Novel method of extracting motion from natural movies

Wataru Suzuki, Noritaka Ichinohe, Toshiki Tanii, Taku Hayami, Naohisa Miyakawa, Satoshi Watanabe, Hiroshige Takeichi

A Department of Ultrastructural Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan
b Ichinohe Neural System Group, Laboratory for Molecular Analysis of Higher Brain Functions, RIKEN Brain Science Institute, RIKEN, Wako, Saitama, Japan
c Computational Engineering Applications Unit, Advanced Center for Computing and Communication (ACCC), RIKEN, Wako, Saitama, Japan

HIGHLIGHTS

- We propose a new method to extract object motion information from any natural movies.
- The extracted motion information can be visualized as random dot movies.
- Cells in area MT of common marmosets elicited comparable responses to the movies.
- This method innovates studies of integration of the shape and motion information.

ARTICLE INFO

Article history:
Received 16 February 2017
Received in revised form 26 June 2017
Accepted 3 August 2017
Available online 9 August 2017

Keywords:
Biological motion
Shape from motion
Three-dimensional structure
Natural stimulus
Common marmoset
Area MT

ABSTRACT

Background: The visual system in primates can be segregated into motion and shape pathways. Interaction occurs at multiple stages along these pathways. Processing of shape–from–motion and biological motion is considered to be a higher-order integration process involving motion and shape information. However, relatively limited types of stimuli have been used in previous studies on these integration processes.

New method: We propose a new algorithm to extract object motion information from natural movies and to move random dots in accordance with the information. The object motion information is extracted by estimating the dynamics of local normal vectors of the image intensity projected onto the x-y plane of the movie.

Results: An electrophysiological experiment on two adult common marmoset monkeys (Callithrix jacchus) showed that the natural and random dot movies generated with this new algorithm yielded comparable neural responses in the middle temporal visual area.

Comparison with existing methods: In principle, this algorithm provided random dot motion stimuli containing shape information for arbitrary natural movies. This new method is expected to expand the neurophysiological and psychophysical experimental protocols to elucidate the integration processing of motion and shape information in biological systems.

Conclusions: The novel algorithm proposed here was effective in extracting object motion information from natural movies and provided new motion stimuli to investigate higher-order motion information processing.

© 2017 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Shape–from–motion as an interaction between pathways

In primates, the motion and shape pathways have been attributed to anatomically and functionally separable cortical streams, which are the dorsal and ventral visual pathways emerg-

http://dx.doi.org/10.1016/j.jneumeth.2017.08.006
0165-0270/© 2017 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
1.2. Introducing a novel stimulus to study the pathway interaction

Higher-order processing of shape-from-motion has been explored using random dot motion illustrating geometrical figures, which are not sufficiently complex to cover the variety of motion embedded in natural vision. Although the light markers used in biological motion experiments enabled the extraction of rich information from their complex motion, the motion covered by the stimuli were generally limited to humans, a few other species such as pigeons, or a part of an animal’s body such as the arm of a monkey (Hatsopoulos et al., 2007; Putrino et al., 2015), because of the difficulty in attaching the light markers. Visual stimuli that provide greater varieties of complex motion have been explored in the investigation of higher-order motion processing and the integration processing of shape and motion in neurophysiological and psychophysical experiments. In this study, we proposed a new algorithm to extract and separate the motion information from natural movies of humans, other animals, or artificial objects, and to move random dots in accordance with the information without any physical constraints. Although the computation in this algorithm was based only on local normal curvature vectors defined in the three-dimensional space \((x,y,(x,y))\), where \((x,y)\) is the luminance value at
position \((x,y)\) in each frame image of a movie, the global movement of a structured object emerged from the extracted image motion.

1.3. Suitability of the new algorithm in neuroscience research

In order to assess whether motion information from original natural movies extracted by the proposed algorithm is suitable for the investigation of higher-order motion processing and the integration processing of shape and motion, responses were recorded from cells in the middle temporal (MT) area of common marmosets. The responses to the random dots that moved in accordance with the extracted motion information were compared with the responses to the original movies. Cells in area MT have direction selectivity (Born and Bradley, 2005; Lui and Rosa, 2015; Solomon et al., 2015; Zavitz et al., 2016) and respond to natural movies largely consistently with predictions based on simple stimuli (Nishimoto and Gallant, 2011). Thus, a similarity in the response pattern of area MT cells between the random dot movies and the original natural movies should indicate the effectiveness of the algorithm to extract the motion information from the original movies.

2. Material and methods

2.1. Detailed protocol

2.1.1. Extraction of motion information: time derivatives of normalized local normal vectors of image projected onto the \(x\)-\(y\) plane

Our basic idea for extracting the motion information of objects in a movie was to estimate the dynamics of local normal vectors by calculating the local normal vectors for each pixel in each frame of a movie and their time derivatives (Fig. 1). Because we are interested in the object shapes in the motion extraction method, the local normal vectors were normalized, and were subsequently projected onto the \(x\)-\(y\) plane, on which the original objects are projected in the movie.

A local normal vector at position \((x,y)/(x,y)\) is depicted as follows:

\[
\left( -\frac{\partial h(x,y,t)}{\partial x} \cdot \frac{\partial h(x,y,t)}{\partial y} \right),
\]

where \((x,y)\) is the pixel value (luminance value) at position \((x,y)\). Thus, the \(x\) and \(y\) components of the normalized local normal vector of the frame image projected onto the \(x\)-\(y\) plane were calculated as follows:

\[
n_x = -\frac{1}{\sqrt{\left(\frac{\partial h(x,y,t)}{\partial x}\right)^2 + \left(\frac{\partial h(x,y,t)}{\partial y}\right)^2 + 1}} \cdot \frac{\partial h(x,y,t)}{\partial x},
\]

\[
n_y = -\frac{1}{\sqrt{\left(\frac{\partial h(x,y,t)}{\partial x}\right)^2 + \left(\frac{\partial h(x,y,t)}{\partial y}\right)^2 + 1}} \cdot \frac{\partial h(x,y,t)}{\partial y}.
\]

Although the normalized local normal vector is related to a standard image feature called the “gradient”, we used the term “normal vector” instead because normal vectors on a three-dimensional object contain geometric information about the shape of the object, and we would like to emphasize the extraction of motion information of objects with a shape.

We applied a spatial low-pass filter before calculating the local normal vectors to remove discontinuities and enable the calculation of the space derivative. Subsequently, the time derivatives of the projected local normal vectors were calculated. The dynamic vector field defined by the time derivatives should reflect some aspects of the shape and motion information in the movie. It should be noted that the time derivatives of the normalized local normal vectors were calculated at each image position \((x,y)\), but not at a specific position that is defined at a fixed point on an object.

A qualitative investigation based on the equations indicated several characteristics of the algorithm. First, a normalized local normal vector markedly changes when the angle between the \(x\)-\(y\) plane and the tangent plane of an object surface markedly changes (e.g., rotation in depth). Second, a local normal vector contains the space derivative of image pixel values, indicating that the calculation of the local normal vector functions as an edge detector. Thus, the time derivative of the normalized local normal vector enlarges when an edge with high contrast is moving. Because the calculation was local, the exact motion of the object was not necessarily extracted. As discussed later, it was not our intention to calculate the ground-truth motion of objects.

The motion information was visualized based on random dot motion. Each dot moves in accordance with the dynamic vector field. We varied the gray level of the dots to establish a correspondence between individual dots in adjacent frames easily. The supplementary movies are examples of movies in which a human actor and a tiger move (Supplementary movies 1 and 2, respectively) and random dots following the dynamic vector field are extracted using the present algorithm (Supplementary movies 3 and 4). The number of square random dots was 160,000 (the side of each dot was 1 pixel). Although each frame of the random dot movies contained no object information, the global movements of the human actor and the tiger emerged when they moved.

2.2. An example of application

2.2.1. Animals

Experiments were performed using two adult common marmoset monkeys (Callithrix jacchus; weighing 300–400 g). This study was approved by the Experimental Animal Committee of the National Institute of Neuroscience and Psychiatry and the animals were cared for in accordance with the “Guiding Principles of the Care and Use of Animals in the Field of Physiological Science” of the Japanese Physiological Society.

2.2.2. Stimuli

We generated 55 random dot movies made from 55 natural movies containing a wide variety of moving objects: animals, human actors, and artificial objects. Motion information from animal movies, human actor movies, and artificial movies (Fig. 2a, Supplementary movies 5 and 6) were extracted by the following procedure to generate random dot movies that traced the extracted motion information. First, various natural movies, e.g., a walking tiger, a jumping human actor, and a moving toy, were recorded with a resolution of 640×480 pixels at 30 frames/s. Subsequently, the movies were trimmed to be one-second long. The colors of the movies were converted to grayscale. After applying a spatial low-pass filter (circular averaging filter, radius: 4 pixels), the normalized local normal vectors of the frame images for each pixel were extracted and projected onto the \(x\)-\(y\) plane using Eqs. (2) and (3). Subsequently, the time derivatives of the projected local normal vectors were calculated as the difference between neighboring frames. The normalization factor in Eqs. (2) and (3) was approximated by the constant value computed for the last frame. In total, 16,000 square dots were placed (the side of each dot was 3 pixels). The gray level of the dots was randomly varied from 0 (black) to 255 (white). The background gray level was 0 (black). Each dot moved in accordance with the time derivative of the projected local normal vector for the corresponding pixel. To prevent dots from clustering in specific regions at the end of the movie, half of the dots were moved prospectively from the first frame to the last frame in accordance with the derivative in the movie and the other half
were moved retrospectively from the last frame to the first frame in accordance with the time reversal of the derivative in the movie. The average lifetime of the dots was set to approximately 0.5 s. Static 8000 random dots were added to mask the shape information. The size of the stimulus movie was approximately 20°. Each stimulus was presented 12 times in a pseudorandom order. The display device was a liquid crystal display (LCD) monitor (EIZO FG2421, Ishikawa, Japan) with a refresh rate of 60 Hz (Ghodrati et al., 2015).

2.2.3. Electrophysiological experiment procedure
The neural data were recorded in two adult common marmosets. All the methods for surgery, anesthesia, and electrophysiological recordings were described in detail previously (Suzuki et al., 2015a,b). In brief, to investigate whether the proposed algorithm extracted the motion information of the natural movies, we recorded 64 multi-unit activities (MUAs) from the area MT of the animals. Surgery and electrophysiological recordings were conducted under anesthesia induced by ketamine hydrochloride (Ketalar, 25 mg/kg, intramuscular [i.m.]) following atropine sulfate (0.15 μg/kg, i.m.). The electrocardiographs (ECGs), expired CO2 level, and rectal temperature were monitored continuously. The animals were placed in a stereotactic apparatus and the head holder and the recording chamber were implanted on the skull. Following the surgery, the stereotactic apparatus was removed and the head was fixed using the head holder. For the electrophysiological recordings, the anesthetic was switched to an intravenous (i.v.) infusion of remifentanil (Ultiva, 0.1 μg/kg/min) with rocuronium bromide (Esplax, 13 μg/kg/min, i.v.) to induce muscular paralysis. The animal was artificially ventilated with a mixture of 70% N2O and 30% O2. Before the recordings, the pupil was fully dilated with topical tropicamide (0.5%) and phenylephrine hydrochloride (0.5%). A contact lens was used to focus the eye contralateral to the recorded hemisphere at a distance of 57 cm.

Electrodes were inserted with reference to the superior temporal sulcus (STS), and retinotopical organization was revealed by intrinsic optical signal imaging (data not shown). A micromanipulator was used to lower a 32-channel multicontact linear-array electrode (Neuronexus, Ann Arbor, Michigan, USA), which contained four shanks (400-μm shank separation). Each shank had eight electrode contacts (impedance: ~1 MΩ at 1 kHz) with an intercontact spacing of 200 μm. We inserted the electrode perpendicularly to the cortical surface of the posterior and ventral STS parts corresponding to the area MT (Lui and Rosa, 2015), where strong responses to visual motion stimuli were observed under anesthesia. MUAs were simultaneously recorded from the 32 contacts. The timings of band-pass (0.3–5 kHz) filtered MUAs and task events (stimulus onset and offset) were recorded at 24 kHz using a TDT signal processing system (RZ2; Tucker-Davis Technologies, Alachua, Florida, USA).

The receptive fields (RFs) and direction selectivity of the MUAs in area MT were identified before presenting the stimulus movies. First, the position and size of the RF for 32 MUAs were determined by presenting a small dot (4.8°) at different positions in the visual field. Second, the direction selectivity for each MUA was evaluated by presenting a traditional stimulus, i.e., random dots moving to eight directions with 100% coherence at a speed of 2.9°/s in a 3.5° window. The centers of the window were either at the fixation point, or at ±4.5° of the fixation point (nine positions).

2.2.4. Data analysis—direction selectivity
The RF of each MUA was estimated to be the window in which the maximum response was elicited when the direction selectivity was investigated using the traditional coherently moving random dots. The preferred direction of motion for traditional coherently moving random dots was defined as the direction that elicited the maximum response in the window. The estimated direction selectivity for each MUA was compared between the traditional coherently moving random dots and the random dot movies generated with the new algorithm. The motion vector of each dot in each frame of the random dot movies was calculated by taking the position differences along the x- and y-axes between neighboring frames. The motion vectors were vector-summed and thresholded: if the norm of the summed motion vector exceeded 2.9°/s, i.e., the speed of the dots in the traditional coherently moving random dots, the vector was normalized to 1; otherwise, vectors were normalized to 0. The preferred direction for each MUA estimated from the random dot movie was defined by the weighted average of the motion vector in the RF window. The weighting coefficient for each frame was the response magnitude of the MUA defined as the aver-
Fig. 3. Comparison of the preferred direction of the random dot movies and example of multiunit activities in middle temporal (MT) area of the common marmoset. a. Histogram of the inner-products for all responsive MUAs (n = 50) is shown. The inner-products are calculated from normalized vectors with preferred direction derived from random dot motion with 100% coherence and those from the random dot movies generated with the present algorithm. b. Rasters and spike density functions show responses to random dot movies that follow the vector field extracted by the present algorithm and to the corresponding original natural movies. The order of the plot is the same as in Fig. 2a. Red indicates the responses to the original movies and black to the random dot movies. c. Scatter plot of the responses to 55 random dot movies and the corresponding original natural movies. MUA, multi-unit activity.

Age firing rate in a temporal window from 67 to 100 ms after the onset of the corresponding frame subtracted by the spontaneous firing rate. Moreover, the weighting coefficient could be positive or negative depending on whether the response was excitatory or inhibitory. Therefore, the estimated preferred direction took into account both strong responses to the optimal direction and weak responses or suppression to the opposite direction. Finally, the estimated preferred directions were compared between the traditional coherently moving random dots and the random dot movies generated with the new algorithm, by taking an inner-product between the two vectors. Their norms were normalized to 1. If the inner-product was 1, the two vectors would thus point to the same direction.

2.2.5. Data analysis–comparison between the random dot and natural movies

The neural responses to the random dot movies and original natural movies were evaluated in three aspects: magnitude (quantitative), effective region of the visual field (spatiotemporal), and population representation (representational aspects). The average response magnitude was calculated as the average firing rate of MUAs in the window from 33 to 1033 ms after the movie onset subtracted by the average spontaneous firing rate in the 300-ms interval immediately before the movie onset. The response in each MUA was considered to be significant when the average firing rate in the window was larger than the mean +3 standard deviation (SD) of the spontaneous firing rate for one or more movies. To compare the stimulus selectivity of each MUA between the random dot movies and the original movies, the coefficient of correlation between the response magnitudes to the random dot movies and the corresponding original natural movies was calculated. The correlation coefficient was considered to be significant when it was larger than the mean +3SD of the distribution of 1000 correlation coefficients with random pairing between the random dot movies and the original natural movies.
Local spatial regions of the visual field with the effective responses of MT cells (effective regions, ERs) were compared between the original natural movies and the random dot movies. First, the ER for each MUA for each of the original natural movies was derived from the weighted average of luminance change images. The luminance change images represented absolute pixel luminance value changes and were generated from differences between neighboring frames in the original natural movies. The absolute values were taken because an important aim of this analysis was to localize the spatial regions contributing to the responses of MT cells and with a high magnitude of luminance changes irrespective of the signs. The luminance change was referred to instead of local motion because computation of the local motion in the natural movie was not a trivial task. While the local spatiotemporal gradients in luminance may be calculated, they do not necessarily correspond to the local motion. For example, the same spatiotemporal gradients in luminance stem from motion of a bar in one direction at the leading edge and motion of another bar with the opposite contrast in the opposite direction at the trailing edge; therefore, it is ambiguous in motion estimation. The images showing luminance changes were binned by 20 pixels. The weighting coefficient for each frame in the movie was the response magnitude of MUAs defined as the average firing rate in the window from 67 to 100 ms after the onset of the corresponding frame subtracted by the spontaneous firing rate. Similarly, the ER in the random dot movies was derived from the weighted average of dot speed images with the weighting coefficient of MUAs. The dot speed images were generated from the random dot movies by averaging the speed of the dots in 20-pixel windows. The speed of each dot was defined by the norm of the motion vector, which was obtained by taking position differences along the x- and y-axes between neighboring frames. The weighting coefficient for each frame in the movie was the response magnitude of MUA defined as the average firing rate in the window from 67 to 100 ms after the onset of the corresponding frame subtracted by the spontaneous firing rate. Because the absolute luminance value change was used in estimating ERs for the natural movies, the absolute velocity (i.e., speed) was used in estimating ERs for the random dot movies. Finally, the correlation of the ERs for each MUA was evaluated between the original natural and random dot movies. The ER was obtained by averaging across the image frames weighted by the neuronal responses, delineating integrated stimulus portions that yielded a response. While this “spike-triggered average method” used in this study was computationally equivalent to the “reverse correlation method” to reconstruct RF and response properties (DeAngelis et al., 1993), the meaning of the result was different because the spatiotemporal frequency distribution of the image set used here was not “white”. In other words, there were spatial correlations in visual attributes, e.g., dot appearance, speed, or direction of the dot motion, in the random dot movies. A response to a frame of a movie could potentially be attributed to several visual attributes. Therefore, this analysis could not unambiguously relate individual stimulus elements, i.e., motion of individual dots to isolated neural responses to reconstruct and compare individual RFs and response properties such as direction selectivity or tuning curves. With a large number of movies, the spatiotemporal correlation might be reduced. Here, we anticipated that ERs represented integrated portions of the stimulus yielding the response, and would be useful in the comparison between the original and random dot movies used in this study.

The random dot and original natural movies were represented in a 32-dimensional population response space in which each dimension was defined by the normalized response magnitude of an MUA. The response magnitude was normalized by calculating the z-score with the mean and standard deviation across all movies. Euclidian distances in the population response space were calculated for all pairs in the random dot movies and in the original natural movies. The similarity of the representation between the random dot movies and the corresponding original natural movies was assessed on the basis of the Euclidian distances for the corresponding pairs. The correlation coefficient of the Euclidian distances for all the corresponding pairs was calculated and considered to be significant when it was larger than the mean + 3SD of the distribution of 1000 correlation coefficients with random assignments of response magnitudes to movies.

2.2.6. Histological processing
After the experiments, the animals were sacrificed by using an intraperitoneal overdose of sodium pentobarbital (Nembutal, 75 mg/kg) followed by intracardial perfusion with 0.1 M phosphate-buffered serum (PBS; pH 7.4) and, subsequently, 4% paraformaldehyde in PBS (Merck, Whitehouse Station, NY, US). The dissected brain blocks were immersed in 0.1 M PBS with 10, 20, or 30% sucrose. Coronal sections of 50-μm thickness were cut on a freezing microtome (Yamato-Koki, Saitama, Japan) and prepared.
Fig. 5. Effective spatial region in the original natural movie and random dot movie. Effective spatial regions with the response of four MUAs in the original natural movie (left) and the random dot movie (right). Pseudo-color (arbitrary unit) indicates the response magnitude. Channel #27 corresponds to MUA described in Fig. 3b. Channel #22, #14, and #4 belongs to different shanks, which are located at intervals of 400, 800, and 1200 μm from the shank of channel #27, respectively. Black circles indicate the centers of gravity. MUA, multi-unit activity.

in a series of three sections. For areal demarcation, one in three consecutive sections was stained with myelin (Pistorio et al., 2006) and the second with the Nissl substrate with thionin. The third section was used for fluorescence imaging of Dil, which marked the electrode tracks after the electrophysiological experiments. We confirmed that all the electrodes penetrated the area MT from the sections stained for myelin (Fig. 2b) (Rosa and Elston, 1998).

3. Results

Among the recorded 64 MUAs, 50 (78%) showed significant excitatory responses to one or more of the 55 original and 55 random dot movies. Fig. 3a shows the inner-products for all responsive MUAs, which were significantly greater than 0 (Mann-Whitney U test, p < 0.00001). These results indicated that the direction selectivity of the MUAs in area MT derived from the random dot movies generated with the new algorithm was mostly consistent with the direction selectivity determined by traditional stimuli. The possible reasons for these discrepancies are discussed in the Discussion section.

Fig. 3b depicts an example of the MUAs to the random dot and the corresponding original natural movies. The random dot movies tended to elicit responses similarly to the corresponding original natural movies, indicating a resemblance in stimulus selectivity between the random dot movies and the original natural movies. The correlation coefficient was 0.68 between the response magnitudes to the random dot movies and those to the original natural movies (Fig. 3c). This was larger than the mean +3SD (0.01 + 0.42) of the distribution of 1000 correlation coefficients with random pairing. Among the 50 MUAs, 33 (66%) showed a significant correlation.
between the responses to random dot movies and the corresponding original natural movies.

The ER of MT cells in this study would be the same between the random dot movies and the original natural movies if the extracted motion information in each local region was similar between the random dot movies and the original natural movies. Before comparing the neural responses to the movies, we compared the movies themselves. First, we verified whether the dot motion in local regions of the random dot movies reflected the pixel value changes in the corresponding regions of the original natural movies. Fig. 4a depicts an image of luminance changes and a dot speed image each averaged across all the sequential frames of all the original natural movies and the random dot movies, respectively (\( n = 1595; 30-1 \) frames for 55 stimuli). These showed the specific spatial bias of luminance change and dot speed in the stimulus movie used in this study. For comparison, the center of gravity was calculated for each of the original movies and the random dot movies. Fig. 4b depicts the scatter plot of the centers of gravity for all the frames of all the movies. The correlation coefficients were 0.90 and 0.92 for the \( x \)- and \( y \)-axes, respectively, and were significant (\( p < 0.00001; n = 1595 \)). Thus, the spatial bias was considered to be similar between the pixel luminance value changes and the dot speed for the stimulus movies used in this experiment. Subsequently, we compared the ERs between the original natural movies and the random dot movies. Fig. 5 shows the ERs for the original natural movies and the random dot movies for four MUAs: one (channel #27) was described in Fig. 3b and the other three, namely channel #22, #14, and #4, were selected from different shanks that were located at intervals of 400, 800, and 1200 \( \mu \)m from the shank of channel #27, respectively. The correlation coefficients of the centers of gravity between the original movies and the random dot movies for all responsive MUAs (\( n = 50 \)) were 0.35 (\( p = 0.013 \)) and 0.38 (\( p = 0.007 \)) for the \( x \)- and \( y \)-axes, respectively.

Representations of the random dot movies and the original natural movies in the population response space were examined by plotting all the movies in the response space spanned by the normalized responses of the recorded MT cells and by calculating the Euclidean distances for all pairs in the random dot movies and in the original natural movies, respectively. The correlation coefficients of the Euclidean distances for the random dot movies and the corresponding original natural movies were 0.60 and 0.64 for the first and second animal, respectively. These were larger than the mean +3SD (−0.17 + 0.21 and −0.21 + 0.23) for the first and second animal, respectively of the distribution of 1000 correlation coefficients with random assignments.

4. Discussion

4.1. Evaluation of the results

In this study, we developed a new algorithm to extract motion information from natural movies and showed that the random dot movies that followed the algorithm elicited neuronal responses similarly to the corresponding original natural movies in area MT of common marmosets, in terms of magnitude, spatial localization in the visual field, and population representation. Area MT is known to play an important role in visual motion processing (Born and Bradley, 2005; Lui and Rosa, 2015), since the cells in area MT are selective for direction, speed, and spatial frequency. The characteristics of the responses to natural movies in area MT are consistent with those to simple stimuli such as bars, gratings, and dots (Nishimoto and Gallant, 2011; Lui et al., 2012). Thus, the similarity between the responses to the random dot movies and those to the original natural movies shown in this study indicated that the proposed algorithm is a useful method for investigating biologically critical motion information from natural movies.

Although the correlation coefficients of the responses between random dot movies and original natural movies were significantly high, the responses were not the same. More specifically, the intercept of the linear regression through the points in Fig. 3c would be far from zero. Thus, the random dot movies tended to generate far greater responses when the spike rates were low. Consistent with this observation, some MUAs showed different direction selectivity between the random dot motion generated with the new algorithm and the traditional coherently moving random dots. These might be due to the sensitivity of the cells in area MT to the spatial frequency and contrast of the stimulus image (Priebe et al., 2003; Lui et al., 2007). The random dot movies obviously consisted of spatial frequency components and image contrasts different from the original natural movies. Testing various types of random dot movies, e.g., contrast-reversed versions (black dots on a white background) or Gaussian-filtered versions (dots containing more low spatial frequency components), would also be required in future studies to validate the notion that the proposed algorithm indeed extracts motion information of natural movies.

The random dot movies introduced here provided a wide variety of complex motion stimuli extracted from deformable objects without any physical constraints. In contrast, conventional stimuli used for shape-from-motion processing and biological motion usually do not, because the stimuli for shape-from-motion are based on geometrical figures (Siegel and Andersen, 1988; Todd, 1984), and multiple light markers used for biological motion are constrained by the moving agents (Cutting, 1978; Johansson, 1973). For example, attaching light markers for preparing a stimulus of the tiger’s biological motion is laborious. In principle, the present algorithm can be applied to any daily movie. Thus, this new visual stimulus set of the random dot movie would be useful for investigating higher-order motion processing and the integration processing of shape and motion in neurophysiological and psychophysical experiments.

4.2. Limitation of the current methods

The aim of using algorithms to extract optical flow (Baker et al., 2011; Barron et al., 1994) is to estimate the ground-truth flow of motion in an image sequence. These usually assume some constraints, such as a continuity equation (Horn and Schunck, 1981), to compute the vector field in the image sequence. In contrast to this, we did not attempt to estimate the ground-truth flow and the present algorithm did not require any assumptions or constraints on the image sequence. Thus, an image sequence can contain nonrigid objects such as animals. In the present algorithm, the field vector at each pixel was calculated from the spatial derivatives; thus, it depended on the change in contrast between nearby pixels. In an extreme case, it would be different if the grayscale was inverted. For example, the local vector direction in a movie in which a white object (e.g., a white bear) moves on a black background (e.g., soil) is opposite to that in a movie in which an identical, but black, bear moves on a white background (e.g., snow). Thus, the present algorithm is different from those used for the extraction of the optical flow.

4.3. Significance of the new algorithm for the visual sciences

A distinguishing feature of the present algorithm is that it consisted of two independent computations: one for the vector field derived from spatial and time derivatives in the original movies and the other for the visualization of the vector field based on the random dot motion. This implies that we can visualize the vector field using a different modality of random dot other than motion.
For example, Supplement movie 7 shows a film clip in which random dots change their size in accordance with the dynamic vector field extracted from Supplement movie 1 using the new algorithm. The lengths along the x- and y-axes of the random ellipse-shaped dots change in accordance with the vector field that has x and y components represented in Eqs. (2) and (3). The two independent computations for making the stimulus movie may be useful for preparing various types of stimulus sets for neurophysiological and psychophysical motion experiments. The presentation of the same information in different modalities may help in designing experiments to identify the response properties.

In the generation of visual stimuli, the present motion extraction algorithm can be applied to natural movies and does not require parameterized geometric shapes, providing a variety of complex visual motion stimuli, including social actions by animals. Because such visual motion stimuli do not contain shape information unless they move, it is possible to dissociate the shape information from the motion factor. Thus, this algorithm is useful for investigating the integration processing of shape and motion in neurophysiological and psychophysical experiments. For example, the functional roles of cortical areas other than area MT in integration processing might be inferred by comparing the neuronal responses to original natural movies with those to their random dot motion obtained using the present algorithm and static frames selected from the original movies. This might further elucidate the neural substrates for the interaction between the shape and motion pathways.

5. Conclusion

We have proposed a new algorithm to extract motion information from natural movies and new stimulus movies in which random dots move in accordance with the algorithm. The random dot movies elicited neuronal responses similarly to the corresponding original natural movies in area MT of common marmosets, indicating that the proposed algorithm is a useful method for investigating biologically critical motion information from natural movies. The algorithm can be applied to natural movies and does not require parameterized geometric shapes, providing a variety of complex visual motion stimuli. Thus, the new algorithm is a useful tool for investigating motion information processing in neurophysiological and psychophysical experiments.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted. This article does not contain any studies with human participants performed by any of the authors.

Acknowledgements

We would like to thank Ueno Zoo for permitting us to take animal videos at their premises, and Kazuhisa Sakai and Miyuki Yasue for their assistance with the experiments. This work was supported by Grant-in-Aid for Scientific Research (C) (26430031), an Intramural Research Grant (Grant No. 23-7) for Neurological and Psychiatric Disorders from National Center for Neurology and Psychiatry (NCNP), and the program for Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS) from Japan Agency for Medical Research and Development (AMED), Japan. We would like to thank Editage (www.editage.jp) for English language editing.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jneumeth.2017.08.006.

References

Baizer, J.S., Ungerleider, L.G., Desimone, R., 1991. Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. J. Neurosci. 11, 168–190.

Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J., Szeliski, R., 2011. A database and evaluation methodology for optical flow. Int. J. Comput. Vision 92, 1–31.

Barron, J., Fleet, D., Beauchemin, S., 1994. Performance of optical flow techniques. Int. J. Comput. Vision 12, 43–77.

Borns, R.T., Bradley, D.C., 2005. Structure and function of visual area MT. Annu. Rev. Neurosci. 28, 157–189.

Burr, D.C., Ross, J., 2002. Direct evidence that speedlines influence motion mechanisms. J. Neurosci. 22, 8661–8664.

Cutting, J.E., 1978. Generation of synthetic male and female walkers through manipulation of a biomechanical invariant. Perception 7, 393–405.

DeAngelis, G.C., Ohzawa, I., Freeman, R.D., 1993. Spatiotemporal organization of simple-cell receptive fields in the cat’s striate cortex. I. General characteristics and postnatal development. J. Neurophysiol. 69, 1091–1117.

Geisler, W.S., 1999. Motion streaks provide a spatial code for motion direction. Nature 400, 65–69.

Ghodrati, M., Morris, A.F., Price, N.S., 2015. The (un)suitability of modern liquid crystal displays (LCDs) for vision research. Front. Psychol. 6, 303.

Hatsopoulos, N.G., Xu, Q., Amit, Y., 2007. Encoding of movement fragments in the motor cortex. J. Neurosci. 27, 5105–5114.

Horn, B.K.P., Schunck, B.G., 1981. Determining optical flow. Artif. Intell. 17, 185–203.

Johansson, G., 1973. Visual perception of biological motion and a model for its analysis. Percept. Psychophys. 14, 201–211.

Kraevit, D.J., Saleem, K.S., Baker, C.J., Mishkin, M., 2011. A new neural framework for visuospatial processing. Nat. Rev. Neurosci. 12, 217–230.

Kraevit, D.J., Saleem, K.S., Baker, C.J., Ungerleider, L.G., Mishkin, M., 2013. The ventral visual pathway: an expanded neural framework for the processing of object quality. Trends Cognit. Sci. 17, 26–49.

Krekelberg, B., Dannenberg, S., Hoffmann, K.P., Bremmer, F., Ross, J., 2003. Neural correlates of implied motion. Nature 424, 674–677.

Liu, L.L., Rosa, M.G., 2015. Structure and function of the middle temporal visual area (MT) in the marmoset: comparisons with the Macaque monkey. Neurosci. Res. 93, 62–71.

Liu, L.L., Bourne, J.A., Rosa, M.G., 2007. Spatial and temporal frequency selectivity of neurons in the middle temporal visual area of new world monkeys (Callithrix jacchus). Eur. J. Neurosci. 25, 1780–1792.

Liu, L.L., Dobiecki, A.E., Bourne, J.A., Rosa, M.G., 2012. Breaking camouflage: responses of neurons in the middle temporal area to stimuli defined by coherent motion. Eur. J. Neurosci. 36, 2063–2076.

Nishimoto, S., Gallant, J.L., 2011. A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. J. Neurosci. 31, 14551–14564.

Pavlova, M., Sokolov, A., 2000. Orientation specificity in biological motion perception. Percept. Psychophys. 62, 889–899.

Perry, C.J., Fallah, M., 2014. Feature integration and object representations along the dorsal stream visual hierarchy. Front. Comput. Neurosci. 8, 84.

Pistorio, A.L., Hendry, S.H., Wang, X., 2006. A modified technique for high-resolution staining of myelin. J. Neurosci. Methods 153, 135–146.

Priebe, N.J., Cassanello, C.R., 2003. The neural representation of speed in macaque area MT/V5. J. Neurosci. 23, 5650–5661.

Putrino, D., Wong, Y.T., Weiss, A., Pesaran, B., 2015. A training platform for many-dimensional prosthetic devices using a virtual reality environment. J. Neurosci. Methods 244, 68–77.

Rosa, M.G., Elston, G.N., 1984. Visuotopic organisation and neuronal response selectivity for direction of motion in visual areas of the caudal temporal lobe of the marmoset monkey (Callithrix jacchus): middle temporal area, middle temporal crescent, and lateralgeniculate cortex. J. Comp. Neurol. 393, 505–527.

Siegel, R.M., Andersen, R.A., 1988. Perception of three-dimensional structure from motion in monkey and man. Nature 331, 259–261.

Solomon, S.S., Chen, S.C., Morley, J.W., Solomon, S.G., 2015. Local and global correlations between neurons in the middle temporal area of primate visual cortex. Cereb. Cortex 25, 3182–3196.

Somi, S., 1984. Upside-down presentation of the Johansson moving light-spot pattern. Perception 13, 283–286.

Suzuki, W., Banno, T., Miyakawa, N., Abe, H., Goda, N., Ichinohe, N., 2015a. Mirror neurons in a new world monkey, common marmoset. Front. Neurosci. 9, 459.
Suzuki, W., Tani, T., Banno, T., Miyakawa, N., Abe, H., Ichinohe, N., 2015b. Functional columns in superior temporal sulcus areas of the common marmoset. Neureport 26, 1133–1139.

Todd, J.T., 1984. The perception of three-dimensional structure from rigid and nonrigid motion. Percept. Psychophys. 36, 97–103.

Ungerleider, L.G., Mishkin, M., 1982. Two cortical visual systems. In: Ingle, D.J., Goodale, M.A., Mansfield, R.J.W. (Eds.), Analysis of Visual Behavior. The MIT Press, Cambridge, pp. 549–586.

Ungerleider, L.G., Galkin, T.W., Desimone, R., Gattass, R., 2008. Cortical connections of area V4 in the macaque. Cereb. Cortex 18, 477–499.

Webster, M.J., Bachevalier, J., Ungerleider, L.G., 1994. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. Cereb. Cortex 4, 470–483.

Zavitz, E., Yu, H.H., Rowe, E.G., Rosa, M.G., Price, N.S., 2016. Rapid adaptation induces persistent biases in population codes for visual motion. J. Neurosci. 36, 4579–4590.