四帧，并且底部面板显示了从影片其余部分选择的若干帧（在这种情况下，每 15 帧中的一个（即，每半秒 1 帧）。

[0075] 图 10 举例说明了显示导航仪性能的轨迹，及其推广到不同环境的能力。如本文中以及图 7 的流程图所描述，用于学习导航任务的主要算法、以两种方法训练卷积神经网络（CNN）：1) 标准方法，即，使用原始视觉环境（原始图像流），以及 2) 使用其降维后的环境，即在它被编码器处理后。（使用的训练环境是田园环境，如图 8 所示）。随后在 3 个新的环

境中测试所述导航仪的性能：与用于训练所述导航仪不同的田园环境、郊区环境和广场环境（每个环境的样品如图 9 所示）A. 当导航仪从原始图像流获知环境时它的性能。值得注意的是杂乱无章的轨迹和碰撞。B. 当导航仪从视网膜图像（编码器所产生的图像流）流获知环境时它的性能。值得注意的是直线路径和避开了障碍。

[0076]　图 11 显示导航仪高性能的进一步证明；具体而言，它显示了高性能不仅推广到不同的环境（从田园环境到郊区环境到广场），而且它也可以推广到环境中不同的照明条件。A 到 F 与太阳的不同位置对应，因此对应广场环境中不同的阴影条件；跨越日出到日落的光线条件，即，环境左侧地平线上 30 度至右侧地平线上 30 度。浅灰色，当用原始图像流训练导航仪时导航仪的性能（使用一个照明条件的田园环境，如图 8 所示）。如本图所示，导航仪被放置在新环境中时，导航仪的性能较低，并且该结论在各种光照条件下仍然保持正确。每条柱的高度与试验分数对应，在所述试验中导航仪成功留在轮胎路线中并不与轮胎中的任何一个碰撞。误差线表示平均数标准误差 (SEM)。深灰色，当用视网膜图像流训练时的导航仪性能（使用同一个光照条件的相同的田园环境，但这次通过编码器处理）。如图所示，导航仪的性能较高，并在各种光照条件下保持这种高性能。因此，用视网膜图像流（即，用编码器产生的降维图像）训练导致高性能，这能同时推广到新环境以及多种照明条件（日出到日落，见上文）。

[0077]　图 12 是用于控制面部识别任务的机器视觉系统的一个实施方式的流程图。

[0078]　图 13 显示的是用于训练面部识别算法（如本文所述的维奥拉 - 琼斯 - 斯诺

[0079]　(Viola-Jones-Snow) 算法）的原始图像流（影片）的若干帧。以每秒 24 帧的速率记录所述图像流；本图中，显示了每 12 帧（每半秒 1 帧）。

[0080]　图 14 显示了用于测试面部识别算法性能的原始图像流（影片）的若干帧。值得注意的是，这与图 13 中的是同一个人，但是在不同的环境中并具有不同的发型，等。如本文所述，面部识别算法的目的是要识别到属于目标人物的新的图像流，即使该算法是用此人的其他图像流进行训练）。如图 13 所示，以每秒 24 帧的速率记录所述图像流；本图中，显示了每 12 帧（每半秒 1 帧）。

[0081]　图 15 显示了，当以下述两种方法训练时，所述面部识别算法的性能：1) 使用标准方法，即，用原始图像数据流进行训练，以及 2) 用本申请中所描述的方法（即，使用编码器处理过的原始图像流）。在这两种情况下，用许多图像流进行面部识别算法训练（来自目标面部的 4-5 个视频的 250-800 两帧图像流和来自其他面部的 >100 个视频的 2000 个两帧图像流）。随后用来自之前没见过的视频，也就是没有在训练组中使用的视频，中的 50-800 个两帧图像流测定性能（来自训练和测试组的样本帧，参见图 13 和图 14)。显示了两组任务的性能，其中一组中标准方法表现非常弱，另一组中表现较好。柱的高度表示试验分数，在所述试验中面部识别仪成功识别了目标面部。误差线表示平均数标准误差 (SEM)。如图所示，当任务具有挑战性时 (A)，本申请所描述的方法比标准方法改进巨大（4 倍）。当任务是不那么具有挑战性时，即当标准方法表现较好时，本申请所描述的方法仍然有改进（1.5 倍）。

[0082]　图 16 显示了将视网膜编码器方法和传统方法用于图像处理时的示例性混合图像处理方法的处理流程。

[0083]　图 17 是使用视网膜编码数据的数字指纹的系统的框图。

[0084] 图 18A-18F 举例说明了用自然场景影片测试时视网膜编码器模型的性能。在各图中，左侧显示的是常规的线性 - 非线性 (LN) 模型的性能，并且右侧显示了本申请中所描述的类型的线性 - 非线性 (LN) 模型的性能。通过光栅图和直方图 (PSTHs) 显示性能。

具体实施方式

[0085] 图 1 显示了示例性的机器视觉系统 100，所述系统配有摄像头 102、编码器模块 104、机器视觉模块 106，以及由所述机器视觉模块控制的系统 108。摄像头 102 接收视觉刺激并将视觉刺激转换为数字图像数据，例如数字图像流。所述数字图像数据在本文中可被称为“原始”图像数据。但是应当理解的是，所述原始图像数据可包括在用视网膜编码器处理之前的任何图像数据。

[0086] 编码器模块 104 接收图像数据，并使用一种或多种本文所述类型的视网膜编码器和 / 或假体应用处理所述数据。被称为“视网膜图像数据”的编码器模块输出，被传送到处理所述视网膜图像数据的机器视觉模块，该机器视觉模块例如通过一项或多项本领域已知和 / 或本文所描述的机器视觉技术处理视网膜图像数据。根据所述机器视觉处理，机器视觉模块 106 产生输出，所述输出可被用于任何合适的目的。如图所示，所述输出控制一个或多个系统 108，例如，机器人系统。在一些实施方式中，可实时或接近实时地进行图像处理和 / 或控制。

[0087] 应当理解的是，图 1 所示的系统是示例性的，并且可以使用各种其它类型的机器视觉系统。例如，在一些实施方式中，所述控制系统 108 可以不存在，例如，当所述机器视觉模块的输出被存储时，输出是用于进一步的处理等，而不是用于控制。在一些实施方式中，摄像头 102 可以被替换为，例如存储的图像数据源。在一些实施方式中，可包括附加的元件，例如，不同的处理器或控制器、用户控件、输入或输出装置等。

[0088] 在各种实施方式中，摄像头 102 可以是能够将视觉刺激转换为数字形式，例如，数字图像流的任何设备。各种实施方式可包括基于电荷耦合器件 (CCD) 的设备 ；有源像素传感器 (APS)，例如互补式金属氧化物半导体 (CMOS) 传感器、薄膜晶体管 (TFT)、光电二极管阵列，以及上述各项的组合。

[0089] 由摄像头 102 所产生的数字图像中的每一个可包括至少 0.01 百万像素、至少 0.1 百万像素、至少 100 万像素、至少 2 百万像素，或更多，例如，在 0.01-1000 百万像素或其任何子范围的范围内。所述数字图像流的特征可以是帧速率 (即，每秒的图像帧数) 为至少 10Hz、至少 50Hz、至少 100Hz 或更多，例如，在范围为 1-1000Hz 或其任何子范围的范围内。所述数字图像可以是彩色、灰度、黑白的，或者其它合适类型的图像。

[0090] 在一些实施方式中，摄像头基于电荷耦合器件 (CCD)。在一个实施方式中，摄像头 100 是 Point Grey Firefly MV 装置 (具有 752x480 像素，8 位 / 像素，每秒 60 帧)(Point Grey Research，里士满，不列颠哥伦比亚省，加拿大)。在另一个实施方式中，摄像头 100 是 E-consystems eCAM50_OMAP_GSTIX，它集成了 OmniVisionOV5642 的摄像头模块，具有 1280x720 像素，8 位 / 像素，每秒 30 帧)。

[0091] 在一些实施方式中，图像由摄像头 102 获取并以足够的速度传送到编码器模块 104，以允许设备 100 无滞后地进行操作。要做到这一点，在一些实施方式中，摄像头 102 和编码器模块 104 之间设置有高带宽连接。例如，可以使用摄像头和处理设备之间

的 USB2.0 接口来实现大于 20 百万字节 / 秒的数据传输。在其它实施方式中，在摄像头和处理设备之间使用并行接口，如集成在 OMAP3530 处理器的摄像头图像信号处理器（Texas Instruments，达拉斯，德克萨斯州）中的并行接口。在各种实施方式中，可以使用其他合适的连接，包括有线或无线连接。摄像头 102 与编码器模块 104 的接口可通过能高速传输数据的任何连接实现，包括但不限于，串行接口，例如 IEEE1394 或 USB2.0；并行接口；模拟接口，例如 NTSC 或 PAL；无线接口。在一些实施方式中，摄像头可以与编码器模块集成到同一块板上。

[0092] 通过本文所述的技术，编码器模块 104 实施图像流的处理，例如包括实施编码器将图像转换为代码，模拟视网膜电路的操作。由编码器确定的转换被应用到系列输入图像，产生编码输出。例如，编码后的输出可以是指示视网膜细胞放电频率的值的形式，所述放电频率为如果图像由视网膜接收时，由视网膜细胞所产生。所述输出也可以是，例如，指示视网膜细胞"发生器电势"的信息，即，视网膜模型的线性部分的输出（具有线性滤波器的图像卷积输出）。编码后的输出可以指示由视网膜细胞所产生的"尖峰"脉冲串。

[0093] 在一些实施方式中，由于存在不同类型的视网膜输出细胞，不同编码器的集合可被用于更好地模拟正常视网膜的处理。差异可与特定细胞类型（例如，ON 细胞或 OFF 细胞）对应，或与视网膜上的细胞位置（例如，在视网膜中央的细胞对周边的细胞）对应。当编码器模块 104 具有不止一个编码器时，所述编码器可以并行操作，这可以独立地或通过至少一个或多个连接机构来实现。

[0094] 图 2 是举例说明编码器模块 104 的示例性实施方式的操作流程图。在步骤 201 中，编码器模块 104 接收来自摄像头 102（或一些其它合适的来源）的一系列图像。在可选的步骤 202 中，这些原始图像经过预处理，例如，重新调整图像的对比度 / 亮度，将噪声滤波器用于图像，裁剪图像等。

[0095] 在步骤 203 中，对原始图像进行处理，以确定指示视网膜细胞响应于所述图像的信息。例如，在一个实施方式中，在图像区域中的各位置，编码器处理所述图像流并输出随时间变化的值，所述值与将所述图像流投射到视网膜上时，由视网膜细胞（或细胞群）所产生的放电频率对应。在一个实施方式中，放电频率输出的格式如下：对于给定的时间 t，输出为比特矩阵，其中在位置 (x, y) 处的元素，与在位置 (x, y) 处的视网膜细胞的放电频率对应。

[0096] 值得注意的是，在一些实施方式中，编码器可以产生使用度量值而不是放电频率指示的视网膜细胞响应的信息。例如，编码器的输出可以与如上文所述的细胞的激活状态、细胞内电势、发生器电势等对应。

[0097] 在步骤 204 中，来自步骤 203 的编码信息被用于生成适于被机器视觉模块 106 处理的图像（在本文中被称为"视网膜图像"或当用来指代随时间变化的图像时所用的"视网膜图像流"或"视网膜图像数据流"）。例如，当所述编码信息作为放电频率矩阵被输出时，如上文所述，可以产生放电频率视网膜图像，其中"视网膜图像"中每个像素的强度由矩阵中对应元素的放电频率值确定（例如参见图 3）。可以使用放电频率和像素强度之间的任何合适的关系，包括线性关系、非线性关系、多项式关系、对数关系，等。放电频率和像素强度之间的转换可以使用任何合适的技术，包括使用查表来实现。在一些实施方式中，视网膜图像中放电频率可以用图像特征而不是强度来表示。例如，在视网膜图像是彩色图像的实施

方式中，每个像素的颜色空间坐标可以与放电频率对应。

[0098] 在可选的步骤 205 中，视网膜图像被后期处理。可以使用任何合适的处理技术，包括，例如，重新调整、滤波、剪切、平滑处理等。在步骤 206 中，视网膜图像被输出到机器视觉模块 106。

[0099] 值得注意的是，在一些实施方式中，可以省略步骤 204 和步骤 205。在这种情况下，编码器的输出可以直接传送到机器视觉算法进行处理。对本领域技术人员而言，如下文中将更加显而易见的是，在某些情况下这可能需要改进已知的机器视觉算法，以接收未格式化为传统图像数据的输入数据。然而，在许多实施方式中，可以通过简单的方式实现，而不需要改进特定算法的核心概念。

[0100] 在一些实施方式中，每个编码器执行预处理步骤，随后为时空转换步骤。所述预处理步骤是重新调整的步骤，这可以在处理设备的预处理器模块中执行，所述预处理步骤将真实世界的图像 I，映射成量 X，X 在时空转换的工作范围内。值得注意的是，I 和 X 是随时间变化的量，即 I(j, t) 代表真实图像中每个位置 j 和时间 t 的强度，并且 X(j, t) 代表预处理步骤的相应输出。预处理步骤可以进行如下映射 ：通过 X(j, t) = a+b I(j, t) 将 I(j, t) 映射到 X(j, t)，其中 a 和 b 是选定的常数，所述常数用于将真实世界图像强度的范围映射为时空转换的工作范围。

[0101] 还可以使用变量史 (variable history) 进行重新调整，以确定 a 和 b 的量，并且可以使用开关设置这些量在不同条件下（例如，不同光照或不同常数）的值。

[0102] 对于灰度图像而言，针对各位置 j 和时间 t，I(j, t) 和 X(j, t) 均只有一个值。

[0103] 对于有色图像而言，采用相同的策略，但是其被分别用于红、绿和蓝各颜色通道。在一个实施方式中，针对各位置 j 和时间 t，强度 I(j, t) 有三个值 (I_1, I_2, I_3)，其中这三个值 I_1, I_2, I_3 分别表示红、绿、和蓝的强度。然后利用上述转换将各强度值重新调整为其对应的 X 值 (X_1, X_2, X_3)。

[0104] 在一个实施方式中，采用线性 - 非线性级联（在 Chichilnisky EJ2001 ；Simoncelli et al2004 中有综述）实现时空转换步骤，其中各神经节细胞 m 的放电频率 λ_m 由下式给出 ：

[0105] $\lambda_m(t;X) = N_m((X*L_m)(j,t)$ (1)

[0106] 其中 * 表示时空卷积，L_m 是线性滤波器，L_m 对应于第 m 个细胞的时空核，并且 N_m 是描述第 m 个细胞非线性的函数，如之前章节所述，X 是预处理步骤的输出，j 是像素位置，t 是时间。随后可使用放电频率 λ_m 来生成如上文所述的放电频率视网膜图像。

[0107] 采用空间函数和时间函数的乘积对 L_m 参数化。例如，在一个实施方式中，空间函数由网格中各像素的权重组成（例如，照相机中的数字化图像），但还可以使用其它替代方案，如网格上正交基函数之和。在一个实施方式中，网格由 10×10 的像素阵列组成，整个视觉空间为 26×26 度（其中在视觉空间中每个像素为 2.6×2.6 度），但是也可以使用其它替代方案。例如，由于对应于视网膜神经节细胞的视觉空间面积随视网膜上空间位置和物种的不同而不同，因而总阵列尺寸可能不同（例如，从为或约为 0.1×0.1 度至 30×30 度，其对应于在 10×10 的像素阵列中，各像素的视觉空间为或约为 0.01×0.01 度至 3×3 度）。可以理解，像素阵列的角度范围和尺寸仅用于解释某个特定的实施方式，在本发明还包括其它的像素阵列角度范围或尺寸。对于任意选定的阵列尺寸，阵列中的像素数还可以

依据细胞代表的视觉空间中区域的形状而不同（例如，从为或约为 1×1 至 25×25 像素的阵列）。类似地，时间函数由若干时间块的权重之和组成，其在其它时间块的对数时间为升余弦函数（Nirenberg 等 2010；Pillow JW 等 2008）。也可以使用其它替代方案，如正交基函数之和。

[0108] 在所述实施方式中，时间样本跨距为 18 个时间块，均为 67 毫秒，总持续时间为 1.2 秒，但也可以使用其它替代方案。例如，由于不同神经节细胞具有不同的时相性质，因而以块计的持续时间跨距和表示细胞动力学所需的块数均可以不同（例如，持续时间为或约为从 0.5 至 2.0 秒，块数为或约为从 5 至 20）。时相性质还可以因物种不同而不同，但是此改变仍包括在上述范围之内。

[0109] 还可以对公式 1 进行修改，以包括修改编码器输出的项，其依据既往史（即，细胞 m 已经产生的峰电位序列）和其它神经节细胞输出的既往史（Nirenberg 等 2010；Pillow JW 等 2008）。

[0110] 在另一个实施方式中，线性滤波器 L_m 被参数化为 Q 项之和，其中各项为空间函数和时间函数的乘积。

[0111] $$L_m = \sum_k^Q S_k \otimes T_k$$

[0112] 其中⊗表示外积，S_k 和 $_Tk$ 分别为 kth 空间和时间函数（k 的范围为 1 至 Q）。

[0113] 在本实施方式中，如前文所描述的，可以对各空间函数进行参数化，例如作为网格上各像素的权重，或作为网格上正交基函数之和。如前所述，也可以对各时间函数进行参数化，例如在若干时间块作为权重之和，以及在其他时间块作为对数时间的升余弦函数。也可以使用其他替代方案，如正交基函数之和。

[0114] 在一个实施方式中，Q 为 2，和 L_m 可以表示为

[0115] $$L_m = S_1 \otimes T_1 + S_2 \otimes T_2$$

[0116] 其中⊗表示外积，S_1 和 T_1 表示第一对空间和时间函数，以及 S_2 和 T_2 表示第二对空间和时间函数。

[0117] 对于 L 的两组参数（空间和时间），可通过两个因素确定分辨率（像素尺寸，块尺寸）和跨距（像素数，时间块数）的选择：需要获得视网膜代码的合理近似的替代，并且需要保持参数的数量足够少，以使其能够通过实际最优化程序确定（例如，在假体应用中详述）。例如，如果参数数量太少或分辨率过低，则替代将不够准确。如果参数数量过多，则最优化程序将出现过度拟合，将无法获得转化结果（公式 1）使用适宜的基函数集合是一种能够减少参数数量并因此避免过度拟合的策略，即“降维”策略。例如，可以通过 10 个权重之和与基函数对时间函数（覆盖 18 个时间块，各为 67 毫秒）进行参数化；参见假体应用的“实施例 1，构建编码器的方法”部分和（Nirenberg 等，2010；Pillow JW 等 2008）。

[0118] 采用三次样条函数对非线性 N_m 进行参数化，但是也可以采用其它参数化方法，如分段线性函数、高阶样条函数、泰勒级数和泰勒级数的商数。在一个实施方式中，用带有 7 个结点的三次样条函数对非线性 N_m 进行参数化。对结点数量进行选择以准确捕获非线性形状，同时避免过度拟合（参见上述关于过度拟合的讨论）。需要有至少两个结点以控制终点，因此结点数的范围可以从约 2 到至少约 12。结点的间距要覆盖模型的线性滤波器输出

给出的数值范围。

[0119] 对于时空转换步骤而言，除了上述线性-非线性(LN)级联以外，替代映射也包括在本发明的范围内。替代映射包括，但不限于，人工神经网络和其它滤波器的组合，如线性-非线性-线性(LNL)级联。此外，时空转换可以加入来自峰电位产生阶段的反馈(见下文)以提供历史相关性和神经元间的相互关系，如(Pillow JW 等 2008 ;Nichols 等，2010)中描述。例如，可以通过将附加滤波器函数与峰电位产生器的输出进行卷积运算，并将这些卷积的结果通过公式 1 中非线性的验证而实现。

[0120] 时空转换步骤还可以使用其它模型。模型的非限制性例子包括以下模型 :Pillow JW 等 2008 中所描述的模型 ;动态增益控制 ;神经网络 ;表示为接近离散时间步长的积分、微分和普通代数公式的模型，其形式和系数通过实验数据确定 ;表示为由线性投射(输入与时空核的卷积)和非线性失真(通过参数化的非线性函数对得到的标量信号进行转换)所组成的顺序步骤结果的模型，通过实验数据确定其形式和系数 ;时空核为少量项之和的模型，所述各项为空间变量函数与空间变量函数与时间变量函数的乘积，其通过实验数据确定 ;所述空间和/或时间函数以一组基函数的线性组合表示的模型，基函数集合的大小小于空间或时间样本的数量，通过实验数据确定其权重 ;非线性函数由一个或数段组成的模型，其均为多项式，其截点和/或系数通过实验数据确定，且模型为上述模型输出的组合，其可能递归地通过如加、减、乘、除、开方、乘方以及超级函数(例如，求幂、正弦和余弦)等计算步骤组合。

[0121] 如在假体应用中所描述，上文所述类型的编码器可以非常近似地模拟真实视网膜细胞的输入/输出功能。如本文所详述，在某些情况下，其特征可能在于，确定在每个像素处的重构视网膜图像的值与对应原始图像的值之间的标准皮尔森相关系数。因此，相关系数为 1 表明原始图像的所有信息被完全保留，而相关系数为 0 则表明重建与真实图像间相似的可能性很小。

[0122] 例如，在一些实施方式中，设置编码器从而使得测试输入刺激和从编码器数据重构的相应刺激之间的皮尔森相关系数为至少约 0.35、0.65，至少约 0.95，或更高，例如，在 0.35-1.0 或其任何子范围的范围内，所述编码数据由编码器响应所述测试输入刺激时产生。在一些实施方式中，所述测试输入刺激包括一系列自然场景(例如，时空转变的场景)。

[0123] 在一些实施方式中，对于较宽范围的输入，本文所述类型的视网膜编码器模拟真实视网膜细胞的输入/输出功能，例如，空间-时间变化的自然场景。在典型的实施方式中，该性能本质上比传统的编码器更好。

[0124] 图 18A-F 举例说明了，当用自然场景的影片，包括风景、人步行等进行测试时，各细胞(分别为细胞 1-6)的视网膜编码器模型的性能。在每幅图中，左侧显示的是常规的线性-非线性(LN)模型的性能，并且右侧显示的是在本申请中所述类型的线性-非线性(LN)模型的性能。通过光栅图和直方图(PSTHs)显示性能。常规的线性-非线性(LN)模型的开发仅基于视网膜细胞对白噪声刺激的实验响应。与此相反，本申请中所述类型的线性-非线性(LN)模型的开发是基于所记录的细胞对于白噪声和自然场景刺激的响应。

[0125] 对于所示的例子，两种类型的模型的输入测试刺激都是在纽约中央公园拍摄的自然场景的影片。如图所示，标准 LN 模型对自然场景的刺激不是非常有效 :也就是说，用白噪声构建的这个模型不会产生与真实细胞的尖峰模式近似匹配的尖峰模式。与此相反，本申

请中所述的LN模型利用白噪声和自然场景的刺激所构建,它是非常有效的。它产生的尖峰模式与真实细胞所产生的近似匹配。(值得注意的是,用于测试模型的自然场景的影片与用于训练模型的不同,因为后者为验证任何模型所需。在每幅图中,值得注意的是,使用相同的真实细胞作为两种类型的模型的基准。最后,值得注意的是,本文所述类型的编码器模型的性能已经通过许多其他的刺激得到证实,包括面部、人行走、孩子们玩耍、山水、树木、小动物等的影片,如假体应用中的图,以及Nirenberg等.Retinal prosthetic strategy with the capacity to restore normal vision,PNAS2012,以及可从

[0126] www.pnas.org/lookup/suppl/doi:10.1073/pnas.1207035109/-/DCSupplemental获得的随附的补充信息部分中所示)。

[0127] 可以从PSTHs得出关于性能的相同结论。浅灰色的踪迹显示的是真实细胞的平均放电频率;深灰色的踪迹显示的是模型细胞的平均放电频率。标准LN模型缺少放电频率的许多特征;图18A-18F中的每幅图均显示了标准模型所缺少区别特征的例子。但是本申请中所描述的模型,可靠地捕获了放电频率的特征,并且对于不同细胞的阵列也是如此(在假体应用中也显示了许多其他的例子)。

[0128] 图3A举例说明了原始图像被转换成视网膜图像。板A显示了由摄像头102获取的原始图像流的若干帧。如图所示,原始图像流包括人步行穿过复杂的环境。板B显示了相应的视网膜图像的若干帧,其中所述视网膜图像的像素强度与由编码器模块104的编码器生成的放电频率对应。显示了四个不同的视网膜图像流,其中每个使用不同的细胞阵列(OFF侏儒细胞、ON侏儒细胞、OFF伞状细胞和ON伞状细胞,如图中所示)。需要注意的是,显示的视网膜图像的若干帧由编码器模块104在短暂的时滞后产生,其与天然视网膜的处理时滞时间对应(如所示,约80毫秒)。

[0129] 需要注意的是,很明显,视网膜图像中所包含的信息总量小于原始图像中的信息总量。信息量的减少可以有利地减少机器视觉的处理负荷。此外,由于编码器模拟视网膜的行为,对于一些机器视觉的应用,保留在视网膜图像中的信息将包括现有机器视觉任务所需的显著特征,从而允许机器视觉模块106高效和有效地运行。

[0130] 图3B-3F显示了与图3A最后一栏对应的原始图像(图3B)和视网膜图像(图3C-3F)的放大图。在原始图像中,一个人像正在一个相对静态、但是复杂的环境中从右向左移动。需要注意的是,在所有的视网膜图像(图3C-3F)中,静态环境已经不再强调改变程度,

[0131] 而一直强调移动的人的模式。此外,在这两个图像中,"移动阴影"型效应明显地尾随人物影像,这指示了运动的方向。因此,虽然图像中包含的信息总量已经减少,但仍然强调特征,重要特征,即移动的人形。

[0132] 需要注意的是,这些效果都不是任何有意设计的编程的结果。也就是说,编码器没有被有意地编程来确定移动的特征。相反,强调这些特征是编码器模拟发生在视网膜上的自然过程这一事实的结果。尽管某些种类的强调特征在本实例中是显而易见的(在静态背景中的人形移动),可以理解的是,对于其他类型的输入图像,视网膜可能强调其他类型的特征。核心概念是,对于任何给定的图像,强调的特征通常是那些基于数百万年视网膜进化而被确定为显著的特征。因此,如下文所详细描述,当所述视网膜图像用于已知生物视觉系统表现良好的机器视觉应用中时,所述视网膜图像会特别有优势(例如,某些类型的图形

识别任务，例如，面部识别、识别相对于复杂环境的人或其它生命形式、复杂环境中的导航、对移动物体的快速跟踪和反应等）。

[0133] 在一些实施方式中，编码器编码图像数据的时间与由正常或接近正常的视网膜进行的编码的时间大致相同。在不同的实施方式中，编码器以可接受的处理迟滞时间运行。如本文所使用的，处理迟滞时间是指由摄像头 102 接收到视觉刺激中事件发生到相应的输出代码（例如，相应的视网膜图像）递送至机器视觉模块 106 之间的时间量。在一些实施方式中，编码模块的迟滞时间小于约 50 毫秒、小于约 20 毫秒、小于约 10 毫秒、小于约 5 毫秒等，例如，在 5-50 毫秒或其任何子范围的范围内。

[0134] 再次参考图 1，机器视觉模块 106 从编码器模块 104 接收视网膜图像，并用任何合适的机器视觉技术处理所述图像。尽管本文提到了许多这样的技术，应当理解的是，这些实施例并非限制性的，也可使用其它技术。例如，在不同的实施方式中，可以使用在 D. A. Forsyth, J. Ponce Computer Vision:A Modern Approach, 普伦蒂斯·霍尔出版社 (Prentice Hall)，第二版，2011 和 / 或 D. H. Ballard, C. M. Brown ;Computer Vision, 普伦蒂斯·霍尔出版社，新泽西，1982（可于 http://homepages.inf.ed.ac.uk/rbf/BOOKS/BANDB/bandb.htm 获 得），R. Szeliski, Computer Vision:Algorithms and Applications, Springer2010, 可于

[0135] http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf 获得）；以及 E. R. Davies, Computer and Machine Vision, 第四版 :Theory, Algorithms, Practicalities, Elsevier2012, 中所描述的一种或多种技术。

[0136] 在各种实施方式中，机器视觉模块 106 可以实施一项或多项可用的计算机视觉算法或软件工具，例如，OpenCV 软件包中包括的任何一项，可在

[0137] http://opencv.willowgarage.com/wiki/ 获得，或甘道夫 (Gandalf) 计算机视觉软件包，可在 http://gandalf-library.sourceforge.net/ 获得。

[0138] 机器视觉模块 106 可以使用视网膜图像以进行任何合适的任务，包括识别任务（例如，对象识别、图像分类、面部识别、光学字符识别、基于内容的图像检索、姿态估计等）、运动分析任务（例如，自我运动判定、运动追踪、光流测定等）、建模任务（例如，场景重建、三维体积识别等）。

[0139] 在一些实施方式中，机器视觉模块 106 可将视场划分为若干域，其尺寸可以是均等或不均等的。所述域可以重叠或不重叠。所述域可以覆盖视场的频带（例如，水平轴上的整个视野和垂直轴上的有限跨度），或者可以覆盖整个视场。

[0140] 在一些实施方式中，机器视觉模块 106 可以将边界边缘检测技术用于视网膜图像，包括，例如，一阶边缘检测技术，如 Canny 边缘检测；二阶边缘检测技术；或者基于相位一致的边缘检测技术。边缘检测可能涉及对视网膜图像应用一项或多项转换，例如，霍夫 (Hough) 转换。

[0141] 在一些实施方式中，机器视觉模块 106 可根据视网膜图像流计算光流。光流可指示视场中由观察者（眼睛或摄像头）与场景之间的相对运动引起的对象、表面和边缘的明显运动的模式。光流可用于任何数量的应用中，包括运动检测、目标分割、碰撞时间和扩展计算焦点等。用于计算光流的方法可以包括，相位相关法、基于块的方法、差分方法（如 Lucas-Kanade、Horn-Schunck、Buxton-Buxton 和 Black-Jepson 方法）、变分法、离散优化方

法等。

[0142] 在一些实施方式中，机器视觉模块 106 可以将一项或多项图像分割技术用于分割视网膜图像（例如，为了识别目标区域）。示例性的分割技术包括阈值化、聚类方法、基于压缩的方法、基于直方图的方法、边缘检测（例如，使用上文所述的边缘检测技术）、区域生长法分割与合并方法、基于偏微分方程的方法（例如，水平集方法）、图划分方法、基于分水线变换的方法、基于模型的分割方法、多尺度分割、半自动分割、基于神经网络的分割等。

[0143] 在各种实施方式中，机器视觉模块 106 可以使用本领域中已知的任何计算机学习技术进行培训。计算机学习技术包括监督学习（例如，包括统计分类技术）、无监督学习、强化学习等。在一些实施方式中，机器视觉模块 106 可包括可以经训练后用于执行各种任务的一种或多种的人工神经网络。

[0144] 图 4 举例说明了用于机器视觉系统 100 的训练机器视觉模块 106 的示例性的训练系统 400。所述训练系统包括原始训练图像源 402（例如，存储图像的数据库）、编码器模块 404 机器视觉模块 108，以及控制器 406，所述编码器模块 404 根据原始训练图像采用本文所述的技术产生视网膜图像，机器视觉模块 108 从编码器接收视网膜图像，控制器 406 监测并根据所监测到的性能改变机器视觉模块的操作。

[0145] 图 5 是举例说明训练系统 400 的操作的流程图。在步骤 501 中，编码器 404 从源 402 接收训练图像。例如，训练图像可以是一系列的肿瘤医学图像，其中图像的第一部分已知与恶性肿瘤对应，而训练图像的第二部分与良性肿瘤对应。

[0146] 在步骤 502 中，编码器将原始训练图像转换为视网膜图像。在步骤 503 中，所述视网膜图像被输出到机器视觉模块 106。

[0147] 在步骤 504 中，控制器 406 监测机器视觉模块 106 在处理视网膜图像以执行任务时的性能。在医学图像的例子中，机器视觉模块 106 可应用图像识别技术，从良性肿瘤图像中区分出恶性肿瘤的图像。控制器监测机器视觉模块 106 在执行任务（例如，计算区分恶性肿瘤的错误率）时的性能。如果性能是可以接受的，则过程在步骤 505 结束。如果性能是不能接受的（例如，如果出错率超过阈值电平），在步骤 506 中，控制器 406 调节机器视觉模块 106（例如，通过修改一个或多个参数、通过改变人工神经网络中的连接等），并且该过程返回到步骤 503。因此，控制器 406 反复调节机器视觉模块 106，直到其性能达到可接受的水平（例如，出错率低于阈值电平）。

[0148] 值得注意的是，在各种实施方式中，可以使用其它合适类型的训练。例如，除了或可供选择地将性能与固定阈值比较，训练还可实施收敛准则（例如，其中迭代训练持续直至每次迭代中性能递增低于阈值电平）。

[0149] 在各种实施方式中，机器视觉模块 106 可以包括任何合适的控制技术，包括使用复杂的基于人工智能的系统。然而，对于许多应用而言，机器视觉模块 106 可以实施比较简单的控制方案。在一些这样的实施方式中，根据对编码器模块接收到的视网膜图像进行相对简单的即时分类，机器视觉 106 控制一个或多个系统操作的一部分或全部（例如，机器人的移动轨迹）。也就是说，控制不依赖于复杂的计划，但仅依赖于暂时的局部分类。有利的是，本领域中已知的学习算法已知适合于这些类型的相对简单的分类任务的性能。

[0150] 例如，参考图 6，在一个实施方式中，机器视觉系统 100 被用来控制机器人 600 来通过设有障碍物的环境，例如，如图所示的迷宫。机器视觉系统的摄像头 102 设置在机器人

600 上，并且具有能捕获机器人面前场景的视场。

[0151] 来自摄像头 102 的视频流由编码器模块 104 处理，以产生视网膜图像流。在一个例子中，编码器模块可以模拟鼠视网膜神经节细胞的性能（例如，使用特征在于如假体应用中题为“小鼠神经节细胞编码器参数示例性集合”部分所述的编码器参数的编码器）。在另一种情况下，编码器模块可以模拟猴视网膜神经节细胞的性能（例如，使用特征在于如假体应用中题为“猴神经节细胞编码器参数示例性集合”部分所述的编码器参数的编码器）。

[0152] 处理视网膜图像流，例如，使用光流技术，以确定图像中不同位置的移动速度。通常，图像中速度较慢的位置或域将与远离机器人 600 的物体对应，而速度更快的位置将与接近机器人的对象对应。为了避免撞上障碍，机器视觉模块 106 控制机器人朝一定方向移动，所述方向与图像中运动速度较慢的位置对应。

[0153] 例如，在一个实施方式（如图 7 所示）中，视场（即，视网膜图像数据流）被图像分割步骤 702 分成 N = 7 的大小相同的区域。在这个实施方式中，各区域不重叠，并且它们对摄像头的水平视场（即 40°）从左至右进行划分，从而使得每个区域水平跨越 5.7°；在垂直方向上，它们被限制在导航仪视场下半部分（即 27°），从而使得这些地区垂直跨越 13.5°）。以规律的间隔（例如，每 2 秒）从视网膜图像序列获取两个连续的视网膜图像，并传送到机器视觉模块 106 进行分类。由于每个视网膜图像被划分成 N 个区域，机器视觉模块接收 N 对区域。每一对区域通过卷积神经网络 (CNN) 704 传送，所述卷积神经网络对所述区域的光流速度进行分类。所述分类的输出可以是每个区域 i 的速度标签 Li，其中 Li 是介于 1 和 M 之间的数，1 代表在所述区域中非常缓慢的平均速度，并且 M 代表非常快的平均速度。例如，M 可以是 8，因此有 8 种不同的速度等级。

[0154] 结果是 N 分类的阵列 706；基于此，由转向决定模块 708 作出转向决定。选择速度分类最慢的区域为“目标区域”（所朝向的区域），也就是数 Li 最小。如果有多个区域都为最慢的速度分类，转向决定模块 708 可选择最接近中心的区域（从而使转向量最小），或根据系统所需要的用途选择其他一些区域。一旦选定目标区域，机器视觉模块 106（具体而言，机器视觉模块 106 中的转向决定模块 708）启动转向使导航仪面对目标区域的中心。

[0155] 上文所述的例子是指机器人的导航。应当理解的是，在各种实施方式中，上文所述技术可以被用于其它类型的导航系统，包括导航穿过虚拟世界，如下文的例子将所描述。

[0156] 例如，通过将视网膜图像流的图像场分成几个区域或域，并将各区域分类为各速度类别，并控制机器人 600 朝与最低速度的类别的图像区域对应的方向移动，机器视觉模块 106 可识别并避免障碍物。可以通过相对简单的训练算法，如上文所述的 CNN 以及在下文实施例所述的算法，或推进算法（例如，AdaBoost 算法，参见 Yoav Freund, Robert E. Schapire “A Decision-Theoretic Generalization of on-Line Learning and Application to Boosting”, 1995)，训练机器视觉模块 106 来执行上述分类任务。

[0157] 通常，所述装置和技术可被用于任何适当的应用，包括医学图像处理（例如，自动或计算机辅助医疗诊断）、机器人控制或导航、工业过程监测和控制、自动分拣应用、基于运动追踪的接口（例如，与计算机游戏系统一起使用）等。本文所述的装置和技术可实时或接近实时地操作，例如，允许上述的应用程序的实际自动化。

[0158] 实施例—虚拟世界导航

[0159] 在评价一种方法对机器视觉的有效性的一个实施例中，使用导航任务，因为

这是特别具有挑战性的（需要同时处理时间和空间）。这种方法应用通常用于导航的多种学习算法的不同方面，例如，LeCun，Y. 等所著 (2010) Convolutional Networks and Applications in Vision. Proc. International Symposium on Circuits and Systems (ISCAS'10), pp. 253-256. IEEE ;Szarvas, M. 等所著 (2005) Pedestrian detection with convolutional neural networks. Proc. Intelligent Vehicles Symposium, pp. 224-229. IEEE ;Jackel, L. D. 等所著 (2006) The DARPA LAGR program:Goals, challenges, methodology, and phase I results. Journal of Field Robotics, 23, 945 - 973，这些文献通过整体引用并入本文。使用这些技术，导航仪被构造成通过卷积神经网络 (CNN)（一种学习算法）来学习它的环境。使用名为 Theano（公众可从 http://deeplearning.net/software/theano/ 获得）的开源数值处理和自动微分包构建 CNN。

[0160] 设计导航仪以学习其所在的训练环境中物体的速度。给予导航仪一个训练环境，并使用它在每一时刻将训练环境划分成 n 个域。随后导航仪学习域中的速度。速度提供对导航有用的信息。如果某些物体以非常快的速度移动，这意味着它非常接近在导航环境（它快速移动经过视网膜）中的虚拟对象。如果它接近，虚拟对象很可能会碰撞。所以导航仪评价环境中的域，然后移向速度最慢（速度最慢的物体是最远和最安全的）的域。在这个实施例中，导航仪不被定向成朝向特定的终点，而是向前移动并且不与任何物体发生碰撞。

[0161] 更具体地说，在本实施例中，使用图 7 所示的方法，当导航仪穿过一个环境，通过图像分割步骤 702，它的视场被分成 7 个大小相等的区域。在本实施方式中，各区域不重叠，并且它们对摄像头的水平视场（即 40°）从左至右进行划分，从而使得每个区域水平跨越 5.7°；在垂直方向上，它们被限制在导航仪视场下半部分（即 27°），从而使得这些地区垂直跨越 13.5°）。

[0162] 在每个决定时间点，基于卷积神经网络 (CNN) 的算法对各区域中的光流速度进行分类（步骤 704）。所述分类的输出是每个域 i 的速度标签 Li，其中 Li 是介于 1 和 8 间的数，1 代表在所述域中非常缓慢的平均速度，并且 8 代表非常快的平均速度

[0163] 如上文所述，根据这些分类，7 个域每个一类，由转向决定模块 (708) 作出导航决定。选择具有最慢速度分类的域作为"目标域"（所朝向的域）。如果有多个域都为速度最慢的分类，导航仪选择最接近中心的域（从而使转向量最小）；如果速度分类仍然相同，导航仪将选择其左边的域。一旦选定目标区域，机器视觉模块 (106) 启动转向使导航仪面对所选定区域的中心。

[0164] 构建虚拟环境以用于培训和使用名为 Panda3D 的开源 3D 绘制架构（公众可从 http://www.panda3d.org/ 获得）进行测试。来自训练集合的若干帧的数据流如图 8 所示，来自三个测试组的帧的数据流如图 9A、9B、9C 所示。如图中所示，训练集合是田园环境。三个测试组如下：与训练集合中田园环境不同的田园环境、郊区环境和广场。

[0165] 在下述两种条件下比较导航仪的性能：1) 当用标准方法训练时，即使用原始图像流作为输入，以及 2) 当使用"视网膜图像流"作为输入进行训练，也就是说，当它使用由本文所述的编码器处理过的图像。在这种情况下，使用猴侏儒和伞状细胞，按照 Nirenberg, S. 和 Pandarinath 所著的 C. (2012) A retinal prosthetic with the capacity to restore normal vision. Proc. Natl. Acad., in press ；以及 Nirenberg, S. 等所著的 (2011) Retina prosthesis and the Prosthesis Applications 中所描述的方法来生成所使用的编码器，所述参考文献中的每一个通过引用而整体并入本文。

[0166] 如图 10A 所示，当导航仪通过原始图像流了解到它的环境时，其性能较低，发生许多碰撞；从训练集合学到的并不能推广到新的环境。如图 10B 所示，当导航仪通过视网膜图像流了解到它的环境时，表现显著提高：值得注意的是直线路径和无碰撞。这明显显示能推广到新的环境（田园、郊区、广场）- 这对于人工导航系统和一般的机器学习算法一直是悬而未解的。

[0167] 图 11 显示了当使用视网膜图像流作为输入时，导航仪的高性能进一步得到证实。具体而言，它表明了高性能不仅能推广到不同的环境（从田园到郊区到广场），它也可以推广到环境中不同的光线条件下。A 到 F 与太阳的不同位置对应，因此广场环境中有不同的阴影条件；跨越日出到日落的光线条件，即，环境左侧地平线上 30 度至右侧地平线上 30 度。如图中所示，当用原始图像流训练导航仪时（使用一种照明条件下的田园环境，如图 8 所示），导航仪的性能并不能推广：它在广场中的性能较低，并且这在各种光照条件下仍然如此。图中每条柱的高度与试验分数对应，在所述试验中导航仪成功留在广场的轮胎路线中而并不与轮胎中的任何一个碰撞。误差线表示平均数标准误差 (SEM)。相反，当用视网膜图像流训练导航仪时（使用同一种光照条件的相同的田园环境，但不同的是通过编码器处理），导航仪的性能较高，并在各种光照条件下保持这种高性能。因此，用视网膜图像流训练（即，用编码器处理过的图像进行训练）导致高性能，这能同时推广到新环境以及多种照明条件（日出到日落，见上文）。

[0168] 需要注意的是，编码器实时操作，这表明处理技术也可以容易地应用到非虚拟环境，例如，为了控制机器人在真实世界环境中的运动。

[0169] 实施例 - 面部识别

[0170] 本实施例评价本申请所述方法对另一个在机器视觉中长期未解决的问题，即在视频中识别面部的有效性。使用通常用于脸部识别和行人检测的学习算法［参见 Viola 和 Jones2001；Viola 和 Snow2005］，构建系统以识别视频中的人脸，也就是说，能够将一个前所未见的作为“目标面部”的图像流与另一个面部或“非目标”面部区分开的系统。可以将同样的方法用于许多其它目的，例如，但不限于，行人检测、对象识别、对象追踪、全人识别、虹膜检测等。通过 Python 编程语言和 NumPy 数值计算软件包实施所述系统。

[0171] 所述方法的实施方式如图 12 所描述。输入视频（原始图像流）通过视网膜编码器 104，产生视网膜图像流。由于任务重点在于面部，随后剪裁所述视网膜图像流以定位包含面部的区域 1202。（编码器处理原始数据流后完成剪裁，从而可以避免进行编码时的边缘效应）在这个实施例中，手动选择包含面部的区域，从而构造已知面部例子的训练和测试集合。在其他实施方式中，通过 Viola-Jones 算法 [Viola 和 Jones, 2001]，可在原始图像流中或处理后的图像数据流中，检测到包含面部的区域。随后剪裁后的视频送至分类器 1206（例如，基于 Haar 过滤器的级联增加的分类器，如 Viola 和 Jones 和 Snow, 2005 所述）。分类器 1206 指定其为“目标面部”（这意味着它是目标个体的面部）或“非目标面部”（这意味着它是不同个体的面部）。

[0172] 图 15 显示了本申请所述方法的有效性的实施例。对于此分析，使用来自 http://www.cs.tau.ac.il/ ? wolf/ytfaces/ 视频中的面部的数据组。参考 Lior

Wolf,Tal Hassner 和 ItayMaoz 所著的 Face Recognition in Unconstrained Videos with Matched Background Similarity.IEEE Conf.on Computer Vision and Pattern Recognition(CVPR),2011。

[0173] 使用这个数据集进行多个面部识别任务。通常的方法是用“目标面部”训练面部识别算法。向所述算法呈现显示人脸（即目标面部）的视频阵列。通过将之前未曾见过的相同人脸视频与其他面部（即“非目标面部”）的视频一起呈现，测试所述算法识别面部的能力。所述算法的工作是对测试视频中目标面部或非目标面部进行正确区分。

[0174] 图 13 和 14 显示了来自实施例视频的视频。图 13 显示来自用于训练面部识别算法的视频的若干帧，并且图 14 显示来自用于测试所述算法的视频的若干帧。如图所示，在测试视频中的人（图 14）与训练视频中的人（图 13）是相同的，但出现在不同的环境中并具有不同的发型等。

[0175] 在下述两种条件下测试所述算法的性能：当用标准方法训练时（即，使用面部的原始图像数据流时），以及用所述面部的视网膜图像流（即，用本申请所述的编码器处理后的原始图像流）训练时。在这两种情况下，都使用了短的（两帧）影片进行培训。训练中使用的两帧影片的数量为，对于目标面部为 250-800（取自 4-5 个不同的视频），对于非目标面部为 2000（取自 >100 视频）。随后使用 50-800 个两帧影片测试性能，所述两帧影片取自之前未曾见过的视频，即未用于培训的视频。

[0176] 如图 15 所示，编码器的使用对性能有明显的影响。显示了两种任务的结果：第一种由非常具有挑战性的任务组成，其定义为标准方法的表现非常差的任务；第二种由更容易的任务组成，其中标准方法的表现适度良好。如图所示，当任务比较困难时（图 15A），采用编码器的方法比标准方法有很大（4 倍）的改进。当任务不是那么具有挑战性时，即当标准方法的表现适度良好时，采用编码器的方法仍然提供了本质上的改进（与标准方法相比提高 1.5 倍）。

[0177] 在一个可选的实施方式中，任务被略加修改，从而绕开面部检测步骤，作为替代，对分类器 1206 而言剪裁为适当大小的视频从输入视频中自动生成，其中无论面部是否呈现在视频的特定部分。随后，对这些新剪裁的视频进行如前所述的分类，或进行修改的分类，其中输出类别是“目标面部”和“非目标面部”或“非面部”。

[0178] 在一个可选的实施方式中，可以使用 N 帧进行分析，其中 N 可以是 1、3 或更多帧，帧数可多达处理器能够处理的程度，而不是图 15 中用于分析的两帧视频。

[0179] 此外，可以由它们自己进行分类，例如警告用户视频中出现了个体，或者它们可以以某种方式进行组合，例如在发出信号前等待在特定的时间窗内发生多个阳性检测（“目标面部”分类）。

[0180] 需要注意的是，虽然已经描述了向机器视觉提供的许多视网膜应用的例子，可使用涉及许多其它应用的实施方式。

[0181] 通常，对于动物（脊椎动物）表现良好的视觉任务，编码器方法可能是有利的，特别是对于已知动物视觉系统比现有机器技术更好的那些任务。如上文所述，在减少来自原始图像流的信息总量（例如，允许处理或更快速度的处理）的同时保持数据的显著特征有利的情况下，编码器方法可能特别有效。例如，如上文所指出的，在一些实施方式中，例如，当用于某些类型的识别任务，例如，面部识别、识别处于复杂背景中的人或其他生物形式、

导航穿过复杂的环境、对移动物体的快速追踪和反应等，编码器方法通常是特别有利的。

[0182] 值得注意的是，对于生物系统通常表现不太好的某些应用，编码器方法可能有局限性。这在需要高度细化的信息或精密测量的情况下可能特别明显。例如，再次参考图 3B-F 所示的视网膜图像，需要注意的是，虽然这些图像有利地强调人像的存在和运动，但是视网膜图像并没提供所述人像的清晰轮廓，对于例如确定精确的生物辨识信息，如人的绝对身高或其他的绝对身体尺寸而言，清晰的轮廓是有用的。为了确定这些类型的信息，将机器视觉算法应用于原始图像可能是更好的。

[0183] 在一些实施方式中，可以使用混合方法以同时获得基于编码器的方法以及传统方法的优点。所述基于编码器的方法用于机器视觉，所述传统方法用于原始图像数据。

[0184] 例如，在一些实施方式中，可以使用任何本文所述的基于视网膜编码器的技术处理原始图像流。可以处理所得到的视网膜图像数据（例如，使用机器视觉算法，如用视网膜图像训练的机器视觉算法），并将结果用于报告随后的相应原始图像分析（例如，使用机器视觉算法，如用原始图像训练的机器视觉算法）。

[0185] 图 16 显示了这种类型的示例性的过程。在步骤 1701 和 1702 中，通过本文所述的任何技术获得原始图像，并用于产生视网膜图像流。在步骤 1703 中，分析所述视网膜图像，例如通过机器视觉算法实现。

[0186] 在步骤 1704 中，视网膜图像的分析结果被用来识别目标视网膜图像（或其片段）。例如，在识别人的任务中，以正常视网膜产生视网膜图像的方式执行图像降维的编码器方法，可允许快速识别身体类型 -- 通过步态、特征鲜明的姿势等。它的一个优势在于，它迅速抽出运动信息，这对于此目的特别有用。因此，编码器方法可作为预筛选方法，来减少与目标个体可能匹配的空间（通过排除具有错误体型、步态、姿势等的候选人）。

[0187] 在步骤 1705 中，可以分析与识别的视网膜图像对应的原始图像（或其片段）。例如，在识别人的任务中，可以将使用原始图像（其中很少或没有使用降维）的算法应用于图像集，从而通过更详细的特征分析（例如，通过提取详细的生物特征信息，如人的准确高度或其它身体尺寸），更可靠对人进行识别。

[0188] 在各种实施方式中，上述方法可以反过来，即先对原始图像进行预筛选，然后是使用视网膜编码器方法的后续分析。在一些实施方式中，迭代技术可以与多轮交替的基于原始和编码器的分析一起应用。在其它实施方式中，可以平行进行不同类型的处理，并综合各结果。一般来说，可以使用传统方法和基于编码器的方法的任何合适的组合。

[0189] 如上所述，在各种实施方式中，进行视网膜处理以减少来自原始图像数据的信息总量（以实现效率，在某种程度上类似于视网膜的方式），同时对于给定的应用保持显著特征。例如，在一些实施方式，即使减少视网膜编码数据中的信息总量，当用于编码数据时，机器视觉算法表现出比用于相应的原始图像数据时更好的性能。在上文所述的两个实施例中都可以观察到这个结果，其中用于“压缩的”视网膜图像的导航和面部识别算法本质上优于应用于原始图像的相同算法。

[0190] 在各种实施方式中，视网膜编码数据可被压缩至少 1.5、至少 2、至少 3、至少 4、至少 5，或更多倍，例如，在 1-100 或其任何子范围的范围内。在一些实施方式中，这种压缩与由编码器产生的降维相对应。例如，在一些实施方式中，视网膜编码器的比特率可被量化，并可以与被编码器用作刺激的原始图像数据的熵（可检测为每单位时间内的比特）进行比

较，并且将比值用于确定压缩比。例如，在假体应用中描述的一些情况下，描述了与 4.9 比特 / 秒的输入原始数据比特率相比，比特率为 2.13 比特 / 秒的编码器。因此，在本实施例中，由编码器产生的数据压缩几乎是 7 倍。

[0191] 在一些实施方式中，本文所描述的处理技术可以用于信息存储和上下文检索。参照图 17，系统 1800 包括存储装置 1801（例如，硬盘驱动器或其他计算存储器），所述存储装置 1801 可操作地与处理器 1802 耦合。存储装置 1801 存储视网膜图像数据，用本文所描述的技术从原始图像数据生成所述视网膜图像数据。如上文所详述，在一些实施方式中，相对于原始数据，视网膜图像数据可以被压缩，同时保持一定的显著特征。因此，在某些实施方式中，所存储的视网膜数据可以用作相应的原始数据的代表，或者“指纹”。在一些实施方式中，存储装置存储指示编码数据和原始图像数据之间对应关系的数据库信息。例如，一个特定的视频剪辑可以被用来产生相应的视网膜图像流，并且在设备 1801 上存储的视网膜图像流具有可以与原始视频剪辑区分开的标记。

[0192] 在一些实施方式中，可以用处理器 1802 将输入数据与存储在存储装置 1801 上的数据相匹配。在一些实施方式中，处理器 1802 可以接收与一系列查询图像对应的查询图像数据（例如，原始视频剪辑）。然后，处理器 1802 可以用视网膜编码器处理查询图像数据，以生成视网膜编码的查询数据。然后，处理器可以将视网膜编码的查询数据与存储在存储装置 1801 上的视网膜编码数据进行比较。如果发现匹配，该处理器可以读取所存储数据的标记，并输出与查询数据视频剪辑关联的信息，其中视频剪辑被用于生成匹配的存储视网膜图像。在一些实施方式，因为视网膜编码数据被压缩和 / 或已具有增强的显著特征，与试图和相应的原始图像剪辑直接匹配相比，编码存储和查询数据间的匹配可以更快和 / 或更精确。

[0193] 在本申请和假体应用中所示的实施例使用从小鼠和猴视网膜得到的数据构建的编码器。然而，可以理解的是，不同的实施方式还可以使用从其他物种构建的编码器，其他物种，例如，但不限于鸟、猫、蛇、兔，可以使用假体应用中完整详细地描述的方法来构造。

[0194] 在各种实施方式中，本文所述技术的整体功能使用由视觉系统（特别是视网膜）进行预处理，以增强机器视觉（特别是降维）。对于一些应用，可以应用由其他物种的视网膜进行的预处理；例如，从鸟视网膜构造的编码器可能对飞行导航仪特别有效；同样，从快速移动的动物（例如虎）构建的编码器，可能对需要在高速下工作的导航特别有效。在一些实施方式中，可以使用基于多个物种的编码器，并将结果结合，以提供有利的协同作用（例如，使用基于鸟的编码器用于基本飞行导航任务，同时，当飞行过程中遇到目标物体时，使用基于猴的编码器来完成物体识别任务）。

[0195] 同样，该方法可以推广到从更高视觉区域构建的编码器，更高视觉区域例如外侧膝状体核、上丘，或视觉皮层。假体应用描述了对视网膜细胞的编码器的构建；可以使用包括数学形式体系的相同方法（同样有完整详细的描述）来获得对于更高视觉区域的编码器，这样的编码器可以作为机器视觉算法的预处理步骤。

[0196] 由于以与视网膜相类似的方式工作，对于基本上任何机器视觉算法，本文所描述的发明技术可以用来作为前端处理（或过滤）。正如视网膜预处理视觉信息以供大脑使用 -- 以允许它执行大量视觉引导的活动，例如导航、对象和人脸识别、图形 - 背景辨别、捕食检测、食品与非食品检测，及其他许多功能 -- 一起形成“虚拟视网膜”的编码器可以对视

觉信息进行预处理以供大量机器算法使用。

[0197] 视网膜的功能本质上是从视觉世界中提取数量惊人的信息，并将其减少至为点，所述要点大脑所必需以用于生物生存。因为编码器非常精确地模拟视网膜的输入 / 输出关系（并为基本上任何视觉输入进行模拟，如假体应用中所示），这意味着编码器以相同的方式减少了视觉世界的信息。因此，在各种实施方式中，本文所述的技术可以为机器视觉算法提供与视网膜提供给大脑相同或接近相同的前端处理，也就是说，它具有相同的速度，效率以及定性和定量过滤。

[0198] 这样做的必然结果是，编码器也影响了机器视觉算法是什么样的，或者，可以被怎样构建。目前算法的构建是使用原始图像作为输入，或其他方式的图像预处理（例如，使用高斯滤波器的差异）。当通过如本文所述的视网膜编码处理图像，其结果是对于机器视觉算法的新型输入，即，之前从未有过的输入。在一些实施方式，这种新的输入可以允许特定类别的算法来以新的方式进行调整或优化。例如，通过一组参数对各种机器视觉算法进行分类，所述参数可以通过图像的训练集合和 / 或在完成给定任务时由算法处理的图像至少部分地被确定。当视网膜图像数据代替原始图像使用时，所得到的该算法的参数将不同于使用相应的原始图像数据所获得的那些参数。在某些情况下，这将导致该算法显示出对于给定任务更高的性能。

[0199] 在某些情况下，由于机器视觉算法一直使用模拟脊椎动物视觉系统的图像来训练，该算法可以有利地适应于获取系统的一些性能品质。例如，由于视网膜处理突出了图像某些方面的显著性，在视网膜编码数据上训练的机器视觉算法可“学会”对这些图像方面更敏感。

[0200] 上文的实施例显示了机器视觉算法的两个实例 -- 导航仪和面部识别器，在这两种情况下，当用于视网膜处理的输入时，所述算法改变其结构。这两种算法都是学习算法，其特征在于一组权重参数，并且发现，当所述算法用于视网膜图像数据时，与图像被用于原始图像数据时相比，这些参数是不同的。在视网膜处理的情况下算法性能的提高（相对于原始图像的情况）大部分或全部是由于权重参数的改变。值得注意的是，所述性能提高能推广到与训练中使用的条件不同的环境中或条件下的导航和识别任务中。这证明了，在某些实施方式中，用视网膜图像数据训练的机器视觉算法的结构，可以以在某种程度上有益并能推广到训练以外的环境和条件的方式发生根本改变。类似地，可以开发新的算法结构以使用所述新的输入数据，也就是说，不仅是当前算法的新权重或参数，而且是能更直接地匹配或者使用本文所述的新的输入数据的新算法。

[0201] 本发明的方法和装置可以处理任何类型的图像数据。例如，响应于可见光可生成图像数据，但也可以通过其他类型的电磁辐射，例如红外、紫外或跨越电磁波谱的其它波长来生成。在一些实施方式中，图像数据可以是人工的或虚拟的图像数据（例如，根据虚拟环境的模型而生成）。在一些实施方式中，人工图像数据可以与任何类型的合适的数据的可视化有关，例如包括医学成像数据（磁共振成像数据、计算机辅助断层扫描数据、地震成像数据等等）。

[0202] 图像数据可以是单个图像或多个图像；此外，图像可以是静态的，或者以时空方式变化。可以使用简单的形状（如图表），或者相对复杂的刺激（如自然场景）。此外，所述图像可能是灰度或彩色，或者灰度和彩色的组合。在一个实施方式中，刺激可以包括白噪声

(“WN”)和/或天然刺激(“NS”),例如自然场景的影片,或者是白噪声和天然刺激的组合。

[0203]　本发明的范围并不限于上文所具体显示和描述的内容。本领域技术人员将认识到,所描述的材料、配置、结构和尺寸的例子具有合适的替代方案。在本发明的说明书和本发明的参考文献列表中引用和讨论了多种参考文献,包括专利和多种出版物。对这些参考文献的引用和讨论仅仅是为了使本发明的描述更清楚,而不是承认任何引用是本文所描述的本发明的现有技术。在本说明书中引用和讨论的所有参考文献均通过引用而整体并入本文。

[0204]　尽管在本文中描述和说明了发明的多种实施方式,但是本领域技术人员将容易想到用于执行本文所描述的功能和/或获得结果和/或一种或多种益处的多种其他方法和/或结构,每个这样的变化和/或改变均视为在本文所描述的发明实施方式的范围内。更概括地说,本领域技术人员将容易理解本文所描述的所有参数、尺寸、材料和配置均是示例性的,实际参数、尺寸、材料和/或配置将依赖于发明教导使用的特定应用。本领域技术人员将认识到,或者能够使用不超过常规实验所能确定的本文所述发明的具体实施方式的多种等效替换。因此,应当理解,前述实施方式仅以示例的方式存在,在本发明的权利要求和等效替换的范围内,可以以不同于具体描述和要求的方式来实施本发明的实施例。本文的发明实施方式针对本文所述的各个单独的特征、系统、物品、材料、试剂盒和/或方法。此外,两种或多种此类特征、系统、物品、材料、试剂盒和/或方法的组合包含在本公开的发明范围内,只要此类特征、系统、物品、材料、试剂盒和/或方法之间不相互矛盾即可。可以多种方式中的任意一种来实施上文所述的实施方式。例如,可以使用硬件、软件或其组合来实施所述的实施方式。当在软件中实施时,可以在任意合适的处理器或处理器集合中执行软件代码,无论该代码在单一计算机中,还是分布在多台计算机中。

[0205]　而且,应当理解,计算机可能包括多种形式中的任意一种,如安装在机架上的计算机、台式计算机、膝上型计算机或平板计算机。此外,还可以将计算机嵌入一般不被当做计算机但具有适宜处理能力的设备,包括个人数字助理(PDA)、智能电话或任意其他适宜的便携式或固定式电子设备。

[0206]　此外,计算机可以具有一个或多个输入和输出设备。此外,可以使用这些设备以显示用户接口。能够用来提供用户接口的输出设备的例子包括用于输出视觉显示的打印机或显示屏和用于输出可收听展示的扬声器或其他声音生成设备。能够用作用户接口的输入设备的例子包括键盘和点击设备,如鼠标、触摸板和数字化输入板。作为另一个例子,计算机可以通过语音识别或其他音响设备接收输入信息。

[0207]　此类计算机可以通过一种或多种任意形式的网络相互连接,包括局域网或广域网,如企业网络,以及智能网(IN)或因特网。此类网络可以基于任意适宜的技术,可以根据任意适宜的协议操作,并且可能包括无线网络、有线网络或光线网络。

[0208]　用于实现本文所描述功能的至少一部分的计算机可以包括存储器、一个或多个处理单元(在本文中也简称为“处理器”)、一个或多个通信接口、一个或多个显示单元、和一个或多个用户输入设备。存储器可以包括任意计算机可读的介质,并且可以存储用于执行本文所描述的各种功能的计算机指令(在本文中也称为“处理器可执行指令”)。可以使用处理单元执行指令。可以将通信接口连接到有线或无线网络、总线或其他通信装置,并可以因此允许计算机发送信息和/或接受来自其他设备的信息。可以提供显示器单元,例如,以

允许用户查看与指令执行相关的各种信息。可以提供用户输入设备，例如，以允许用户进行手动调整、进行选择、输入数据或多种其他信息，和／或在指令执行过程中以多种方式中的任意一种与处理器交互作用。

[0209] 本文列出的各种方法或过程可以被编码为软件，其可以在采用多种操作系统或平台的任意一种的一个或多个处理器上执行。此外，此类软件可以使用多种合适的编程语言，和／或编程或脚本工具中的任意一种编写，并且也可以被编译成可执行的机器语言代码或在框架或虚拟机上执行的中间代码。

[0210] 在这方面，各种发明的概念可以被实施为被一个或多个程序编码的计算机可读存储介质（或多个计算机可读存储介质）（例如，计算机存储器、一个或多个软盘、压缩盘、光盘、磁带、闪存、现场可编程门阵列或其他半导体设备中的电路配置、或其他非临时性介质或有形计算机存储介质），当在一个或多个计算机或其他处理器上执行时，执行实施本发明上文中讨论的多种实施方式。计算机可读介质或媒体可以是可传输的，这样存储在其上的程序可以被加载至一个或多个不同的计算机或其他处理器上，以实施本发明上文中讨论的各个方面。

[0211] 本文所使用的术语“程序”或“软件”在一般意义上指任意类型的计算机代码或计算机可执行指令集，可以将其引入编程计算机或其他处理器以实施上文中讨论的实施方式的各个方面。此外，应当认识到，根据一个方面，一个或多个计算机程序当执行本发明的方法时不需要驻留在一台计算机或处理器上，也可以以模块化的方式分布在多台不同计算机或处理器上，以实施本发明的各个方面。

[0212] 计算机可执行指令可以是多种形式，如程序模块，由一个或多个计算机或其他设备执行。通常地，程序模块包括例程、程序、对象、组件、数据结构等，其执行特定任务或实施特定的抽象数据类型。典型地，可以根据不同实施方式的需要对程序模块的功能进行组合或分布。

[0213] 此外，可以将数据结构存储在任意适当形式的计算机可读介质中。为了简化说明，可以将数据结构显示为具有字段，其由于在数据结构中的位置关联。类似地，这些关系可以通过为在计算机可读介质中具有位置的字段分配存储来实现，其传达了字段间的关系。然而，可以使用任意合适的机制来建立数据结构字段信息之间的关系，包括通过使用在数据元素之间建立关系的指针、标签或其他机制。

[0214] 此外，各种发明构思可以被实施为一种或多种方法，已提供了这样的例子。作为所述方法的一部分，执行的动作可以以任意适当的方式排序。因此，可以构建实施方式，其中行为的执行顺序与已说明的不同，在示例性实施方式中可以包括同时执行某些行为，甚至通过示例性实施例中所述的顺序行为实施。

[0215] 如本文所用的，自然场景应当被理解为是指自然的环境，例如在 Geisler WS 所著的 Visual perception and the statistical of properties of natural scenes. Annu. Rev. Psychol.

[0216] 59:167-92(2008) 中所述。在一些实施方式中，自然场景可以被替换为任何合适的复杂图像，例如，特征在于基本上符合频率的平方反比定律的空间和／或时间频率功率谱的图像。在一些实施方式中，例如，其中使用视频短片的实施方式，复杂图像的光谱可以在某种程度上偏离平方反比定律。例如，在一些实施方式中，复杂的图像可能具有 1/f^x 形

式的空间或时间功率谱，其中 f 是频率，x 在，例如 1-3，或其任何其子范围（例如 1.5-2.5、1.75-2.25、1.9-2.1 等）的范围内。

[0217] 白噪声图像指空间频率功率谱基本上平坦的噪声图像。

[0218] 如本文所使用的，术语“光”以及相关术语（例如，“光学”、“视觉”）应当被理解为包括可见光谱以内和以外的电磁辐射，包括，例如紫外和红外辐射。

[0219] 除非明确指出相反，如本文的说明书和权利要求书中所使用的，不定冠词“一”(a) 和“一”(an)，应该理解为是指“至少一个”。

[0220] 如在本说明书和权利要求中使用的短语“或”，应当理解为是指如此结合的元素的“任一个或两者”，即在某些情况下联合存在并且在其它情况下分开出现。用“或”列出的多个元素应当以相同的方式理解，即“一个或多个”如此结合的要素。除了由“或”的从句明确指出的元素，可任选地存在其它元素，无论是否与那些明确指出的元素相关或不相关。因此，作为非限制性实例，当用开放式语言如“包括”(including) 联合使用时，提及“A 或 B”可以指的是，在一个实施例中，只有 A（任选地包括除 B 之外的元素）；在另一个实施方式中，只有 B（任选地包括除 A 之外的元素）；在又一个实施方式中，同时指 A 和 B（任选地包括其它元素）等。

[0221] 如在本说明书和权利要求中所使用的，术语“或”应当被理解为与上文所定义的“或”具有相同的含义。例如，当将列表中的项目分开时，“或”或“或者”应当被解释为包括的，即包括至少一项，但也包括多个或一系列元素，并且任选地，其他未列出的项目中的至少一项。只有当术语明确指示相反含义时，如“只有…中的一项”或“正好…中的一项”，或者在权利要求中使用时，“由 ...... 组成”是指正好包括多个或一系列元素中的一项。通常，当冠以排他性的项目，如“任一”、“其中之一”、“中的仅一项”或“正好一项”，本文所用的术语“或”应当仅被解释为，表示排他性替代方案（即“一个或另一个但不包括两者”）。权利要求书中使用的“基本由 ...... 组成”应具有如在专利法领域中所用的普通含义。

[0222] 在权利要求书中，以及在上文的说明书中，所有过渡词语如“包括”、“包含”、“带有”、“具有”、“含有”、“涉及”、“持有”、“由…组成”等，应当将被理解为开放式的，即意指包括但不限于。只有过渡性短语“由 ...... 组成”和“基本上由 ...... 组成”应当是封闭式或半封闭式的过渡性短语，如同专利审查程序的美国专利局手册中 2111.03 部分所述。

[0223] 如本文所定义和使用的，所有定义均应被理解为字典中的定义、通过引用并入的文件中的定义和 / 或所定义术语的通常含义。

[0224] 在不违背本发明的主旨和范围的情况下，本领域的普通技术人员可对本发明的描述进行变更、修改或者其它的补充说明。尽管已描述和说明了本发明的某些实施方式，但是在不违背本发明的主旨和范围的情况下，本领域技术人员可以很清楚地知道可以对其进行各种改变和修改。在上述说明书及附图中提到的客体只是说明性，而不是限制性的。

[0225] 参考文献

[0226] Ahuja A, Dorn J, Caspi A, McMahon M, Dagnelie G, Dacruz L, Stanga P, Humayun M, Greenberg R (2010) Blind subjects implanted with the Argus II retinal prosthesis are able to improve performance in a spatial-motor task. Br J Ophthalmol.

[0227] Arenkiel et al., In vivo light-induced activation of neural circuitry in transgenic mice expressing channelrhodopsin-2. Neuron(2007)54(2):205-18.

[0228] Ausubel et al., Current Protocols in Molecular Biology, John Wiley&Sons, New York, 1989Bach, M et al(2008)Visual evoked potential-based acuity assessment in normal vision,

[0229] artificially degraded vision, and in patients. Br J Ophthalmol92:396-403

[0230] Barnstable et al., Thy-1antigen:a ganglion cell specific marker in rodent retina. Neuroscience(1984)11(4):847-55.

[0231] Bi A, Cui J, Ma Y-P, Olshevskaya E, Pu M, Dizhoor AM, Pan Z-H(2006)Ectopic expression of a microbial-type rhodopsin restores visual responses in mice with photoreceptor degeneration. Neuron50:23-33.

[0232] Bomash I, Roudi Y, Nirenberg S. (2010)A virtual retina that works on a broad array of stimuli including natural scenes:A tool to simplify the problem of population coding. Society for Neuroscience. Program No. 891. 5.

[0233] Bookstein R et al(1990)Promoter deletion and loss of retinoblastoma gene expression in human prostate carcinoma. Proc. Nati. Acad. Sci. USA87(19):7762-7766

[0234] Busskamp V, et al(2010)Genetic reactivation of cone photoreceptors restores visual responses in retinitis pigmentosa. Science329:413-417.

[0235] Cai et al. (2010)Gene delivery to mitotic and postmitotic photoreceptors via compacted DNA nanoparticles results in improved phenotype in a mouse model of retinitis pigmentosa. FASEB J. 24:1178-1191.

[0236] Campagnola L, Wang H, Zylka MJ. (2008)Fiber-coupled light-emitting diode for localized photostimulation of neurons expressing channelrhodopsin-2. Journal of Neuroscience Methods. 169:27-33.

[0237] Cardin JA, et al(2010)Targeted optogenetic stimulation and recording of neurons in vivo using cell-type-specific expression of Channelrhodopsin-2. Nat Protoc5(2):247-54.

[0238] Chader GJ, Weiland J, Humayun MS(2009)Artificial vision:needs, functioning, and testing of a retinal electronic prosthesis. Prog Brain Res175:317-332.

[0239] Chiappa, K. (1997)Evoked Responses in Clinical Medicine, Third Edition, Lippincott-Raven

[0240] Chichilnisky EJ. (2001)A simple white noise analysis of neuronal light responses. Network12(2):199-213

[0241] Chopdar A, Chakravarthy U, Verma D(2003)Age related macular degeneration. BMJ326:485-488.

[0242] Cover T and Thomas J. (2006)Elements of Information Theory, 2nd Edition. Hoboken, NJ:Wiley

[0243] Dann JF, Buhl EH. (1987)Retinal ganglion cells projecting to the accessory optic system in the rat. J Comp Neurol262(1):141-58.

[0244] Dedek K, et al(2008)Ganglion cell adaptability:does the coupling of horizontal cells play a role ? PLoS One. 3(3):e1714.

[0245] Douglas RM et al(2005)Independent visual threshold measurements in the two eyes of freely moving rats and mice using a virtual-reality optokinetic system. Vis Neurosci. 22(5):677-84.

[0246] Duda RO, Hart PE(2001)Pattern Classification(2nd Edition)Wiley, NY,

[0247] Enroth-Cugell et al., (1966)The contrast sensitivity of retinal ganglion cells of the cat. J Physiol187(3):517--52.

[0248] European Patent Application No. 19891976

[0249] Famulare M, Fairhall A. (2010)Feature selection in simple neurons:how coding depends on spiking dynamics. Neural Comput22(3):581-98

[0250] Field et al., (2007)Information processing in the primate retina:circuitry and coding. Annu Rev Neurosci30:1--30.

[0251] Fitzgerald et al. (1994)Retinal signal transmission in Duchenne muscular dystrophy. J Clin Invest 93:2425-30.

[0252] Foley JM, Legge GE(1981)Contrast detection and near-threshold discrimination in human vision. Vision Res. 21(7):1041-53.

[0253] Fried S, Werblin F, McMahon MJ(2006)US Pat. 2006/0129207Mimicking neural coding in retinal ganglion cells with short pulse electrical stimulation. In:(US, ed).

[0254] Friedman DS, O'Colmain BJ, Muñoz B, Tomany SC, McCarty C, de Jong PTVM, Nemesure B, Mitchell P, Kempen J, Eye Diseases Prevalence Research Group(2004)Prevalence of age-related macular degeneration in the United States. Arch Ophthalmol122:564-572.

[0255] Geisler WS(200). Visual perception and the statistical properties of natural scenes. Annu. Rev. Psychol. 59:167-92(2008)

[0256] Gerding H, Benner FP, Taneri S(2007)Experimental implantation of epiretinal retina implants(EPI-RET)with an IOL-type receiver unit. J Neural Eng4:S38-49.

[0257] Giolli RA, Blanks RHI, Lui F. (2005)The accessory optic system:basic organization with an update on connectivity, neurochemistry, and function. Prog Brain Res151:407-40.

[0258] Golan L, Reutsky I, Farah N&Shoham S. (2009)Design and characteristics of holographic neural photo-stimulation systems, Journal of Neural Engineering6066004, (2009)

[0259] Graham-Rowe D(2009)A Brighter Future for Retinal Implants. In:Technology Review, http://www.technologyreview.com/biomedicine/23539/. Boston, MA:MIT.

[0260] Greenberg KP, et al. (2011). Differential Targeting of Optical Neuromodulators

[0261] to Ganglion Cell Soma and Dendrites Allows Dynamic Control of Center-Surround Antagonism. Neuron69, 713 - 720.

[0262] Grinstead CM and Snell JL(1997) Introduction to Probability. American Mathematical Society ;2nd Revised edition

[0263] Grossman N, Poher V, Grubb MS, Kennedy GT, Nikolic K, McGovern B, Palmini RB, Gong Z, Drakakis EM, Neil, MAA, Dawson MD, Burrone J, Degenaar P. (2010) Multi-site optical excitation using ChR2and micro-LED array. J. Neural Eng, 7(1):1-13.

[0264] Han et al, (2009), Millisecond-Timescale Optical Control of Neural Dynamics in the Nonhuman Primate Brain, Neuron62, 191-198.

[0265] Hand DJ. (1981) Discrimination and classification. Wiley Series in Probability and Mathematical Statistics.

[0266] Huberman AD, Manu M, Koch SM, Susman MW, Lutz AB, Ullian EM, Baccus SA, Barres BA(2008) Architecture and activity-mediated refinement of axonal projections from a mosaic of genetically identified retinal ganglion cells. Neuron. 2008Aug14 ;59(3):425-38.

[0267] Huberman AD, Wei W, Elstrott J, Stafford BK, Feller MB, Barres BA(2009) Genetic Identification of an On-Off Direction-Selective Retinal Ganglion Cell Subtype Reveals a Layer-Specific Subcortical Map of Posterior Motion. Neuron. 62(3):327-334.

[0268] Ivanova E, Pan Z-H(2009) Evaluation of the adeno-associated virus mediated long-term expression of channelrhodopsin-2in the mouse retina. Molecular Vision15:1680-1689

[0269] Izhikevich EM(2007) Dynamical systems in neuroscience:the geometry of excitability and bursting. Cambridge, MA:MIT Press

[0270] Izhikevich EM(2010) Hybrid spiking models. Review. Phil. Trans. R. Soc. A(2010) 368, 5061 - 5070

[0271] Jacobs AL et al(2009), Ruling out and ruling in neural codes. Proc Natl Acad Sci U S A. 106(14):5936-41.

[0272] Jeffreys, Harold(1961). The Theory of Probability. The Oxford University Press.

[0273] Kass RE, Ventura V, Brown EN. (2005) Statistical issues in the analysis of neuronal data. JNeurophysiol94(1):8-25.

[0274] Kawasaki et al., Variability of the relative afferent pupillary defect. Am J Ophthalmol(1995). 120:622 - 633.

[0275] Kay MA, Glorioso JC, Naldini L. (2001) Viral vectors for gene therapy:the art of turning infectious agents into vehicles of therapeutics. Nat Med. 7(1):33-40. Review.

[0276] Kelly S, Shire D, Chen J, Doyle P, Gingerich M, Drohan W, Theogarajan L, Cogan S, Wyatt J, Rizzo JI (2009) Realization of a15-channel, hermetically-encased wireless subretinal prosthesis for the blind. In, pp200--203.

[0277] Kibbel S, Harscher A, Wrobel W-G, Zrenner E, Rothermel A (2009) Design and Performance of an improved active subretinal chip. In: World Congress on Medical Physics and Biomedical Engineering, September7-12, 2009, Munich, Germany (Kim SI, Suh TS, Dössel O, Schlegel WC, eds), pp192-195: Springer Berlin Heidelberg.

[0278] Koilkonda RD, Hauswirth WW, Guy J. (2009) Efficient expression of self-complementary AAV in ganglion cells of the ex vivo primate retina. Mol Vis. 15:2796-802.

[0279] Kuffler SW. (1953) Discharge patterns and functional organization of mammalian retina. J Neurophysiol16(1):37-68.

[0280] Lagali PS, Balya D, Awatramani GB, Münch TA, Kim DS, Busskamp V, Cepko CL, Roska B (2008) Light-activated channels targeted to ON bipolar cells restore visual function in retinal degeneration. Nat Neurosci11:667-675.

[0281] Lesica NA et al. (2007) Adaptation to stimulus contrast and

[0282] correlations during natural visual stimulation. Neuron55(3):479-491.

[0283] Lettvin et al., (1959) What the frog's eye tells the frog's brain. Proceedings of the Institute of Radio Engineers47(11):1940-51.

[0284] Liao et al. (2007) In vivo gene delivery in the retina using polyethylenimine. BioTechniques2007, 42:285-288.

[0285] Loewenstein JI, Montezuma SR, Rizzo JF, III (2004) Outer Retinal Degeneration: An Electronic Retinal Prosthesis as a Treatment Strategy. Arch Ophthalmol122:587-596.

[0286] Maguire et al. Safety and efficacy of gene transfer for Leber's congenital amaurosis. N Engl J Med (2008) 358:2240 - 2248.

[0287] Mancuso et al., (2009) Gene therapy for red-green colour blindness in adult primates. Nature461(7265):784-7.

[0288] Martin et al. 2002. Gene delivery to the eye using adeno-associated viral vectors. Methods28:267-275.

[0289] McGowan MH et al (1998) Characterization of the Mouse Aldose Reductase Gene and Promoter in a Lens Epithelial Cell Line. Mol Vis1998;4:2

[0290] McLaughlin SK, Collis P, Hermonat PL, Muzyczka N. (1988) Adeno-associated virus general transduction vectors: analysis of proviral structures. J Virol. 62(6):1963-73.

[0291] Meytlis M, Bomash I, Pillow JW, Nirenberg S. (2009) Assessing the importance of correlated firing using large populations of neurons. Society for Neuroscience. Program No. 165.3.

[0292] Morgans CW, et al (2009) TRPM1is required for the depolarizing light response in retinal ON-bipolar cells. Proc Natl Acad Sci U S A106(45):19174-8.

[0293] Nanduri D, Humayun M, Greenberg R, McMahon M, Weiland J(2008) Retinal prosthesis phosphene shape analysis. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp1785--1788. Vancouver, BC.

[0294] Nichols Z, Meytlis M, Nirenberg S. (2010) Correlations play a negligible role in coding white noise and natural scene stimuli in complete retinal populations. Submitted.

[0295] Nirenberg S(2000) Photoablation of cells expressing beta-galactosidase. Methods Mol Biol. 135:475-80

[0296] Nirenberg S and Cepko, C(1993). Targeted ablation of diverse cell classes in the nervous system in vivo. J Neurosci. 13(8):3238-51.

[0297] Nirenberg S and Latham PE. (1998) Population coding in the retina. Curr. Opin. Neurobiol. 8(4):488-493

[0298] Nirenberg S and Meister M. (1997). The light response of retinal ganglion cells is truncated by a displaced amacrine circuit. Neuron18:637-650

[0299] Nirenberg S et al(2001) Retinal ganglion cells act largely as independent encoders. Nature 411(6838):698-701.

[0300] Nirenberg S et al. (2010) Heterogeneous response dynamics in retinal ganglion cells: the interplay of predictive coding and adaptation. J Neurophysiol103(6):3184-94

[0301] Norcia, AM, and Tyler, CW(1985) Spatial frequency sweep VEP: visual acuity during the first year of life. Vision Res. 25(10):1399-408

[0302] Norcia, AM, et al(1989). Measurement of spatial contrast sensitivity with the swept contrast VEP. Vision Res. 1989 ;29(5):627-37.

[0303] Okuyama et al. (1993). Binocular infrared optometer for measuring accommodation in both eyes simultaneously in natural-viewing conditions Applied Optics, Vol. 32. No22, p. 4147

[0304] Pandarinath et al(2010a) A novel mechanism for switching a neural system from one state to another. Front Comput Neurosci. 31 ;4:2.

[0305] Pandarinath et al(2010b) Symmetry breakdown in the ON and OFF pathways of the retina at night: functional implications. J Neurosci30(30):10006-14.

[0306] Paninski L, Pillow J, Lewi J. (2007) Statistical models for neural encoding, decoding, and optimal stimulus design. Prog Brain Res. 165:493-507.

[0307] Paninski L. (2004) Maximum likelihood estimation of cascade point-process neural encoding models. Network15(4):243-62

[0308] Panzeri S, et al(2007) Correcting for the sampling bias problem in spike train information measures. J Neurophysiol. 98(3):1064-72. Review.

[0309] Pelli DG, Robson JG, &Wilkins AJ(1988) The design of a new letter chart for measuring contrast sensitivity. Clinical Vision Sciences2, 187-199

[0310] Perry VH, Silveira LC. (1988) Functional lamination in the ganglion cell layer of the macaque's retina. Neuroscience. 25(1):217-23.

[0311] Petrs-Silva et al., (2009) High-efficiency transduction of the mouse retina by tyrosine-mutant AAV serotype vectors. Mol Ther17(3):463-71.

[0312] Petersen-Jones et al., (2009) AAV retinal transduction in a large animal model species: comparison of a self-complementary AAV2/5with a single-stranded AAV2/5vector. Mol Vis15:1835-42.

[0313] Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP. (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature454(7207):995-9

[0314] Prusky GT, et al (2004) Rapid quantification of adult and developing mouse spatial vision using a virtual optomotor system. Invest Ophthalmol Vis Sci. 45(12):4611-6.

[0315] Published PCT Application WO1996013598

[0316] Published PCT application WO1998048027

[0317] Published PCT Application WO2000015822

[0318] Published PCT application WO2001094605

[0319] Published PCT application WO2002082904

[0320] Published PCT application WO2003047525

[0321] Published PCT application WO2003080648

[0322] Published PCT application WO2003093479

[0323] Published PCT application WO2003104413

[0324] Published PCT application WO2005080573

[0325] Published PCT application WO2007127428

[0326] Published PCT application WO2010011404

[0327] Pun L(1969), Introduction to Optimization Practice, ISBN471-70233-1

[0328] Purpura K, Tranchina D, Kaplan E, Shapley RM. (1990) Light adaptation in the primate retina: analysis of changes in gain and dynamics of monkey retinal ganglion cells. Vis Neurosci4(1):75--93.

[0329] Rolls ET, Baylis GC, Leonard CM. Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey. Vision Res. 1985 ;25(8):1021-35.

[0330] Sambrook et al, "Molecular Cloning. A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory, New York(1989)

[0331] Sauer B. (1987) Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. Mol Cell Biol. 7(6):2087-96.

[0332] Shapley RM, Victor JD. (1981) How the contrast gain control modifies the frequency responses of cat retinal ganglion cells. J Physiol. 318:161-79.

[0333] Sharpee TO et al. (2008) On the Importance of Static Nonlinearity in Estimating Spatiotemporal Neural Filters With Natural Stimuli. J Neurophysiol99(5):2496-509

[0334] Sheridan C(2011) Gene Therapy finds its niche Nature Biotechnology29(2):121-128

[0335] Siegert S, Scherf BG, Punta KD, Didkovsky N, Heintz N, Roska B(2009). Genetic address book for retinal cell types. Nature Neuroscience. 12:1197-1204.

[0336] Simoncelli et al. (2004) Characterization of neural responses with stochastic stimuli. The cognitive neurosciences:327—38

[0337] Simonelli et al. (2010) Gene Therapy for Leber's Congenital Amaurosis is Safe and Effective Through1.5Years After Vector Administration, Molecular Therapy183, 643 - 650.

[0338] Sinclair JR, et al (2004). Selective ablation of a class of amacrine cells alters spatial processing in the retina. J Neurosci. 24(6):1459-67.

[0339] Sjostrand et al. (1999). Morphometric study of the displacement of retinal ganglion cells subserving cones within the human fovea. Graefe's Arch Clin Exp Ophthalmol 237:1014 - 1023.

[0340] Soucy ER et al. (1998) A novel signaling pathway from rod photoreceptors to ganglion cells in mammalian retina. Neuron21:481-493

[0341] Stone et al., (1993) Response properties of ganglion cells in the isolated mouse retina. Vis Neurosci10(1):31-9.

[0342] Strong SP, et al (1998) On the application of information theory to neural spike trains. Pac Symp Biocomput. 621-32.

[0343] Thyagarajan S, van Wyk M, Lehmann K, Löwel S, Feng G, Wässle H(2010) Visual function in mice with photoreceptor degeneration and transgenic expression of channelrhodopsin2in ganglion cells. J Neurosci30:8745-8758.

[0344] Tomita H, Sugano E, Isago H, Hiroi T, Wang Z, Ohta E, Tamai M(2010) Channelrhodopsin-2gene transduced into retinal ganglion cells restores functional vision in genetically blind rats. Exp Eye Res90:429-436.

[0345] Troy JB, Bohnsack DL, Chen J, Guo X, Passaglia CL. (2005) Spatiotemporal integration of light by the cat X-cell center under photopic and scotopic conditions. Vis Neurosci22(4):493--500.

[0346] Troy JB, Bohnsack DL, Diller LC. (1999) Spatial properties of the cat X-cell receptive field as a function of mean light level. Vis Neurosci16(6):1089--104.

[0347] Turchinovich et al. (2010) Non-viral siRNA delivery into the mouse retina in vivo. BMC Ophthalmology10:25.

[0348] U.S. Patent7149586

[0349] U. S. Patent5, 856, 152

[0350] U. S. Patent5, 871, 982

[0351] U. S. Patent6533798

[0352] U. S. Patent Publication20080221653

[0353] U. S. Patent Publication20080249588

[0354] U. S. Patent Publication20090088399

[0355] U. S. Patent Publication20090326623,

[0356] U. S. Patent Publication20100272688

[0357] U. S. Patent Publication Nos. 20070261127

[0358] Ueda et al, (1997)The mGluR65'upstream transgene sequence directs a cell-specific and developmentally regulated expression in retinal rod and ON-type cone bipolar cells. J Neurosci. 17(9):3014-23.

[0359] van Adel et al. (2003)Delivery of ciliary neurotrophic factor via lentiviral-mediated transfer protects axotomized retinal ganglion cells for an extended period of time. Hum. Gene Ther. 14:103-115.

[0360] Victor JD, Shapley RM. (1979)The nonlinear pathway of Y ganglion cells in the cat retina. J Gen Physiol. 74(6):671-89.

[0361] Victor JD. (1987)The dynamics of the cat retinal X cell centre. The Journal of Physiology386(1):219.

[0362] Volgyi B, Deans MR, Paul DL, Bloomfield SA(2004)Convergence and Segregation of the Multiple Rod Pathways in Mammalian Retina. J Neurosci24(49):11182 - 11192.

[0363] Walther W, Stein U. (2000)Viral vectors for gene transfer:a review of their use in the treatment of human diseases. Drugs. 60(2):249-71. Review.

[0364] Wassle H. (2004)Parallel processing in the mammalian retina. Nat Rev Neurosci5(10):747-57. Wells et al. (2005)Optical stimulation of neural tissue in vivo. Optics Letters30(5):504-506, Winter JO, Cogan SF, Rizzo JFI(2007)Retinal prostheses:current challenges and future outlook.

[0365] J Biomater Sci Polym Ed18:1031-1055.

[0366] Wright AF. (1997)Gene therapy for the eye. Br J Ophthalmol81(8):620-623Review.

[0367] Yonehara K, Ishikane H, Sakuta H, Shintani T, Nakamura-Yonehara K, et al. (2009)Identification of Retinal Ganglion Cells and Their Projections Involved in Central Transmission of Information about Upward and Downward Image Motion. PLoS ONE4(1):e4320.

[0368] Yonehara K, Shintani T, Suzuki R, Sakuta H, Takeuchi Y, et al. (2008) Expression of SPIG1 Reveals Development of a Retinal Ganglion Cell Subtype Projecting to the Medial Terminal Nucleus in the Mouse. PLoS ONE3(2):e1533.

[0369] Zhang Y, Ivanova E, Bi A, Pan Z-H(2009)Ectopic expression of multiple microbial rhodopsins restores ON and OFF light responses in retinas with photoreceptor degeneration. J Neurosci29:9186-9196.

[0370] Zrenner E, et al (2009) Subretinal Microelectrode Arrays Allow Blind Retinitis Pigmentosa Patients to Recognize Letters and Combine them to Words. BMEI'09. 2nd International Conference on Biomedical Engineering and Informatics. Issue Date:17-19Oct. 2009. ISBN:978-1-4244-4132-7. Pages1-4.

[0371] Ballard DH et al. (1982) Computer Vision, Prentice-Hall Inc New Jersey, 1982

[0372] Chichilnisky EJ. (2001) A simple white noise analysis of neuronal light responses. Network12(2):199-213

[0373] Forsyth DA et al. (2003) Computer Vision:A Modern Approach, Prentice Hall, Nirenberg S et al. (2010) Heterogeneous response dynamics in retinal ganglion cells: the interplay of predictive coding and adaptation. J Neurophysiol103(6):3184-94.

[0374] Freund Y et al. (1995) A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting Journal of Computer and System Sciences55, 119-139

[0375] Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP. (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature454(7207):995-9

[0376] Simonelli et al. (2010) Gene Therapy for Leber's Congenital Amaurosis is Safe and Effective Through1.5Years After Vector Administration, Molecular Therapy183, 643 - 650

[0377] Nichols Z, Meytlis M, Nirenberg S. (2010) Correlations play a negligible role in coding white noise and natural scene stimuli in complete retinal populations. Submitted.

[0378] Nirenberg S et al. (2010) Heterogeneous response dynamics in retinal ganglion cells: the interplay of predictive coding and adaptation. J Neurophysiol103(6):3184-94。

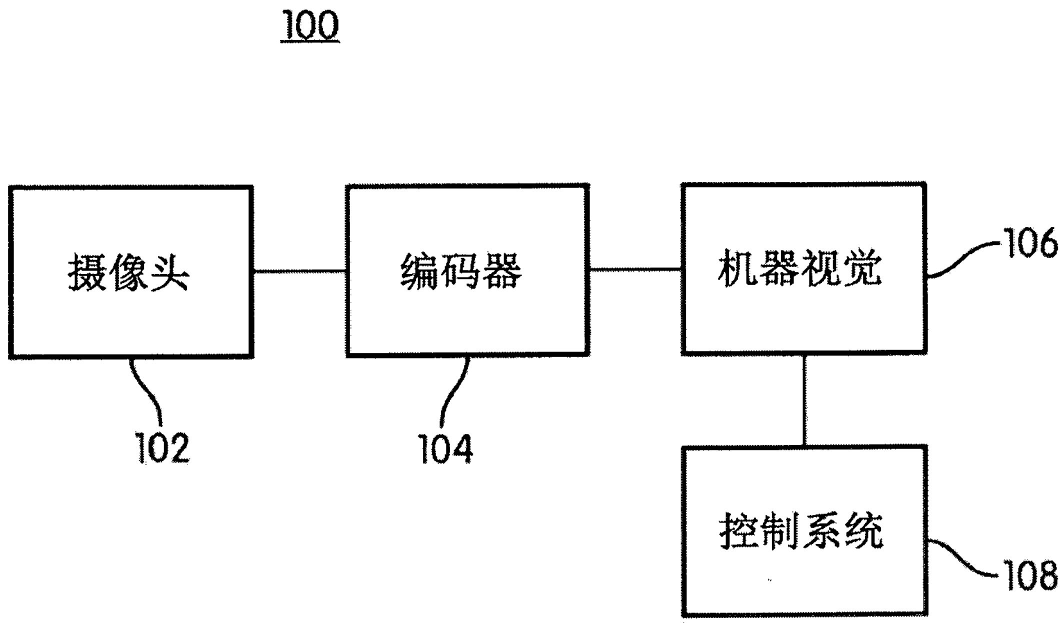

100
摄像头
编码器
机器视觉
控制系统
102
104
106
108

图 1

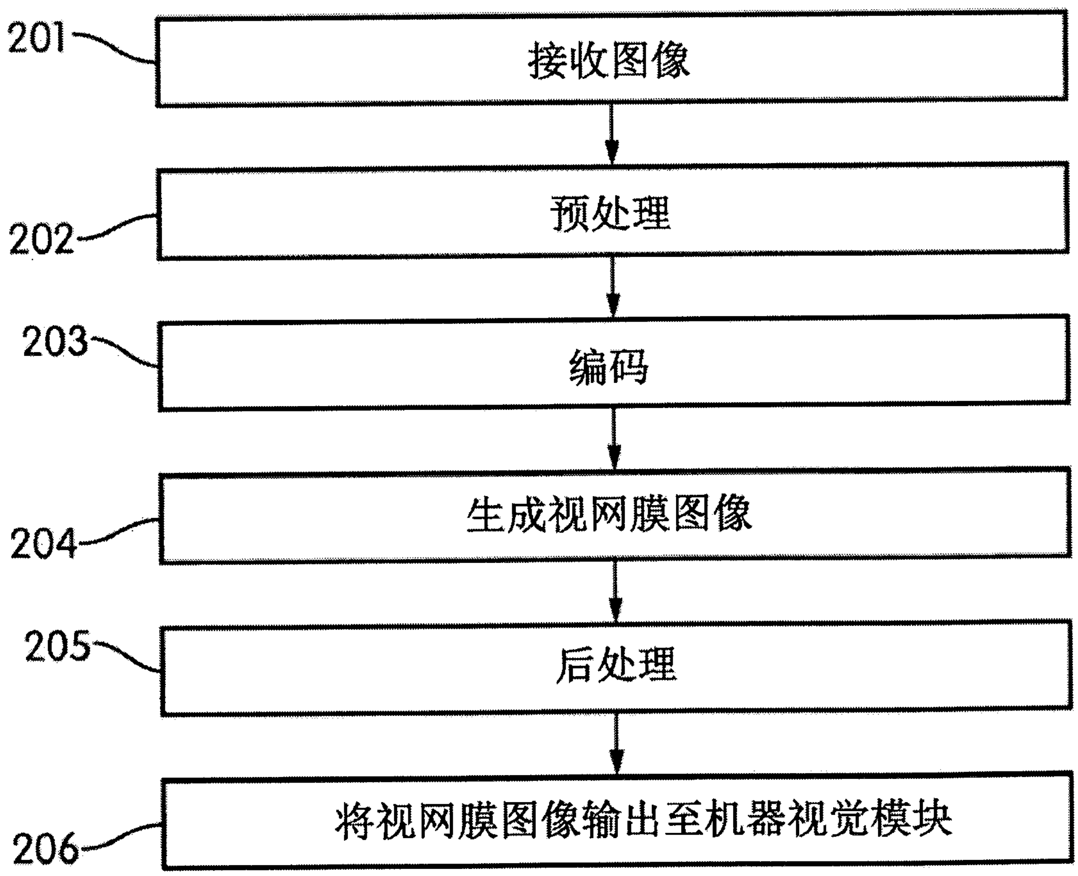

201
接收图像
202
预处理
203
编码
204
生成视网膜图像
205
后处理
206
将视网膜图像输出至机器视觉模块

图 2

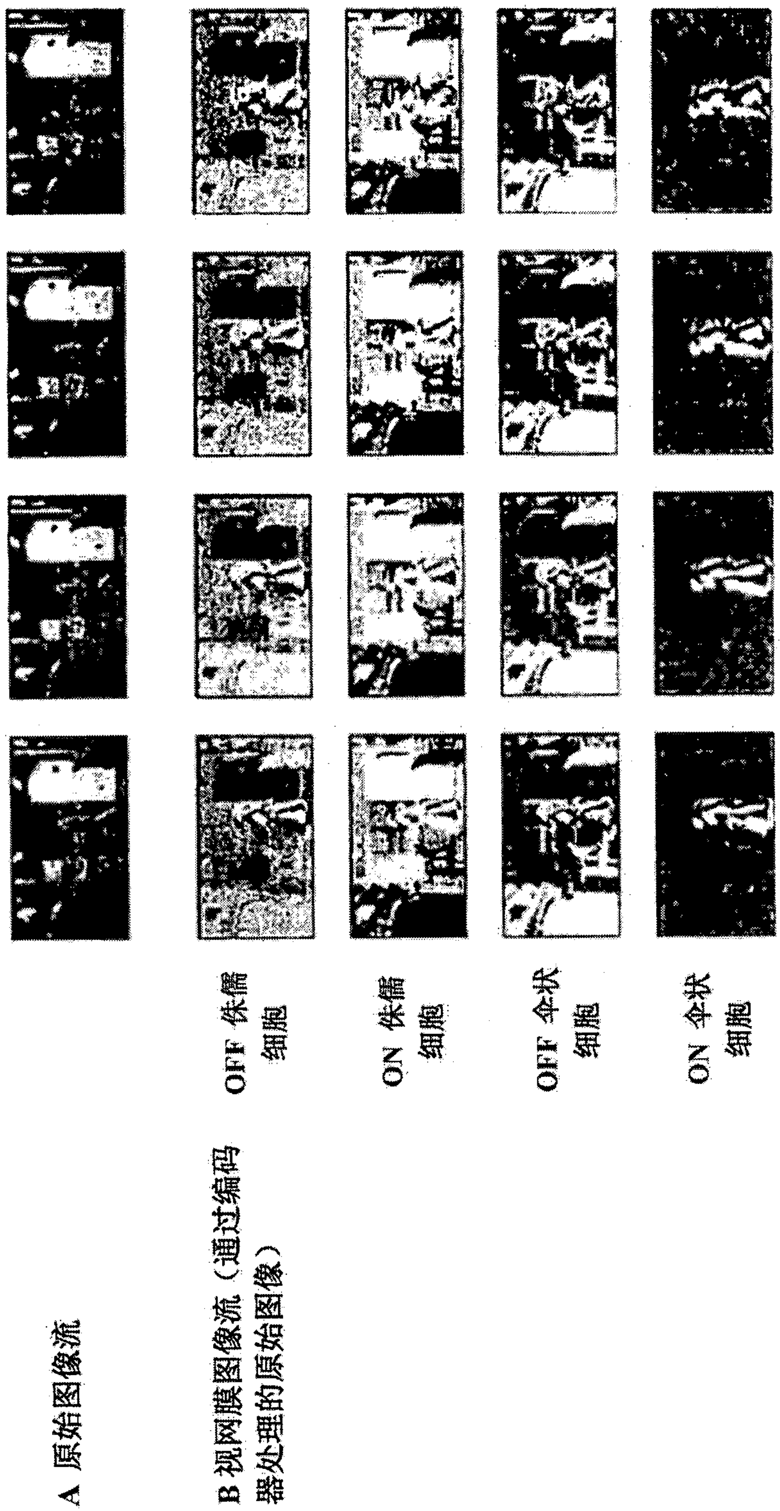

A 原始图像流
B 视网膜图像流（通过编码器处理的原始图像）
OFF 侏儒细胞
ON 侏儒细胞
OFF 伞状细胞
ON 伞状细胞

图 3A

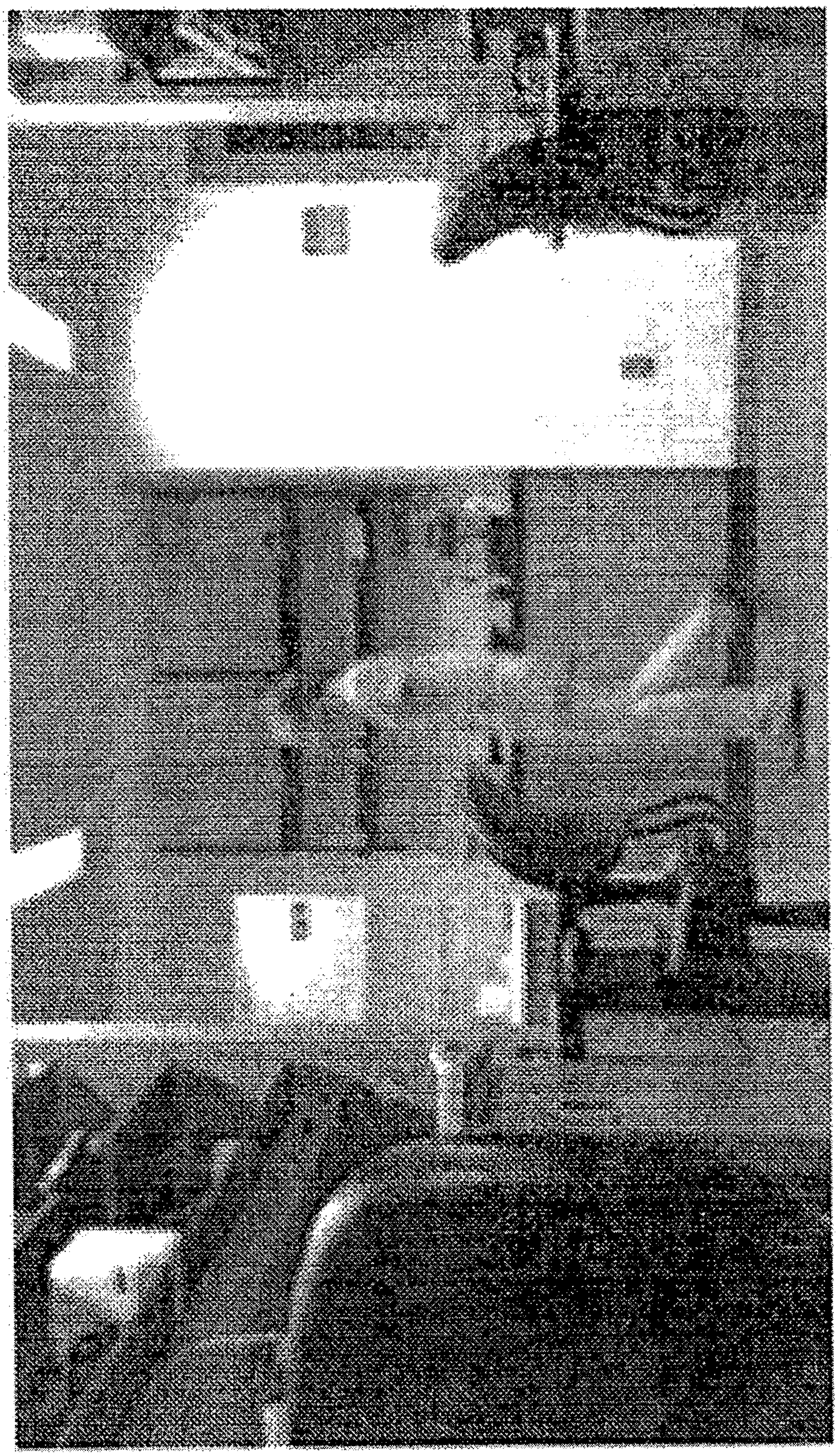


图 3B

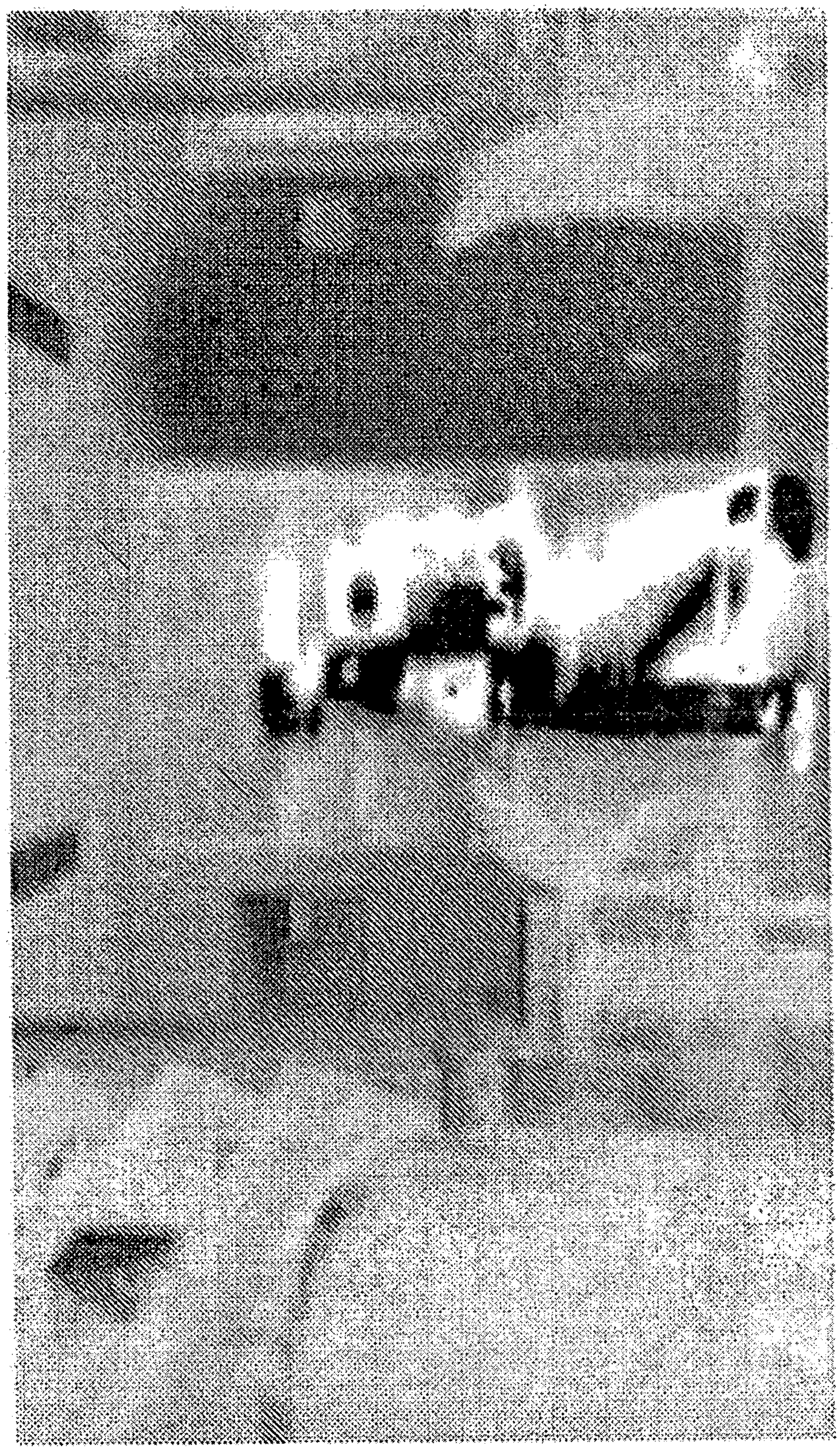


图 3c

图 3D

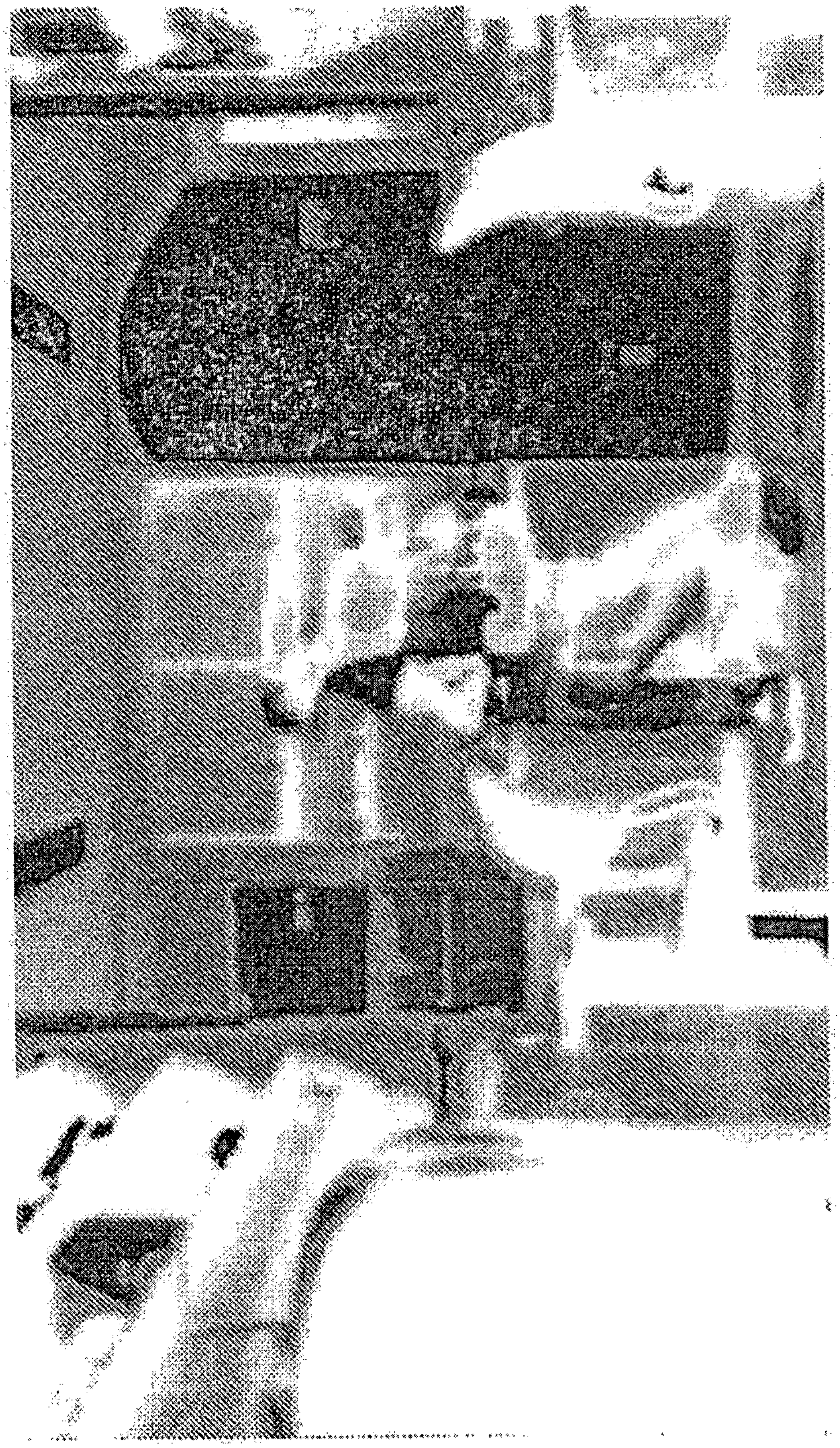


图 3E

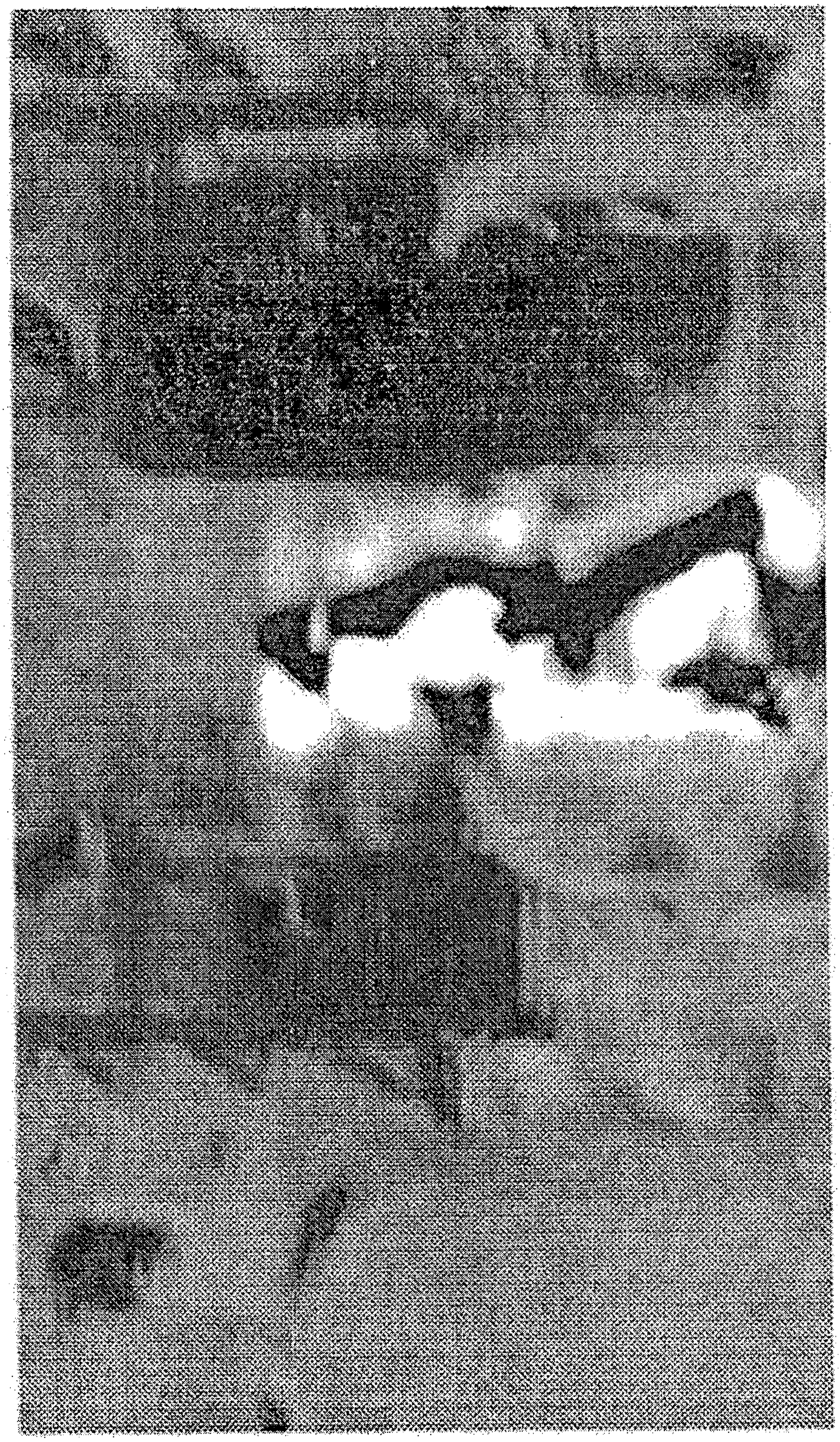


图 3F

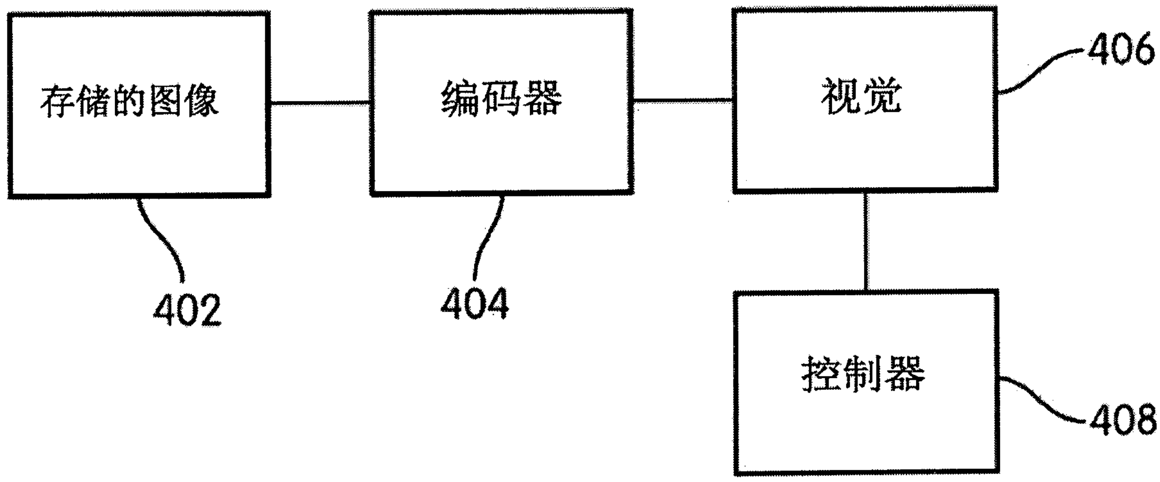

存储的图像
编码器
视觉
406
402
404
控制器
408

图 4

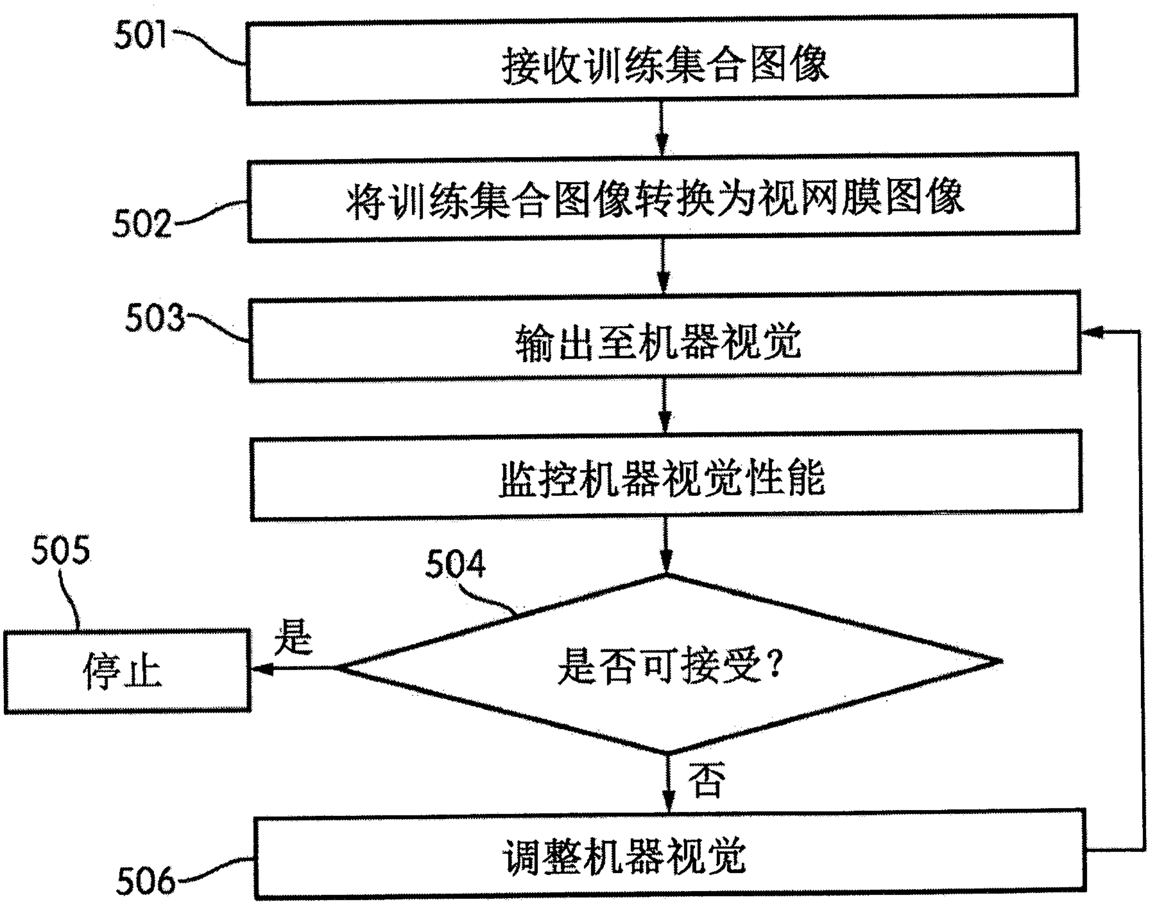

501
接收训练集合图像
将训练集合图像转换为视网膜图像
502
503
输出至机器视觉
监控机器视觉性能
505
504
是
停止
是否可接受?
否
调整机器视觉
506

图 5

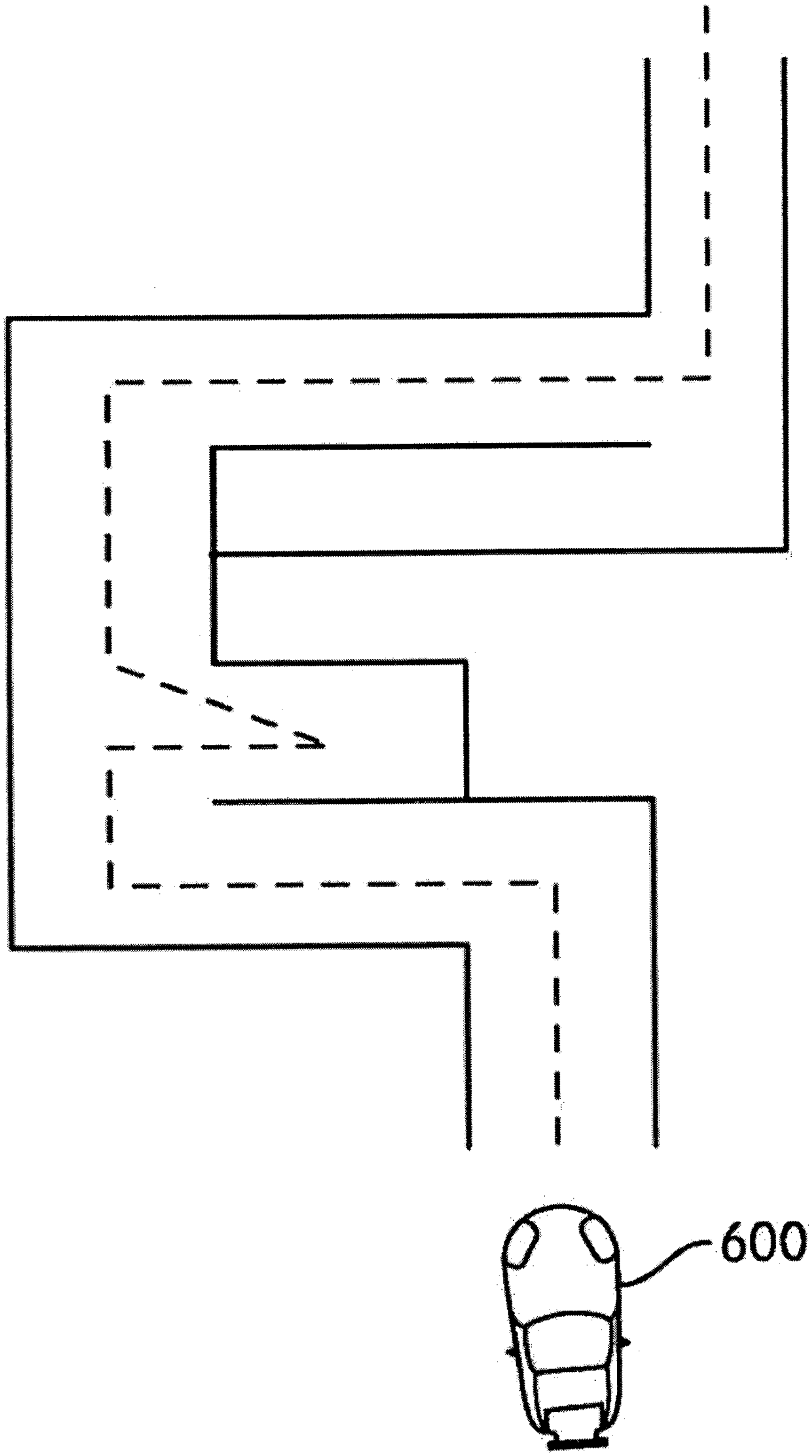


图 6

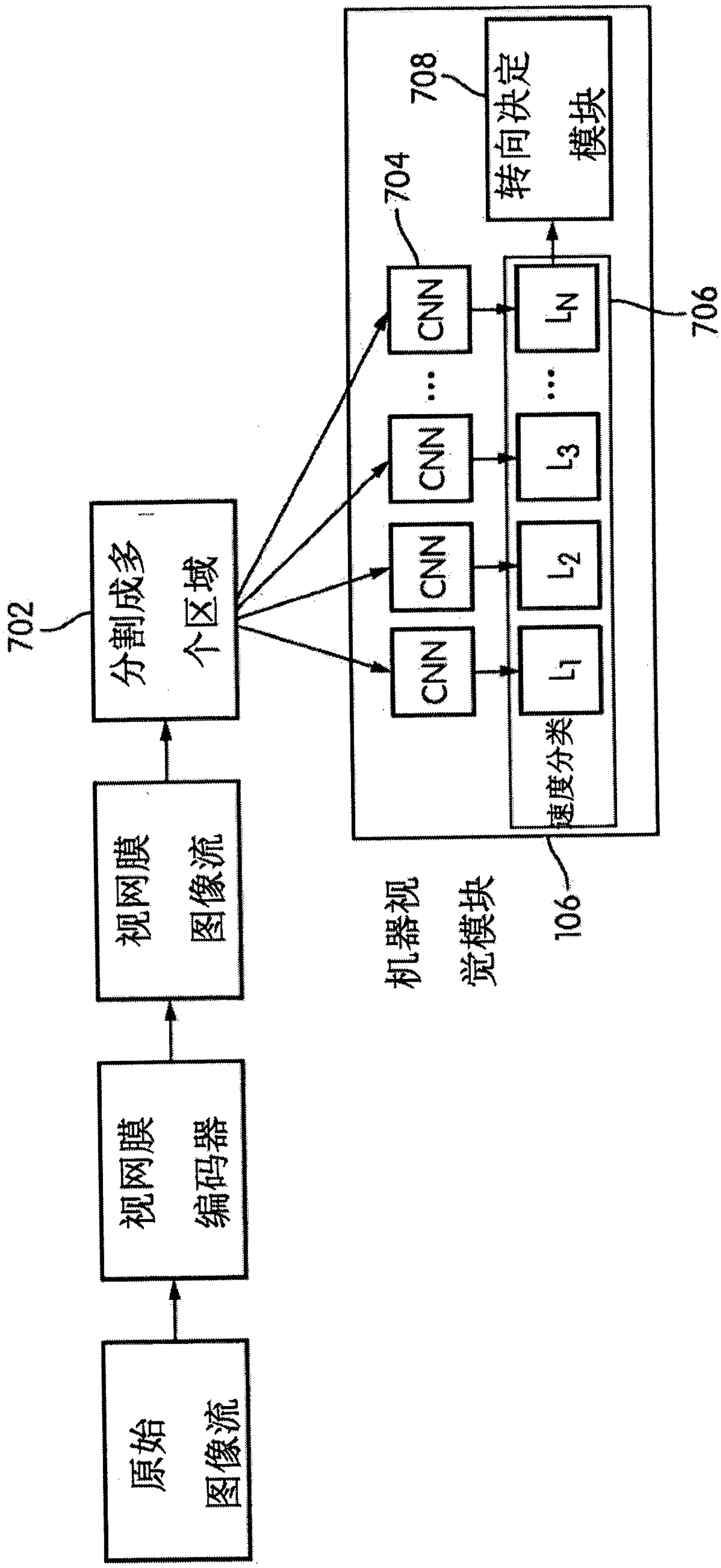

原始图像流
视网膜编码器
视网膜图像流
分割成多个区域
702
CNN
CNN
CNN
CNN
704
L_1
L_2
L_3
L_N
速度分类
706
转向决定模块
708
机器视觉模块
106

图 7

前 5 帧或训练序列

取样更稀疏以覆盖影片其余部分的后续帧

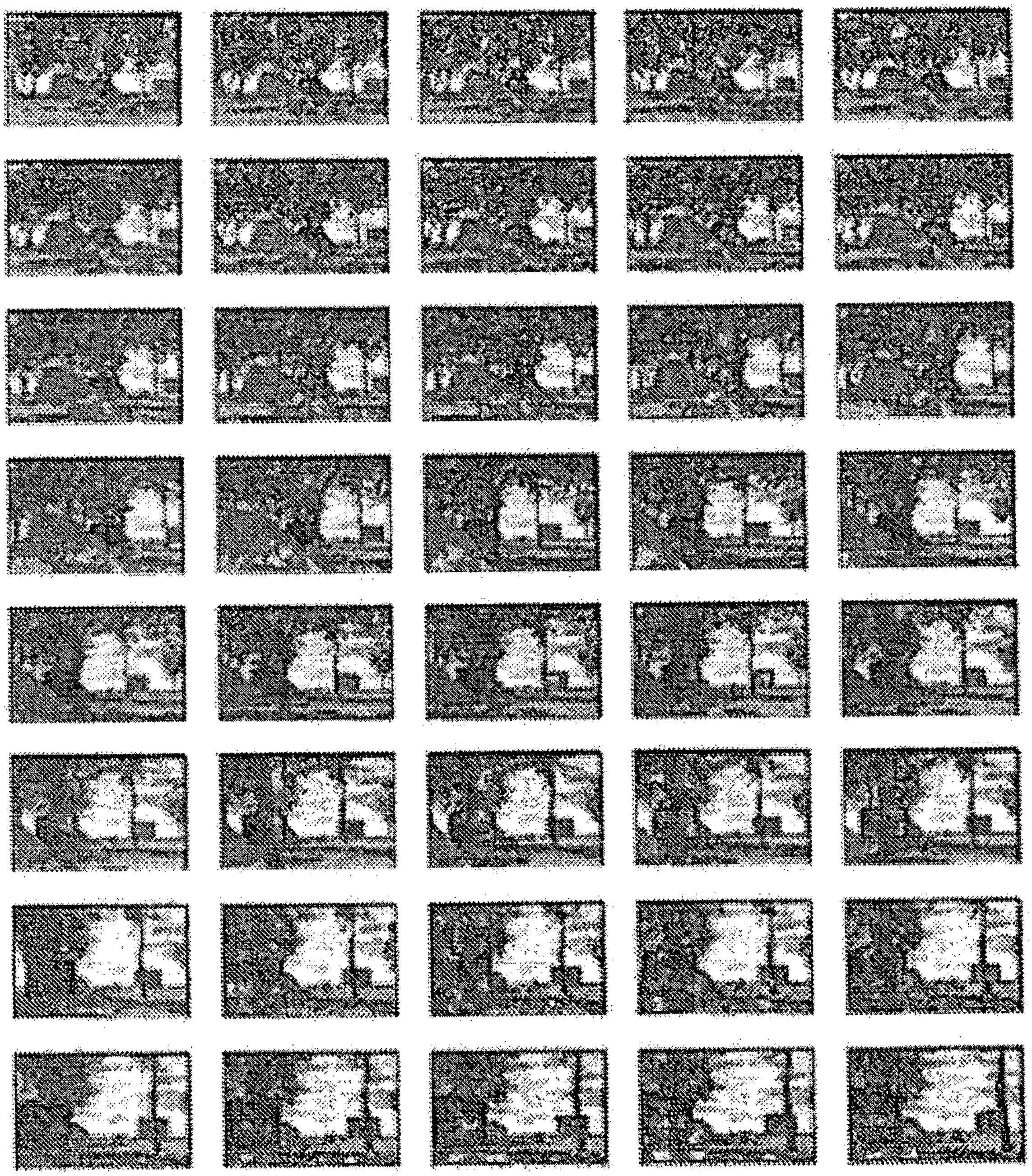
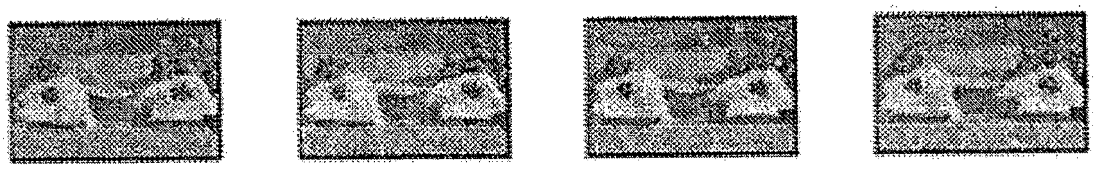


图 8

测试集 1：田园序列的前 4 帧（有意与训练序列不同）



取样更稀疏以覆盖影片其余部分的后续帧

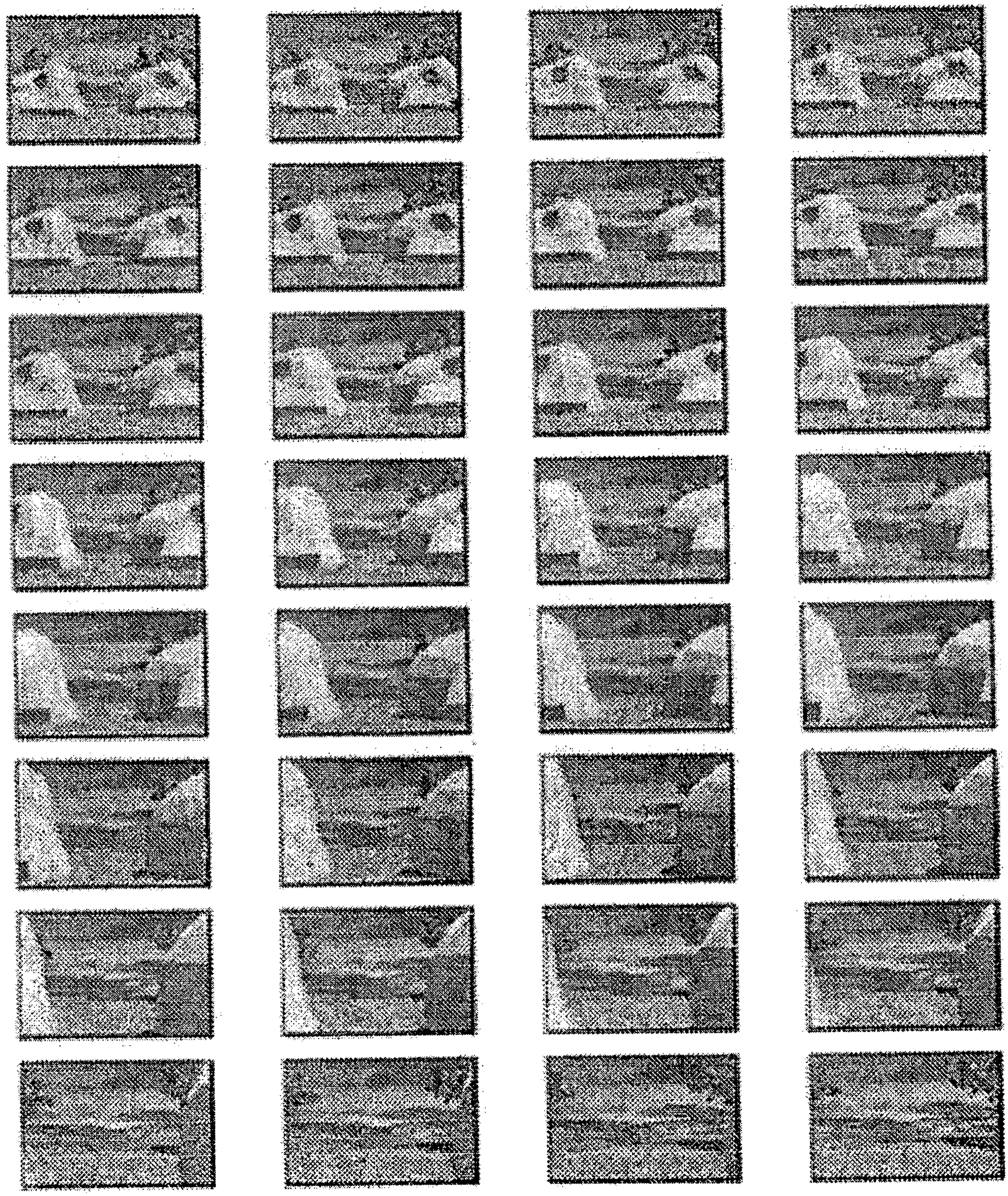
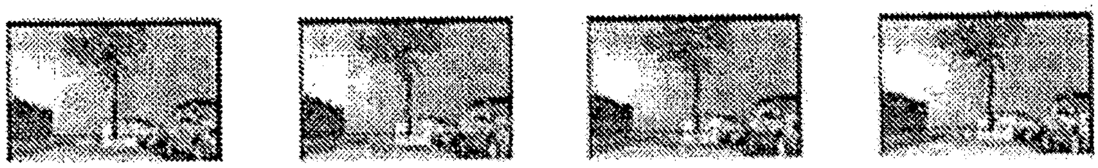


图 9A

测试集 2：郊区序列的前 4 帧



取样更稀疏以覆盖影片其余部分的后续帧

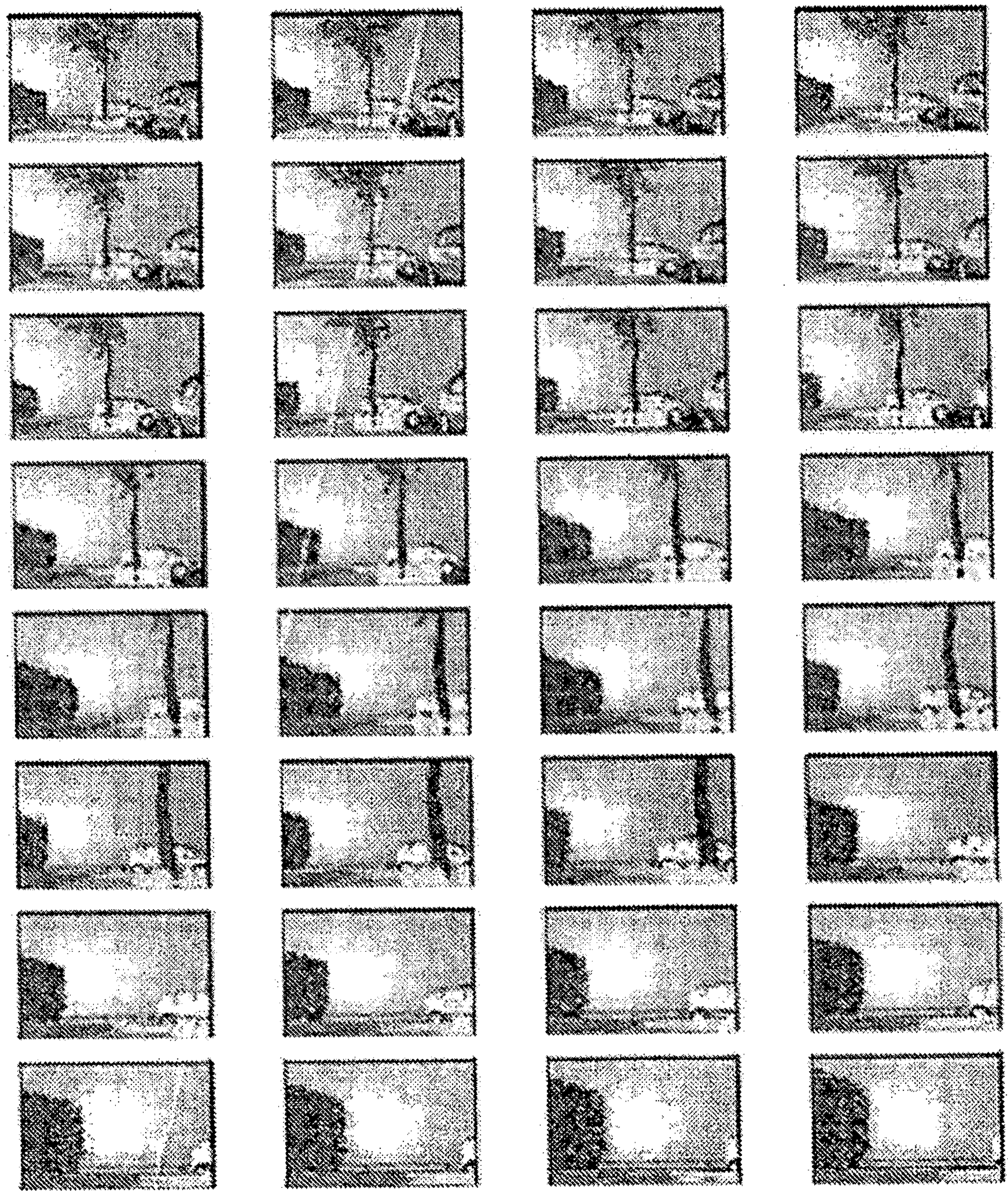
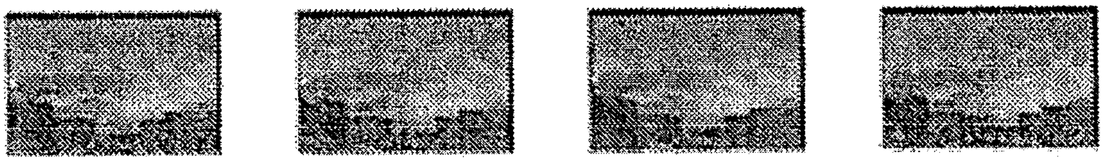


图 9B

测试集 3：广场序列的前 4 帧



取样更稀疏以覆盖影片其余部分的后续帧

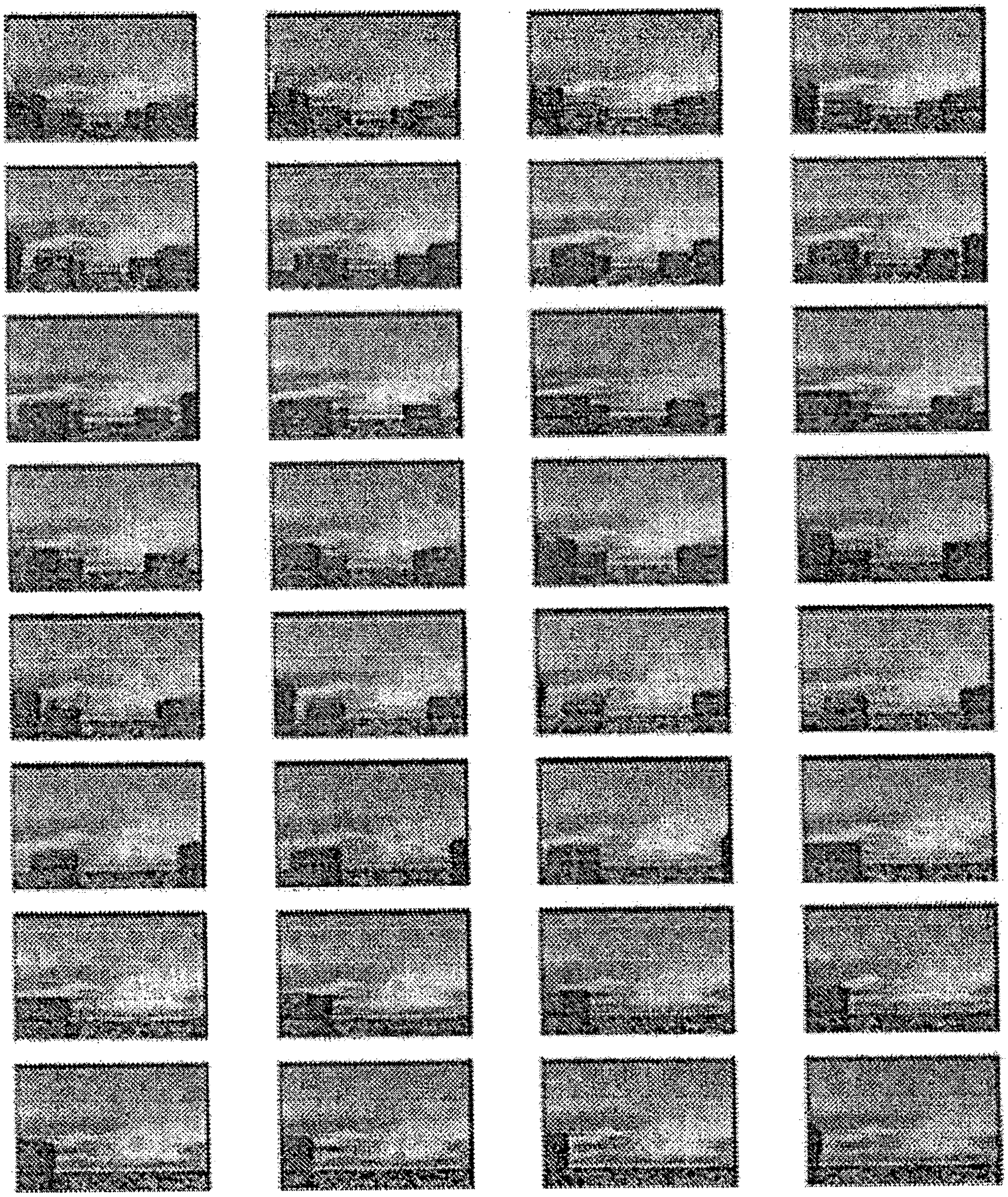


图 9C

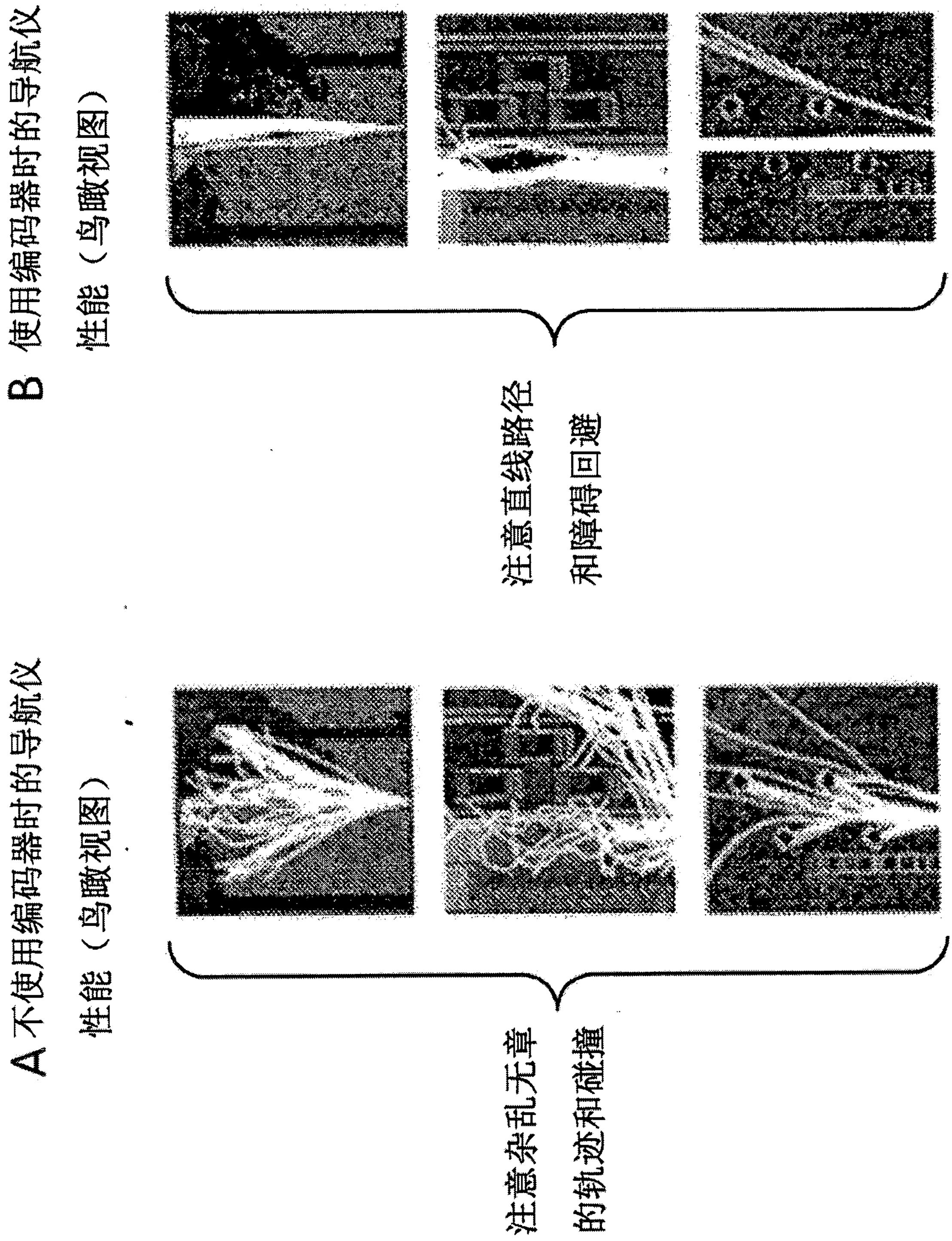

A 不使用编码器时的导航仪性能（鸟瞰视图）
注意杂乱无章的轨迹和碰撞
B 使用编码器时的导航仪性能（鸟瞰视图）
注意直线路径和障碍回避

图 10

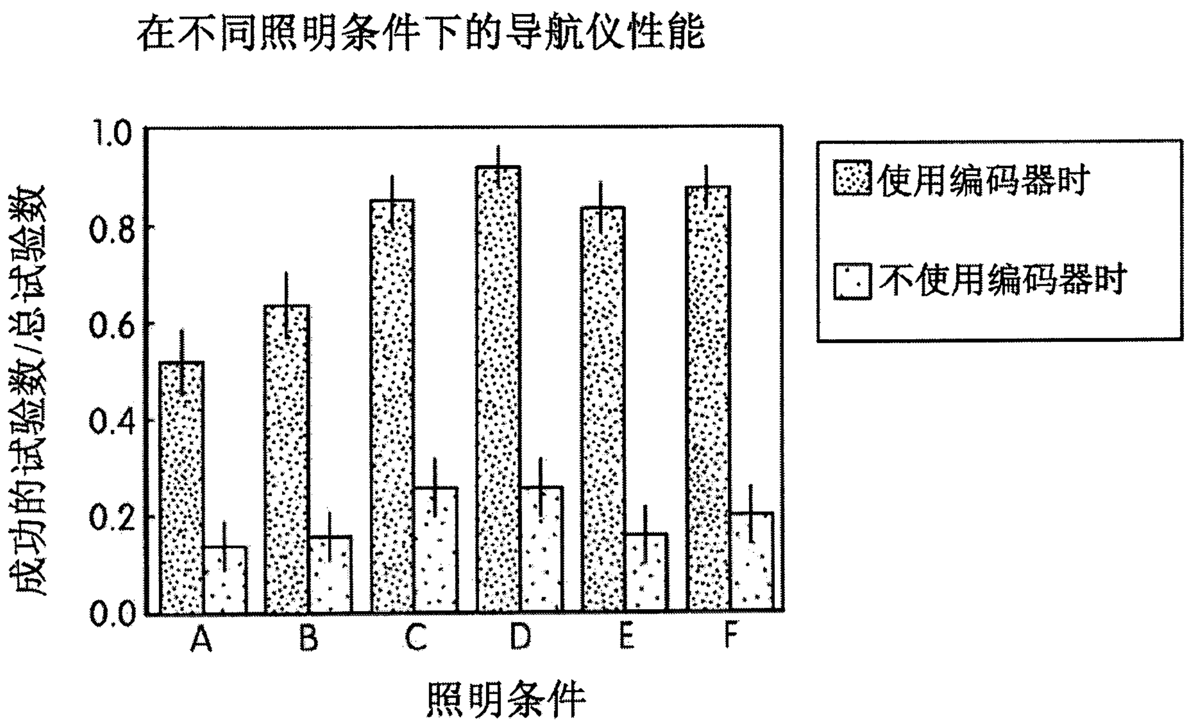

在不同照明条件下的导航仪性能
成功的试验数/总试验数
1.0
0.8
0.6
0.4
0.2
0.0
A
B
C
D
E
F
照明条件
使用编码器时
不使用编码器时

图 11

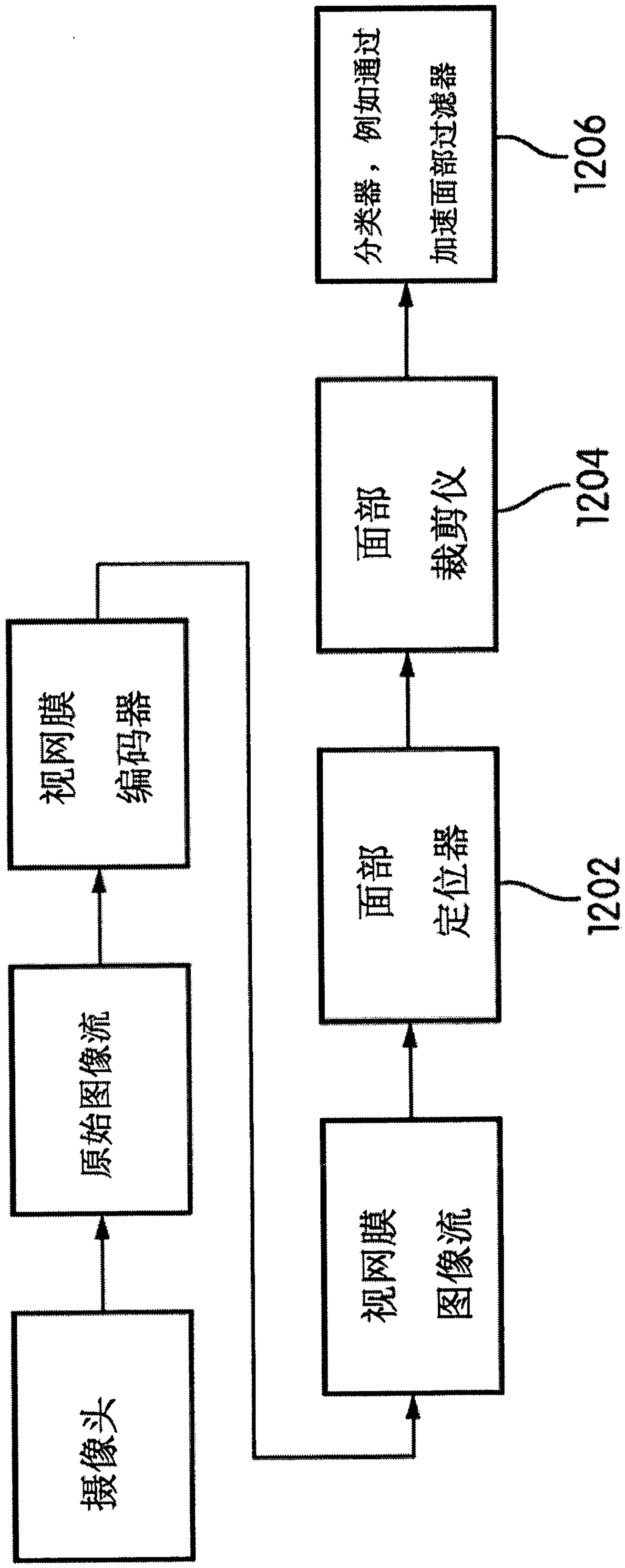

摄像头
原始图像流
视网膜编码器
视网膜图像流
面部定位器
1202
面部裁剪仪
1204
分类器，例如通过加速面部过滤器
1206

图 12

来自用于训练面部识别器的图像流的帧

图 13

来自用于测试面部识别器的图像流的帧

图 14

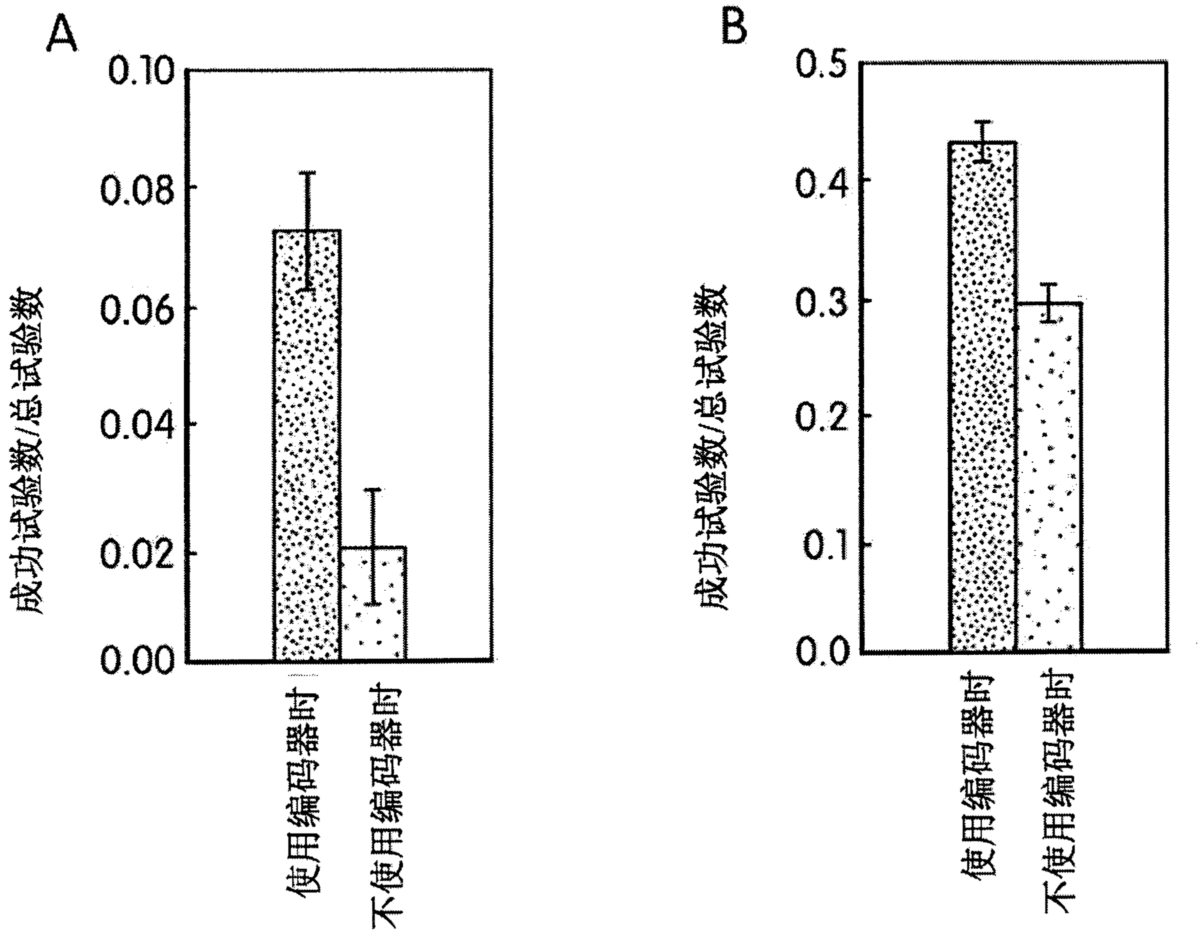

A
0.10
0.08
0.06
0.04
0.02
0.00
成功试验数/总试验数
使用编码器时
不使用编码器时
B
0.5
0.4
0.3
0.2
0.1
0.0
成功试验数/总试验数
使用编码器时
不使用编码器时

图 15

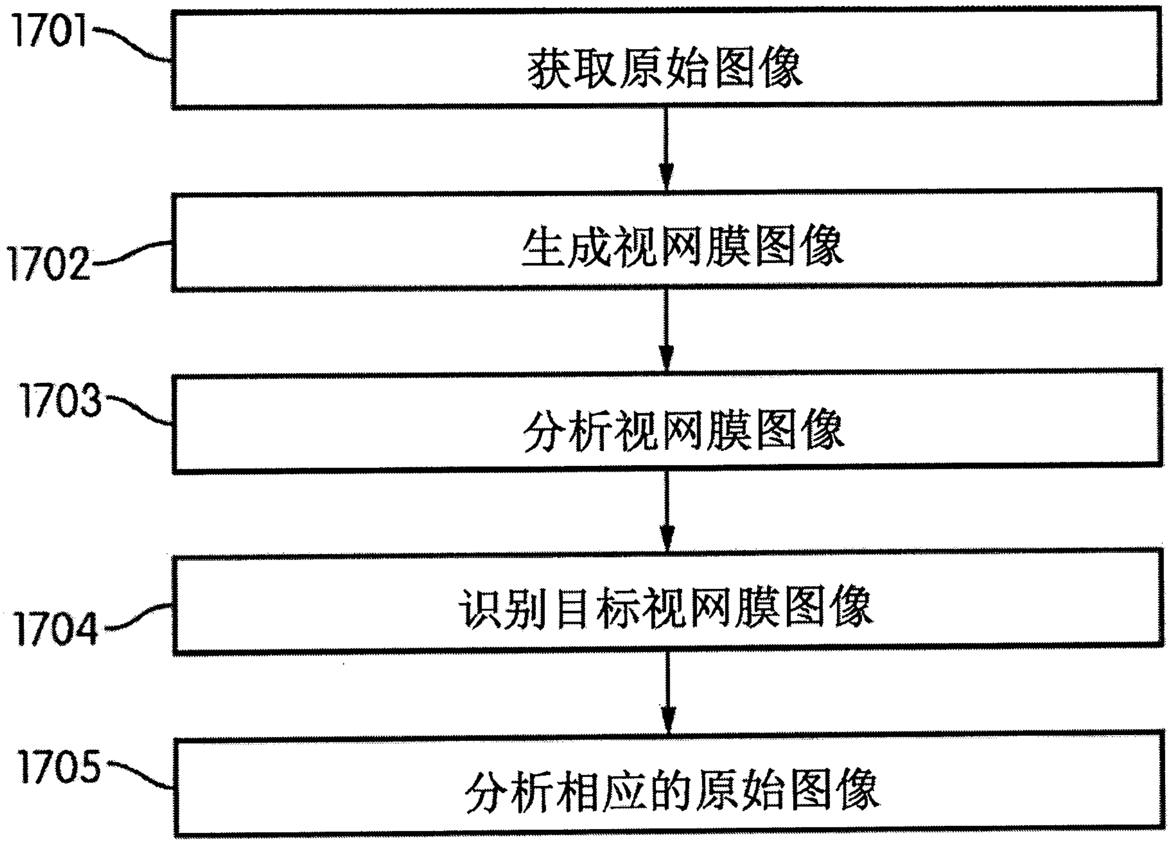

1701
获取原始图像
1702
生成视网膜图像
1703
分析视网膜图像
1704
识别目标视网膜图像
1705
分析相应的原始图像

图 16

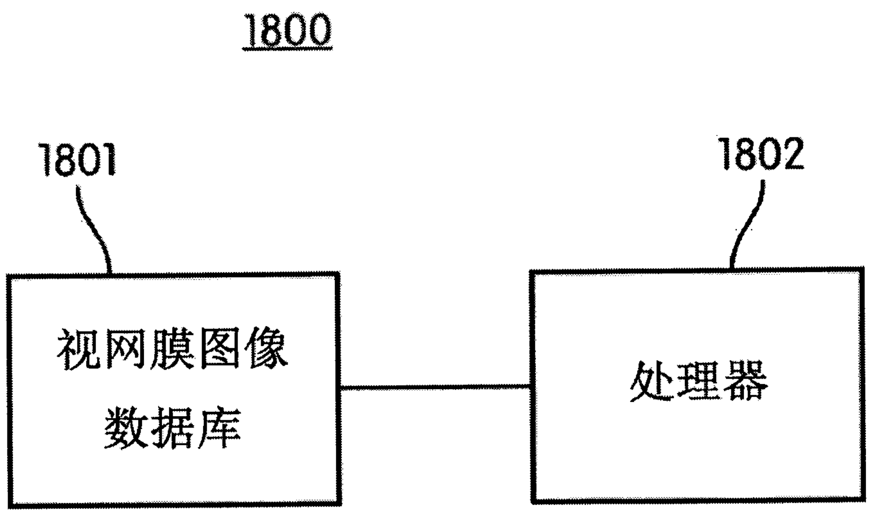

1800
1801
视网膜图像数据库
1802
处理器

图 17

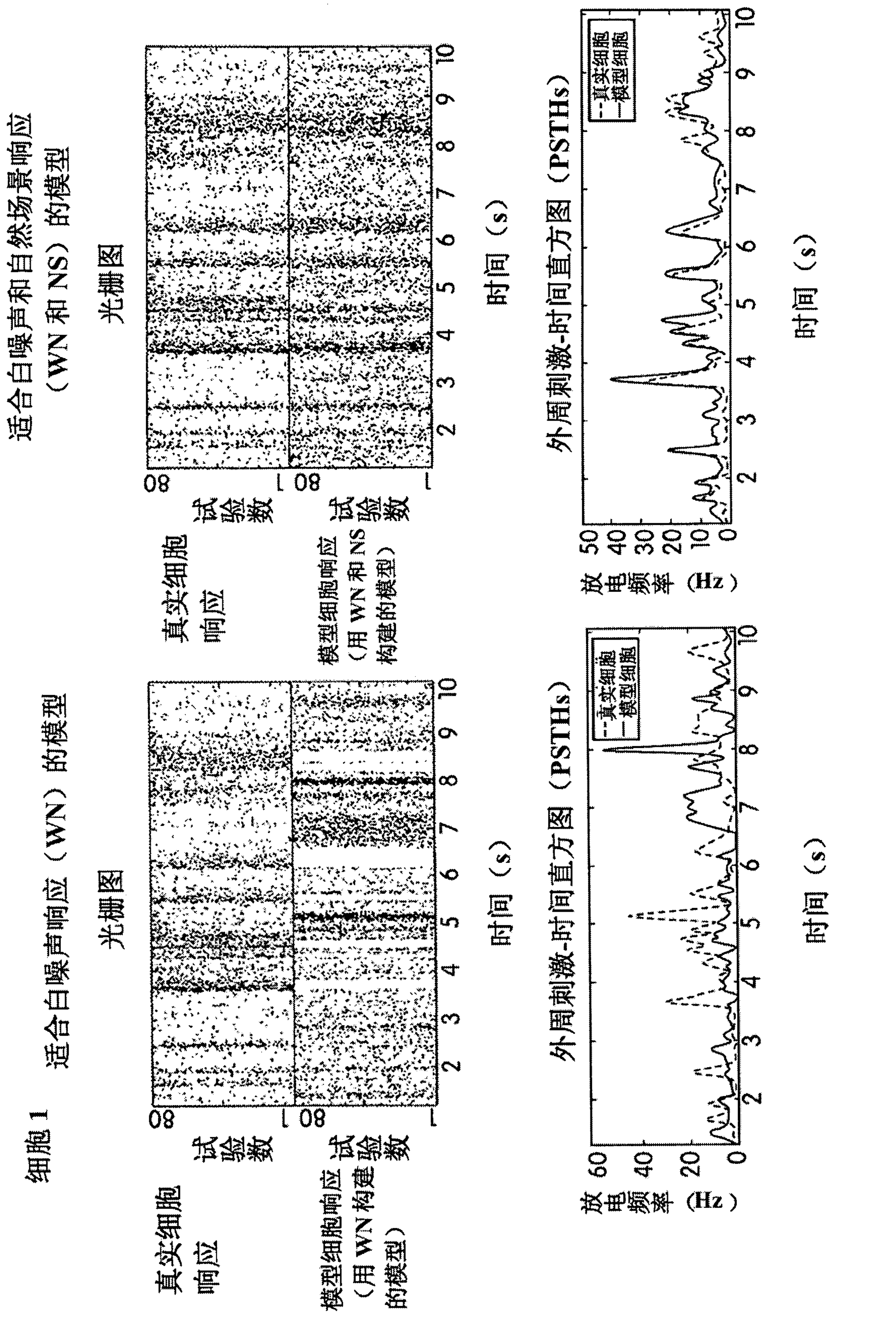

细胞 1
适合白噪声响应（WN）的模型
光栅图
真实细胞响应
试验数
模型细胞响应（用 WN 构建的模型）
时间（s）
外周刺激-时间直方图（PSTHs）
放电频率（Hz）
--真实细胞
—模型细胞
适合白噪声和自然场景响应（WN 和 NS）的模型
光栅图
真实细胞响应
模型细胞响应（用 WN 和 NS 构建的模型）
时间（s）
外周刺激-时间直方图（PSTHs）
放电频率（Hz）
--真实细胞
—模型细胞

图 18A

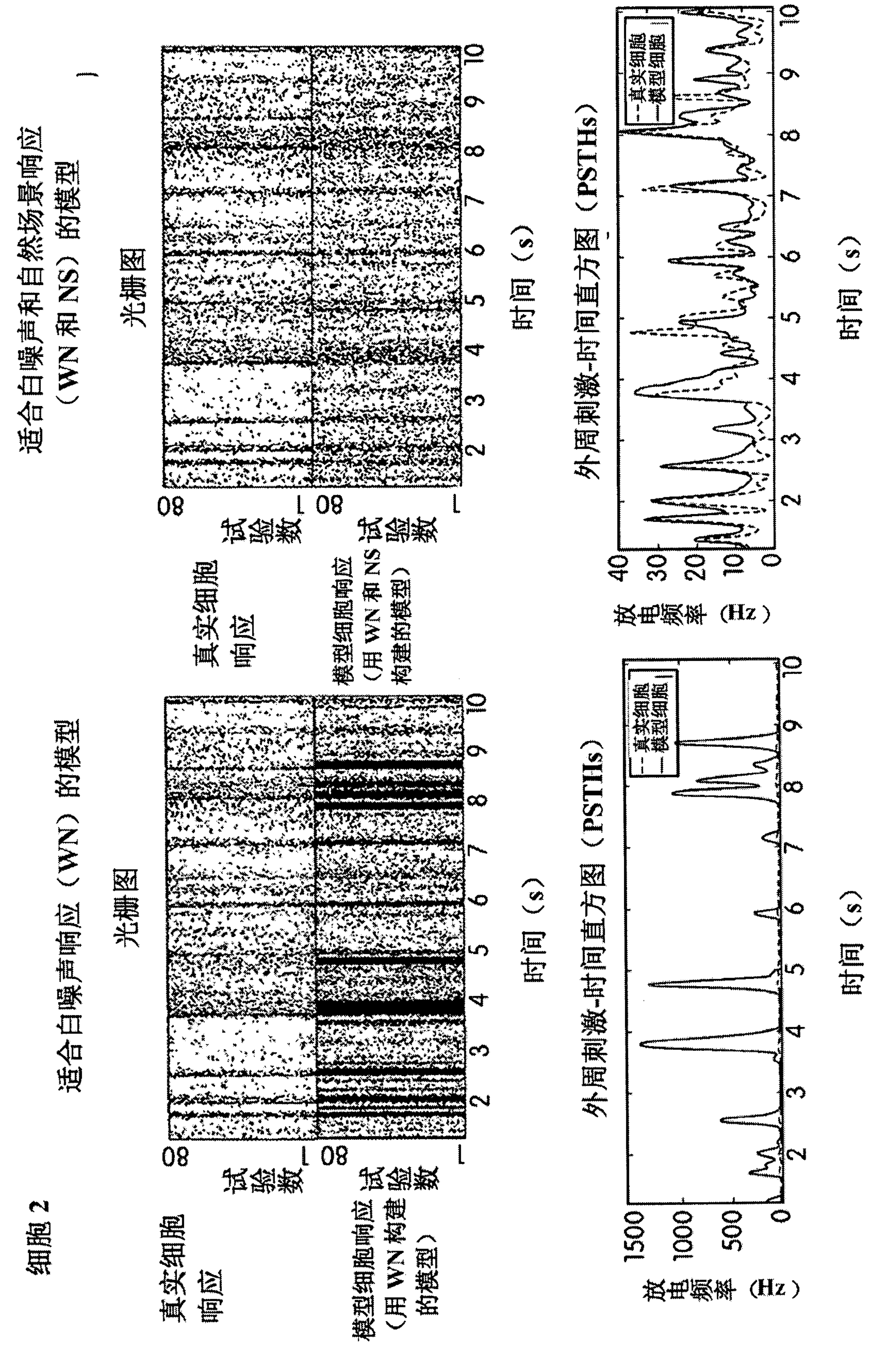

细胞 2
适合白噪声响应（WN）的模型
光栅图
真实细胞响应
模型细胞响应（用 WN 构建的模型）
试验数
时间（s）
外周刺激-时间直方图（PSTHs）
真实细胞
模型细胞
放电频率（Hz）
适合白噪声和自然场景响应（WN 和 NS）的模型
光栅图
真实细胞响应
模型细胞响应（用 WN 和 NS 构建的模型）
试验数
时间（s）
外周刺激-时间直方图（PSTHs）
真实细胞
模型细胞
放电频率（Hz）

图 18B

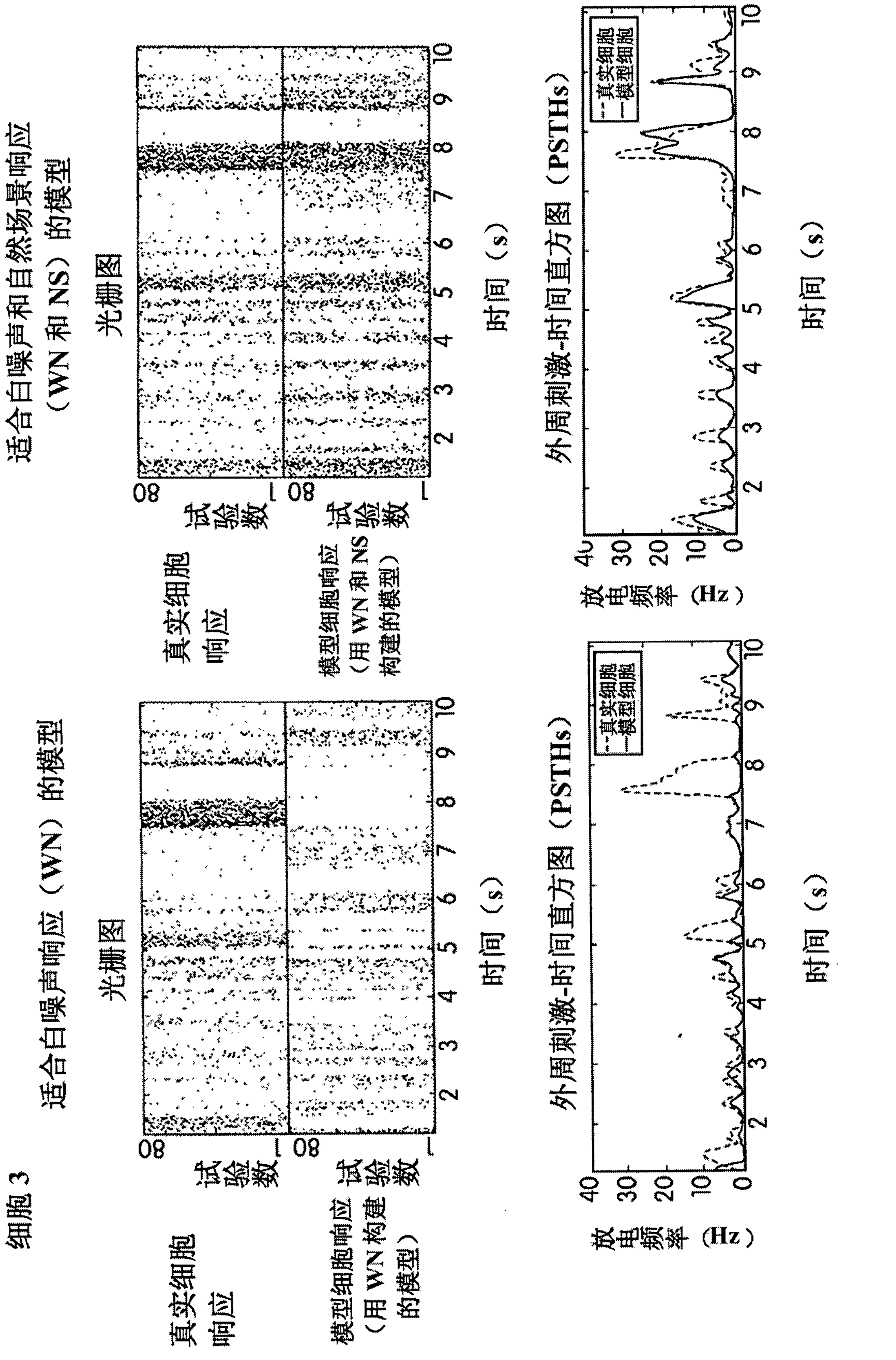

细胞 3
适合白噪声响应（WN）的模型
适合白噪声和自然场景响应（WN 和 NS）的模型
光栅图
真实细胞响应
模型细胞响应（用 WN 构建的模型）
模型细胞响应（用 WN 和 NS 构建的模型）
试验数
时间（s）
外周刺激-时间直方图（PSTHs）
放电频率（Hz）
--真实细胞
—模型细胞

图 18C

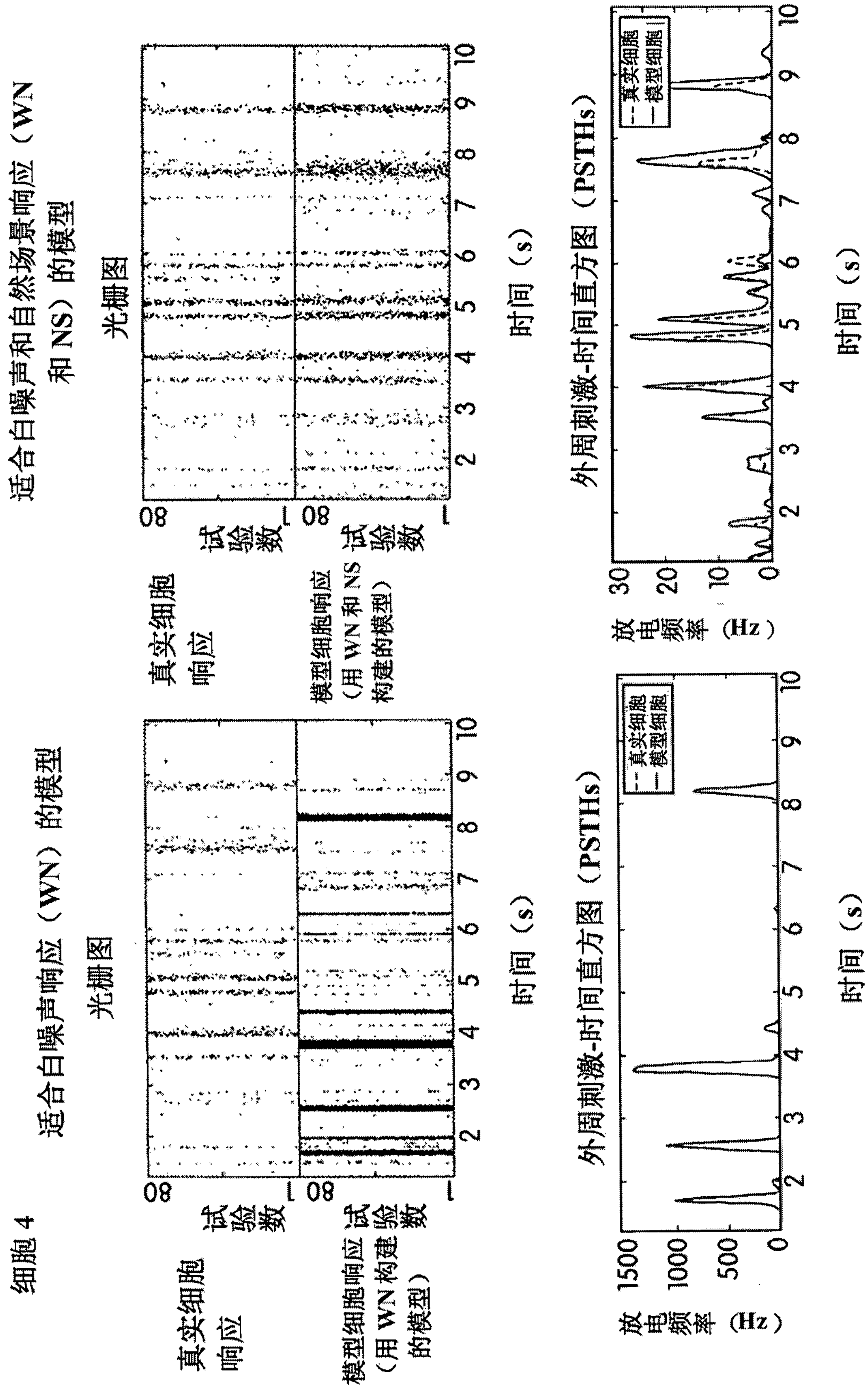

细胞 4
适合白噪声响应（WN）的模型
光栅图
真实细胞响应
模型细胞响应（用 WN 构建的模型）
试验数
时间（s）
外周刺激-时间直方图（PSTHs）
放电频率（Hz）
真实细胞
模型细胞
适合白噪声和自然场景响应（WN 和 NS）的模型
光栅图
真实细胞响应
模型细胞响应（用 WN 和 NS 构建的模型）
试验数
时间（s）
外周刺激-时间直方图（PSTHs）
放电频率（Hz）
真实细胞
模型细胞

图 18D

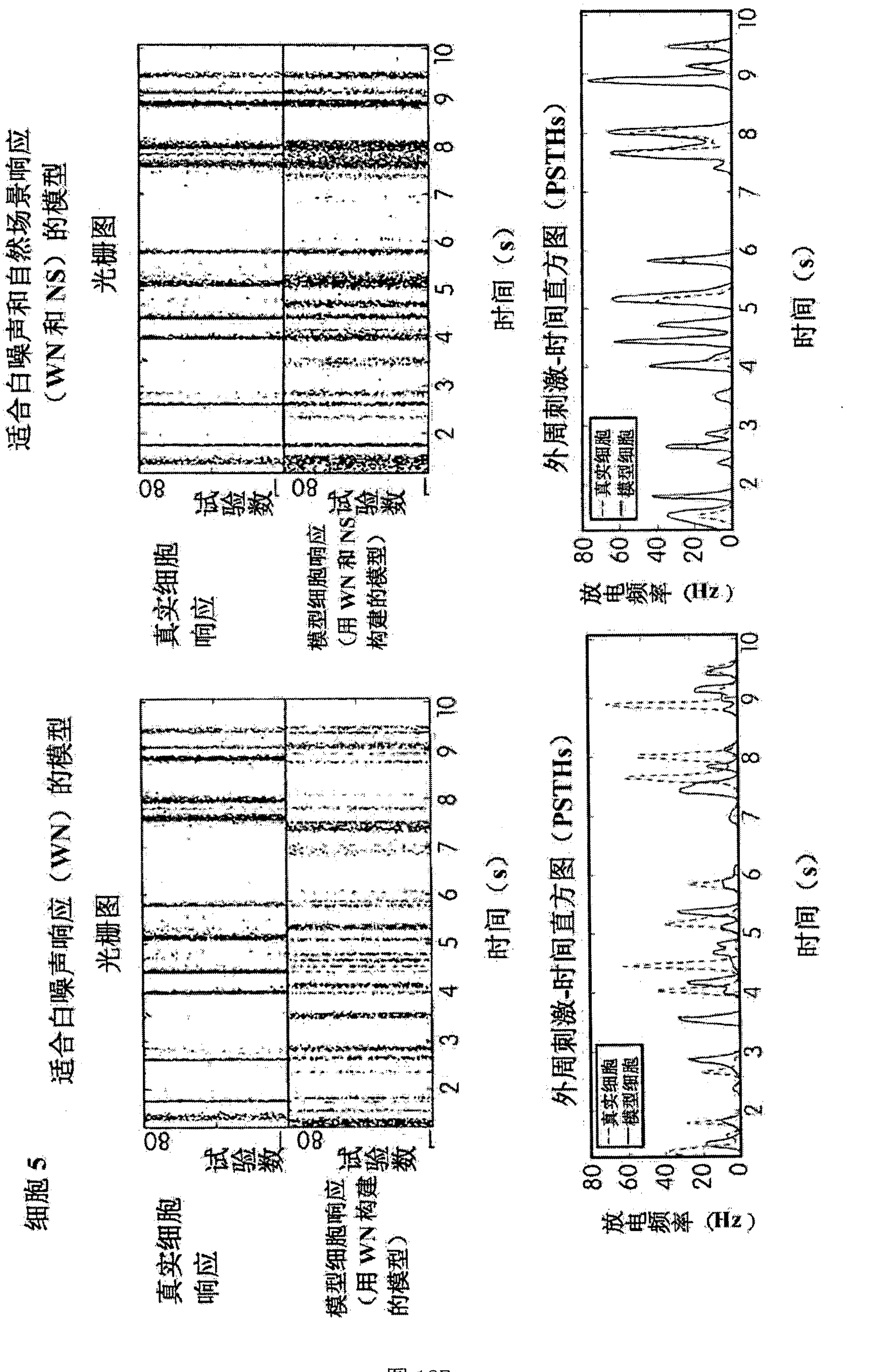

细胞 5
适合白噪声响应（WN）的模型
光栅图
真实细胞响应
模型细胞响应（用 WN 构建的模型）
试验数
时间（s）
外周刺激-时间直方图（PSTHs）
一真实细胞
一模型细胞
放电频率（Hz）
适合白噪声和自然场景响应（WN 和 NS）的模型
模型细胞响应（用 WN 和 NS 构建的模型）

图 18E

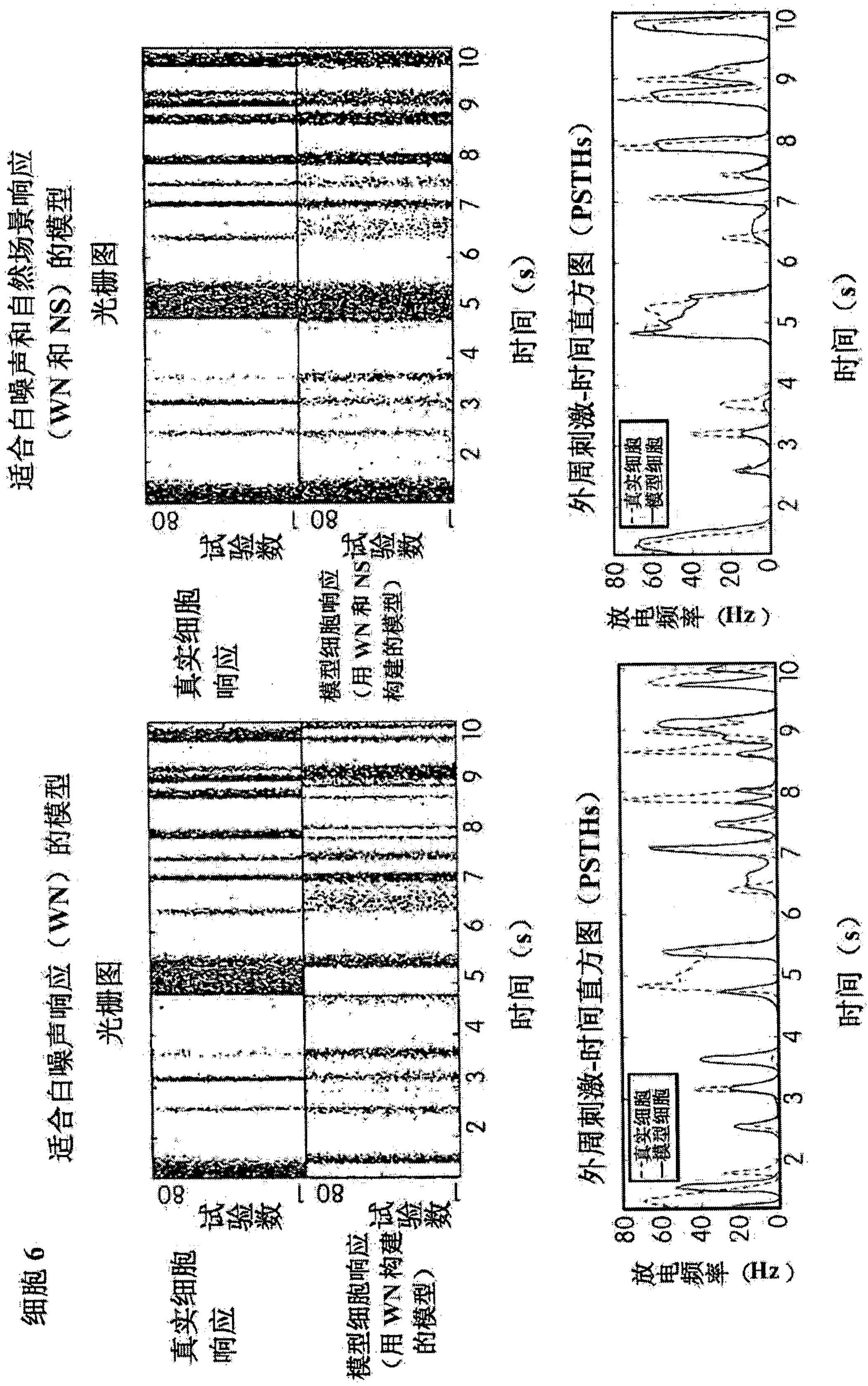

细胞 6
适合白噪声响应（WN）的模型
光栅图
真实细胞响应
模型细胞响应（用 WN 构建的模型）
试验数
时间（s）
外周刺激-时间直方图（PSTHs）
放电频率（Hz）
真实细胞
模型细胞
适合白噪声和自然场景响应（WN 和 NS）的模型
光栅图
真实细胞响应
模型细胞响应（用 WN 和 NS 构建的模型）
试验数
时间（s）
外周刺激-时间直方图（PSTHs）
放电频率（Hz）
真实细胞
模型细胞

图 18F

ABSTRACT

RETINAL ENCODER FOR MACHINE VISION

A method is disclosed including: receiving raw image data corresponding to a series of raw images; processing the raw image data with an encoder to generate encoded data, where the encoder is characterized by an input/output transformation that substantially mimics the input/output transformation of one or more retinal cells of a vertebrate retina; and applying a first machine vision algorithm to data generated based at least in part on the encoded data.

摘 要

用於機器視覺的視網膜編碼器

公開了一種方法，該方法包括：　接收與一系列原始圖像對應的原始圖像數據；用編碼器處理原始圖像數據以生成編碼數據，其中所述編碼器的特徵在於輸入/輸出轉換，所述輸入/輸出轉換本質上模擬脊椎動物視網膜的一個或多個視網膜細胞的輸入/輸出轉換；以及將第一機器視覺算法用於至少一部份基於所述編碼數據而產生的數據.