



Figures and figure supplements

Asymmetric ON-OFF processing of visual motion cancels variability induced by the structure of natural scenes

Juyue Chen *et al*

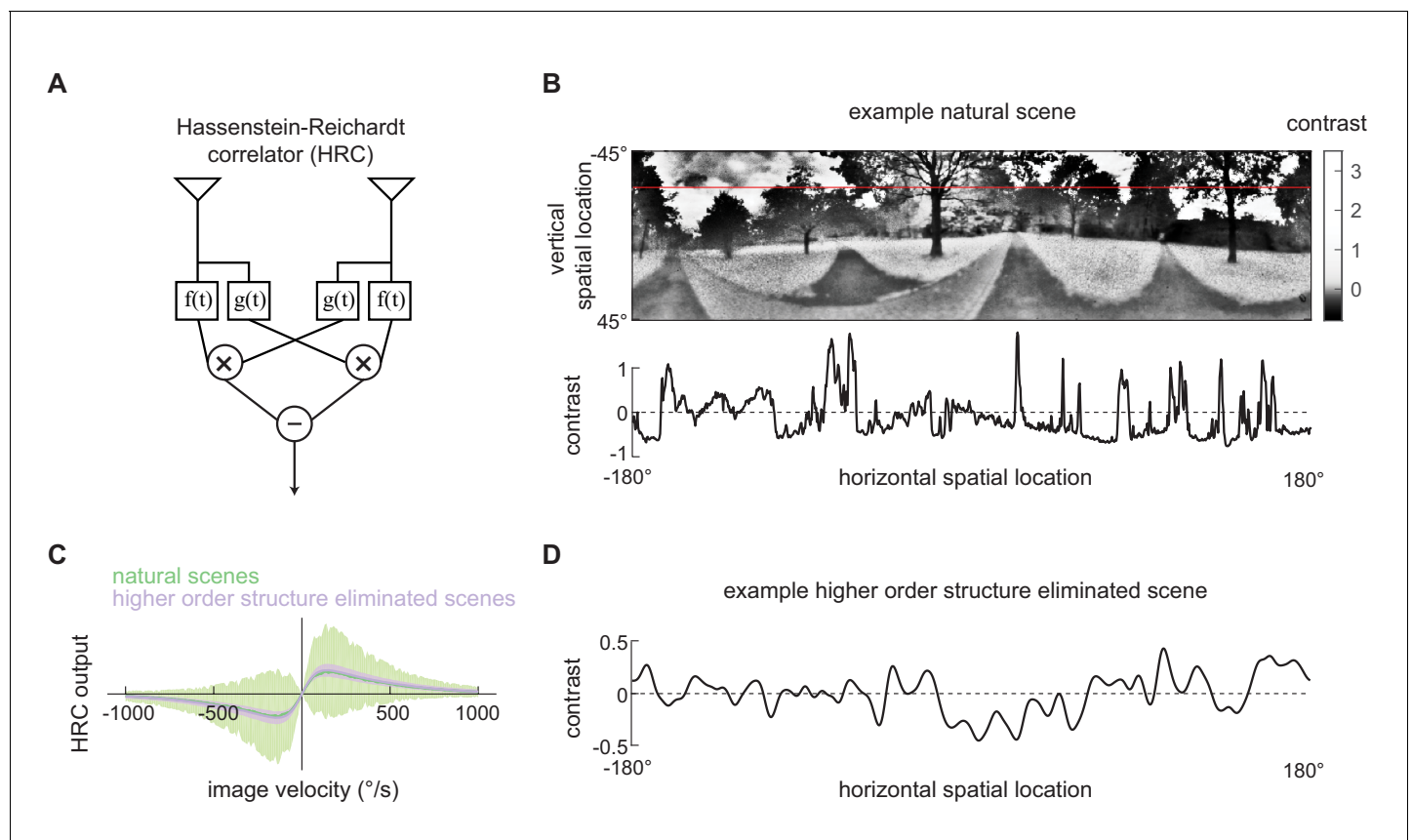


Figure 1. Second-order motion detectors perform poorly with natural scene inputs. **(A)** Schematics of the Hassenstein-Reichardt correlator (HRC). Each half of the HRC receives inputs from two nearby points in space, which are filtered in time with filters $f(t)$ and $g(t)$, and then multiplied together. The full HRC receives outputs from two symmetric halves with opposite direction tuning and subtracts two outputs. **(B)** An example two-dimensional photograph from a natural scene dataset (top), including a one-dimensional section (image) through the photograph (bottom), indicated by the red line. So that the image can be viewed clearly, the contrasts in the photograph were mapped onto gray levels so that an equal number of pixels were represented by each gray level. **(C)** Average response (line) and variance (shaded) of the outputs of an HRC (equivalent to a motion energy model; Adelson and Bergen, 1985) when presented with naturalistic motion at various velocities. Images were sampled from natural scenes (green) or from a synthetic image dataset in which all higher order structure was eliminated (purple, see Materials and methods). **(D)** Example synthetic image in which all higher order structure was eliminated.

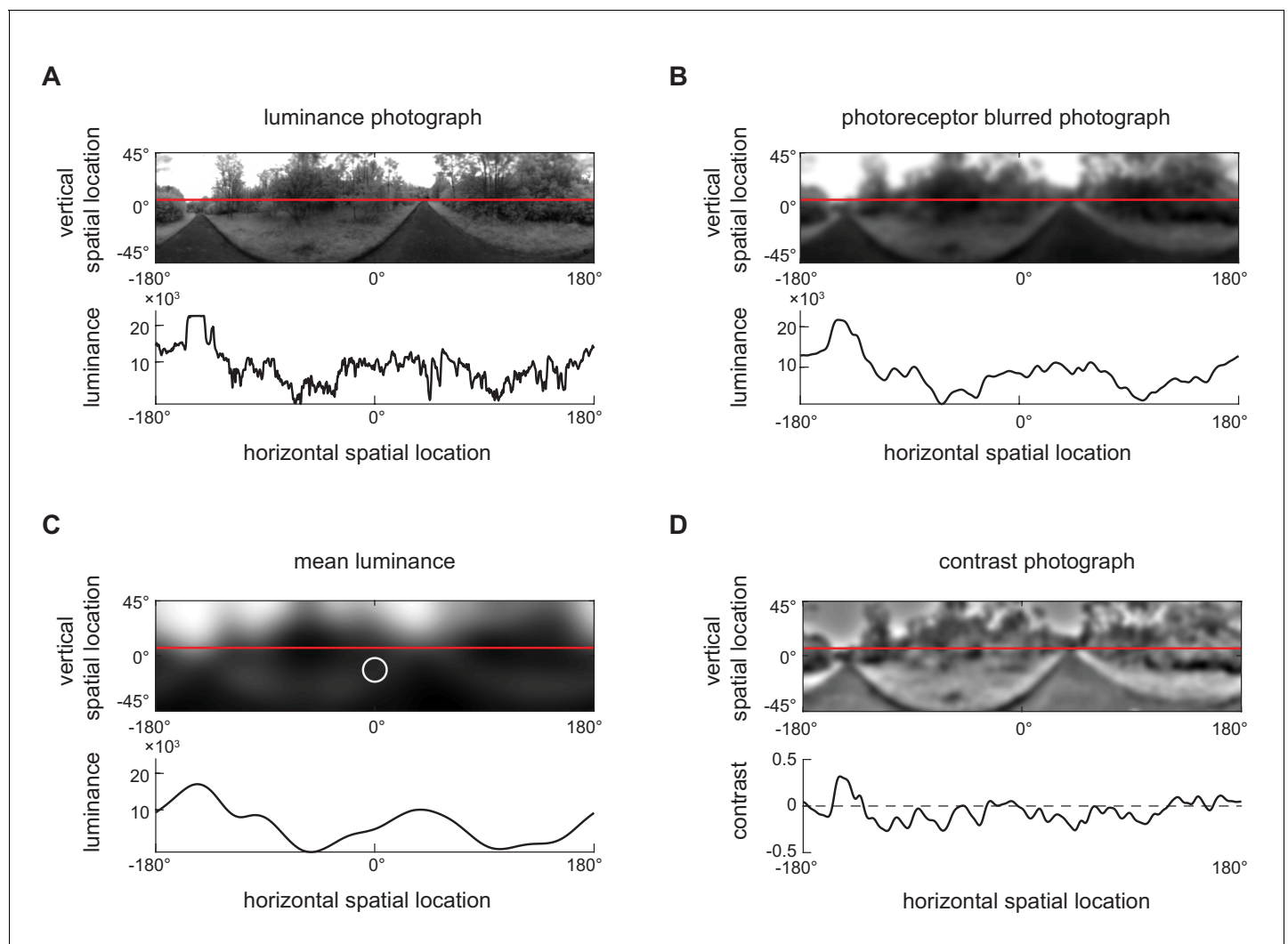


Figure 1—figure supplement 1. Converting luminance signals into contrast signals (see Materials and methods). (A) An example two-dimensional luminance photograph (*upper*) from the dataset, including a one-dimensional section through the photograph (*bottom*), indicated by the red line. The luminance is shown in log scale in the 1-dimensional image. (B) The two-dimensional photograph from (A) filtered with a two-dimensional Gaussian filter (FWHM = 5.3°) to represent the spatial filtering of the fly's ommatidia. This yielded the photograph I_{blur} . (C) The two-dimensional photograph from (B) filtered with a two-dimensional Gaussian filter (FWHM = 25°) to represent local mean luminance. The white circle represents the FWHM of the filter. This yielded the photograph I_{mean} . (D) The two-dimensional photograph from (A) in units of contrast. We calculated the contrast by computing $c = \frac{I_{\text{blur}} - I_{\text{mean}}}{I_{\text{mean}}}$.

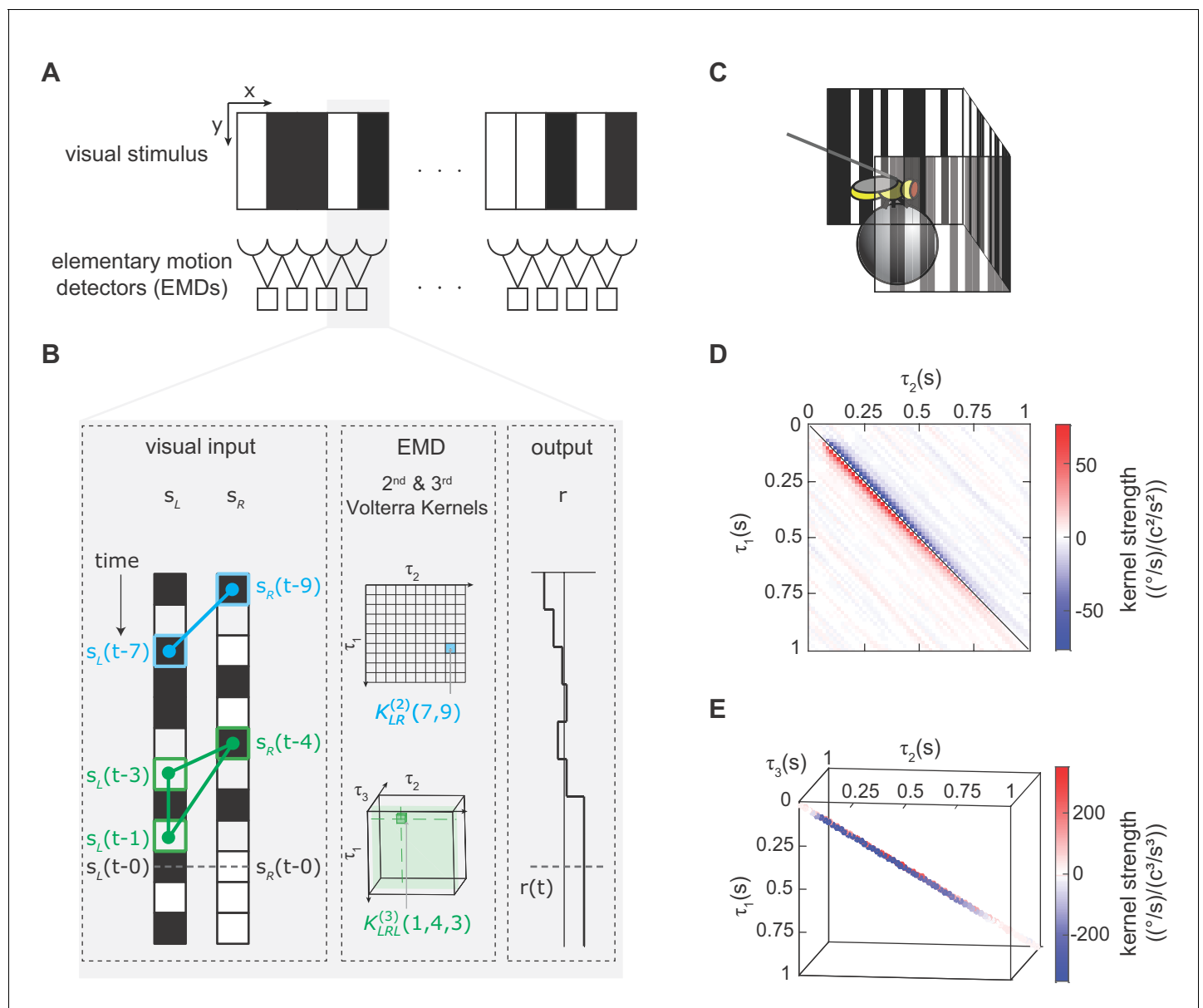


Figure 2. We modeled the fly's motion computation algorithm with second- and third-order Volterra kernels and extracted the kernels using reverse-correlation (Appendix 1, Materials and methods). **(A)** Visual depiction of our model, where the fly's motion computation system consists of a spatial array of elementary motion detectors (EMDs), and each EMD receives inputs from two neighboring spatial locations. We presented flies with vertically uniform stimuli with 5° -wide pixels, roughly matching ommatidium spacing. **(B)** Diagram showing how the output of one EMD at time t , indicated by gray dashed line, is influenced by the second- and third-order products in the stimulus. *Left:* visual inputs of one EMD, s_L and s_R . The visual stimulus contained products of pairwise and triplet points with various spatiotemporal structure. One specific pairwise product is highlighted (blue barbell) and one specific triplet product is highlighted (green triangle). *Middle:* The motion computation of the EMD is approximated by the second-order kernel (blue) and the third-order kernel (green). The second-order kernel (blue) $K_{LR}^{(2)}(\tau_1, \tau_2)$ is a two-dimensional matrix. For example, the response at time t is influenced by the products of $s_L(t-7)$ and $s_R(t-9)$ with weighting $K_{LR}^{(2)}(7, 9)$. The third-order kernel (green) $K_{LRL}^{(3)}(\tau_1, \tau_2, \tau_3)$ is a three-dimensional tensor. The response at time t is influenced by $s_L(t-1)s_R(t-4)s_L(t-3)$ with weighting $K_{LRL}^{(3)}(1, 4, 3)$. *Right:* turning response at time t is influenced by all pairwise and triplet products in the visual stimulus, with weightings given by the second- and third-order kernel elements. **(C)** Diagram of the fly-on-a-ball rig. We tethered a fly above a freely-rotating ball, which acted as a two-dimensional treadmill. We presented stochastic binary stimuli, and measured fly turning responses. **(D)** The extracted second-order kernel. The color represents the magnitude of the kernel, with red indicating rightward turning and blue indicating leftward turning to positive pairwise spatiotemporal correlations. Above the diagonal line, the matrix represented left-tilted pairwise products (example in B) and below the diagonal line represents right tilted pairwise products. **(E)** The extracted third-order kernel. For visualization purposes, we show only the two diagonals with the largest magnitude.

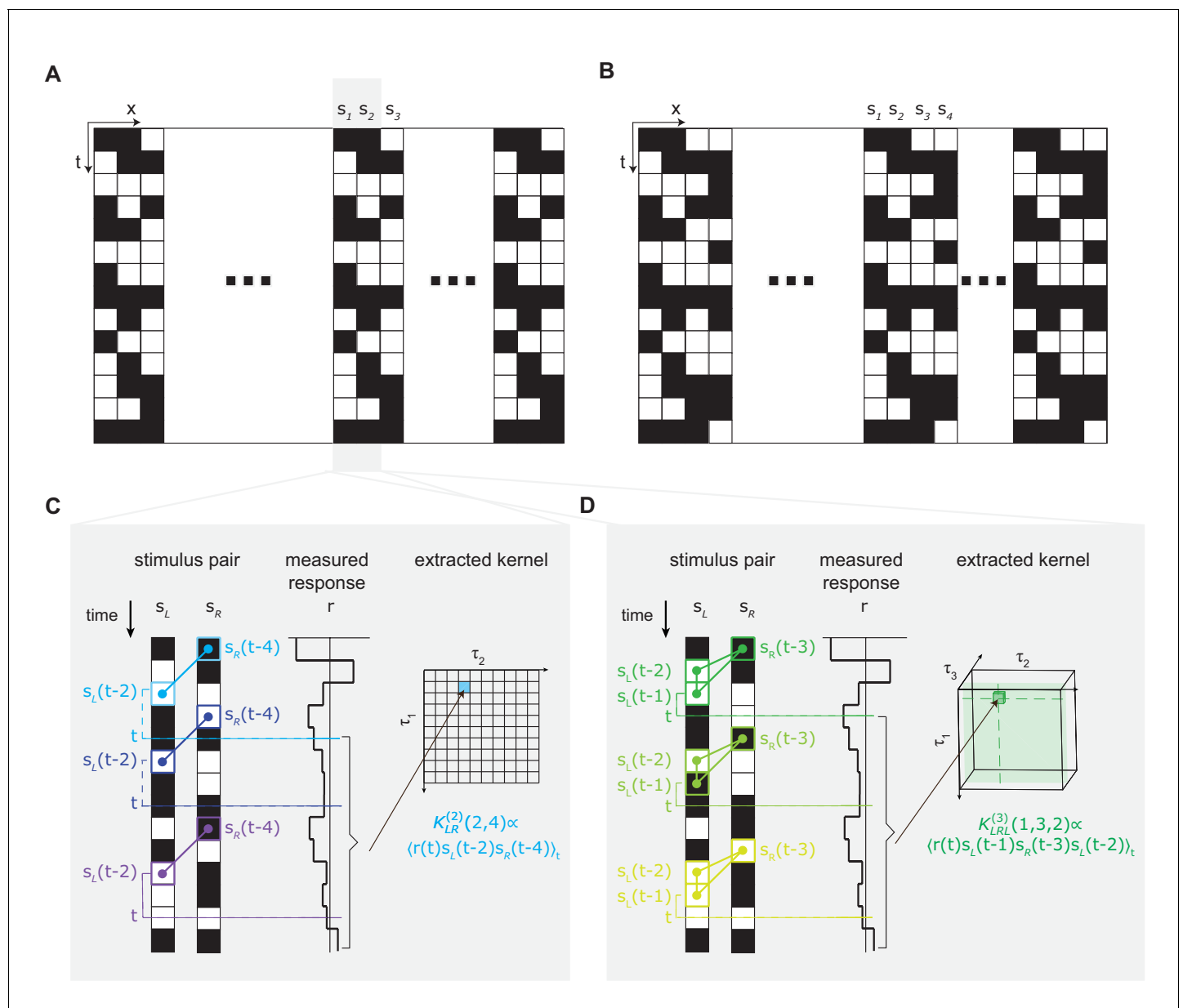


Figure 2—figure supplement 1. Using reverse-correlation to extract second- and third-order kernels from the measured turning response to stochastic binary stimulus (Materials and methods, Appendix 1). (A) The three-bar-block stochastic binary stimulus (Materials and methods). The screen is 270° wide, and discretized into 5° bars. Within each three-bar-block, contrasts of three bars were randomly and independently sampled to be black or white over space and time. The block repeats itself across space 18 times. (B) The four-bar-block stochastic binary stimulus (Materials and methods), and block repeated itself across space 14.5 times. (C) Extracting the second-order kernel using reverse correlation. Left: visual stimulus at two neighboring spatial locations s_L and s_R , and examples of a second-order feature $s_L(t-2)s_R(t-4)$ at three different times t . Middle: Measured optomotor turning response of a fly. Right: To calculate a given element ($\tau_1 = 2$, $\tau_2 = 4$) in the second-order kernel, we calculated the product between the turning response $r(t)$ at time t and the corresponding second-order feature $s_L(t-\tau_1)s_R(t-\tau_2)$, and averaged the products over all times t . (D) Extracting the third-order kernel with reverse correlation. Left: visual stimulus at two neighboring spatial locations, s_L and s_R , and examples of the third-order feature $s_L(t-1)s_R(t-3)s_L(t-2)$ at three different times t . Middle: Measured optomotor turning response of a fly. Right: To measure a given element ($\tau_1 = 1$, $\tau_2 = 3$, $\tau_3 = 2$) in the third-order kernel, we calculated the product between the turning response $r(t)$ at time t and the corresponding third-order feature $s_L(t-\tau_1)s_R(t-\tau_2)s_L(t-\tau_3)$, and averaged the products over all times t .

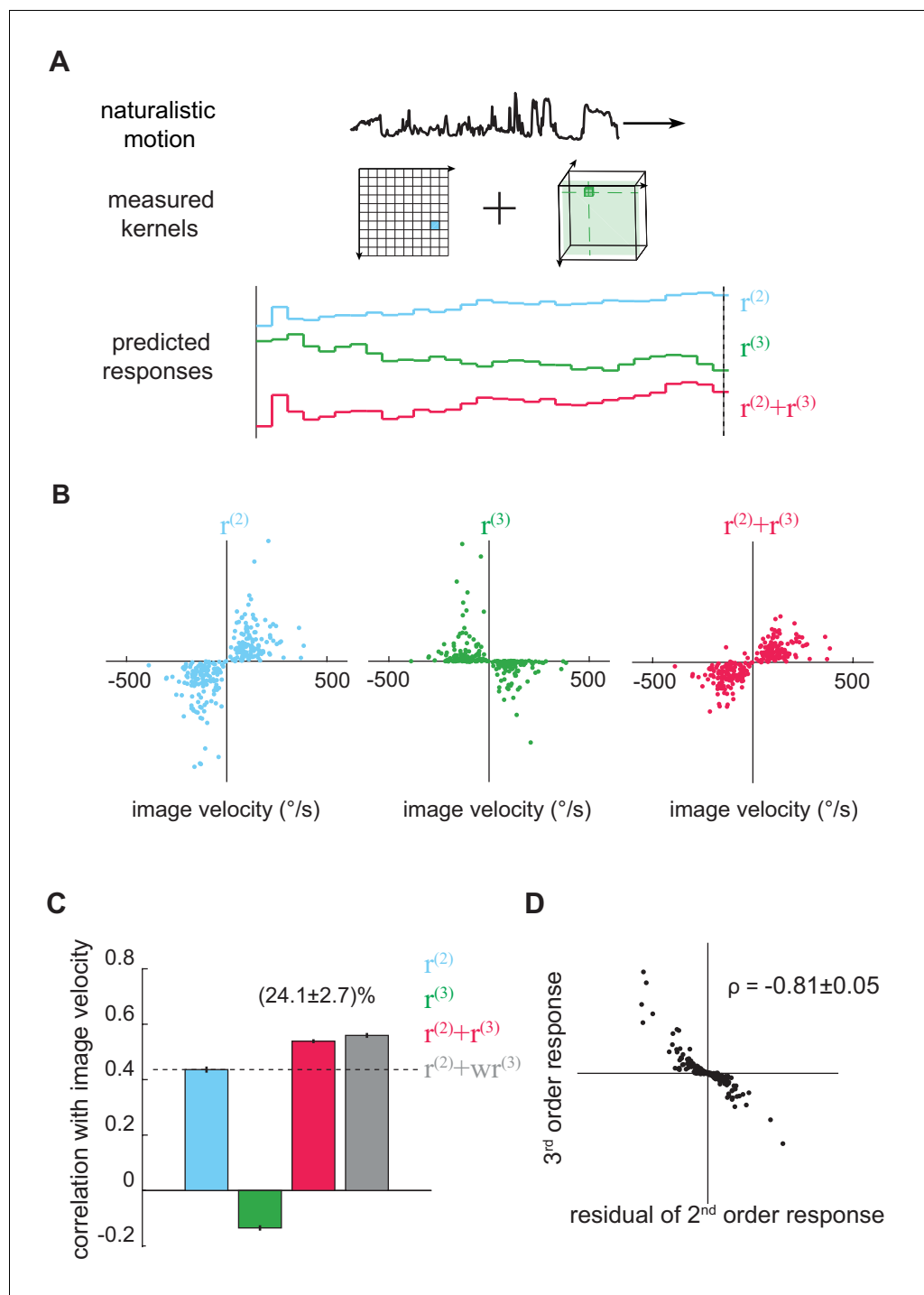


Figure 3. The third-order kernel improved motion estimation in natural scenes. (A) Predicting responses of the second- and third-order kernels to rigidly moving scenes. *Top*: natural scenes rigidly translating with constant velocities. *Middle*: cartoon of the second- and third-order kernels. *Bottom*: second-order response (blue), third-order response (green), the predicted motion estimate (red) is the summation of $r^{(2)}$ and $r^{(3)}$. (B) Scatter plot of $r^{(2)}$, $r^{(3)}$ and $r^{(2)} + r^{(3)}$ against image velocity over the ensemble of moving images. 10,000 independent trials were simulated, and 1000 trials were plotted here. (C) Pearson correlation coefficients between responses of each kernel and the true image velocities ($\rho = 0.44 \pm 0.01$, -0.14 ± 0.01 , 0.54 ± 0.01 , 0.56 ± 0.01 , from left to right; $w = 1.39 \pm 0.01$; mean \pm SEM across 10 groups of 1000 trials). (D) Scatter plot between $r^{(3)}$ and the residual in $r^{(2)}$, computed by subtracting a scaled image velocity from $r^{(2)}$ (Materials and methods). ρ represents the Pearson correlation coefficient mean \pm SEM across 10 groups (Materials and methods).

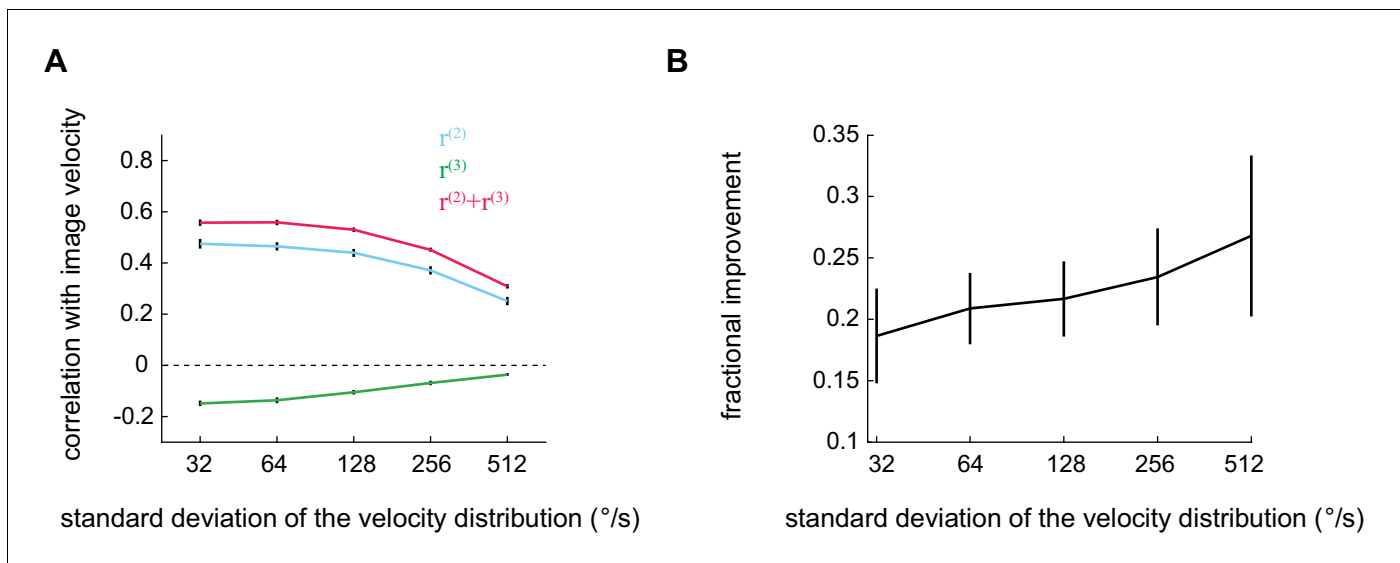


Figure 3—figure supplement 1. The improvement added by the third-order kernel persists across a wide range of velocities. (A) Pearson correlation coefficient between the true image velocity and the second-order, third-order, and full outputs as a function of the standard deviation of the velocity distribution. (B) The improvement added by the third-order kernel as a function of the velocity standard deviation.

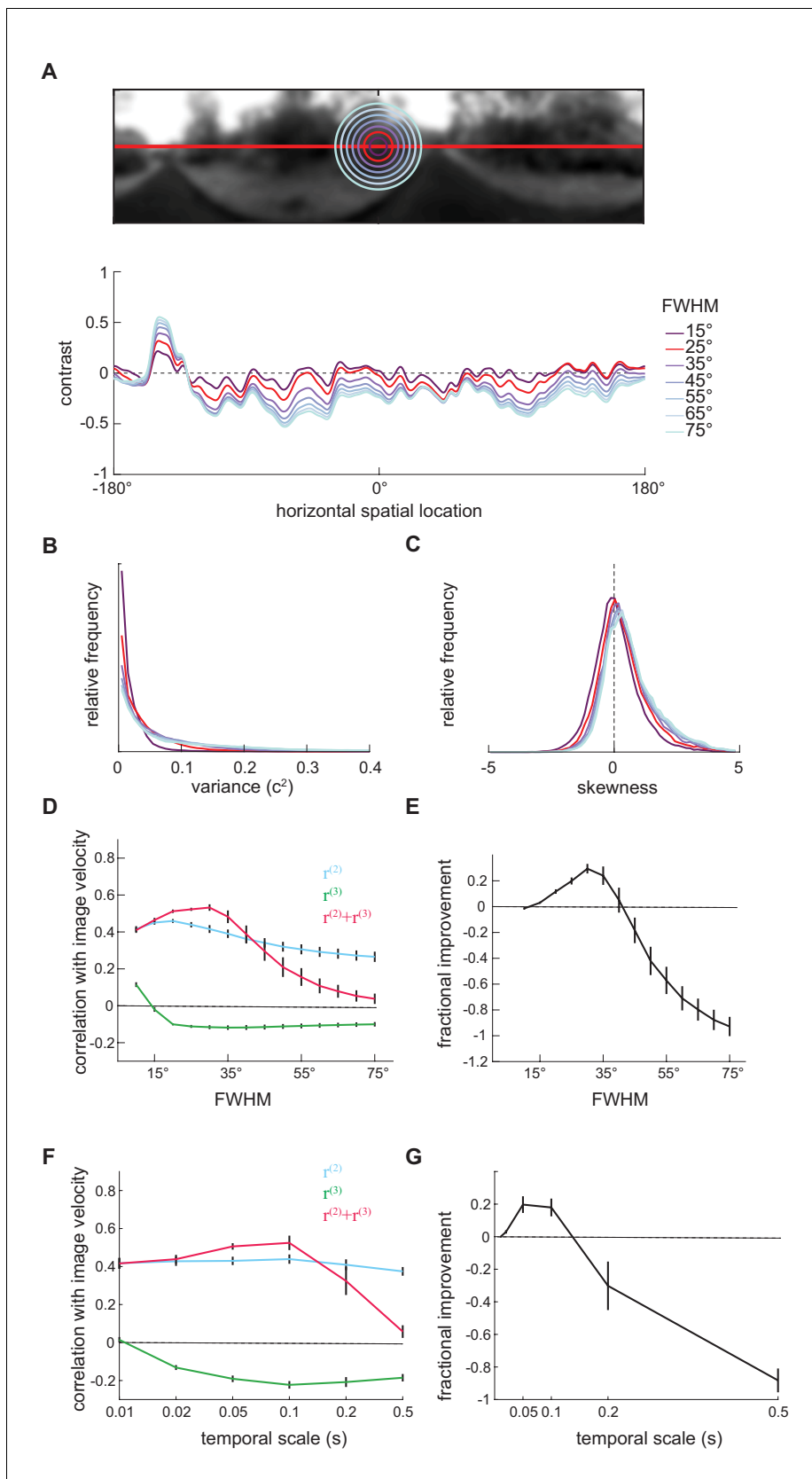


Figure 3—figure supplement 2. The length scale of local mean luminance computation affected the performance of the measured kernels. (A) Local mean luminance was computed by filtering the luminance picture with a Gaussian filter (Materials and methods). We varied the full-width-half-maximum

Figure 3—figure supplement 2 continued on next page

Figure 3—figure supplement 2 continued

(FWHM) of this Gaussian filter from 15° to 75°. *Top*: Example two-dimensional blurred luminance photographs I_{blur} . The radius of the circle represents the length scale of the filter. *Bottom*: Example one-dimensional contrast images computed with corresponding filters. (B) As the length scale of the Gaussian filter increased, the distribution of the variances of individual images shifted to higher values. (C) As the length scale of the Gaussian filter increased, the distribution of the skewness of individual images shifted to more positive values. (D) The performance of measured kernels changed as the length scale of the Gaussian filter increased. (E) The improvement added by the third-order response to the second-order response existed when FWHM varied from 15° to 40°, and peaked around 30°. (F) The performance of second- and third-order responses as a function of the timescale of the contrast computation. (G) The degree of improvement added by the third-order responses changed with the timescale of the contrast computation.

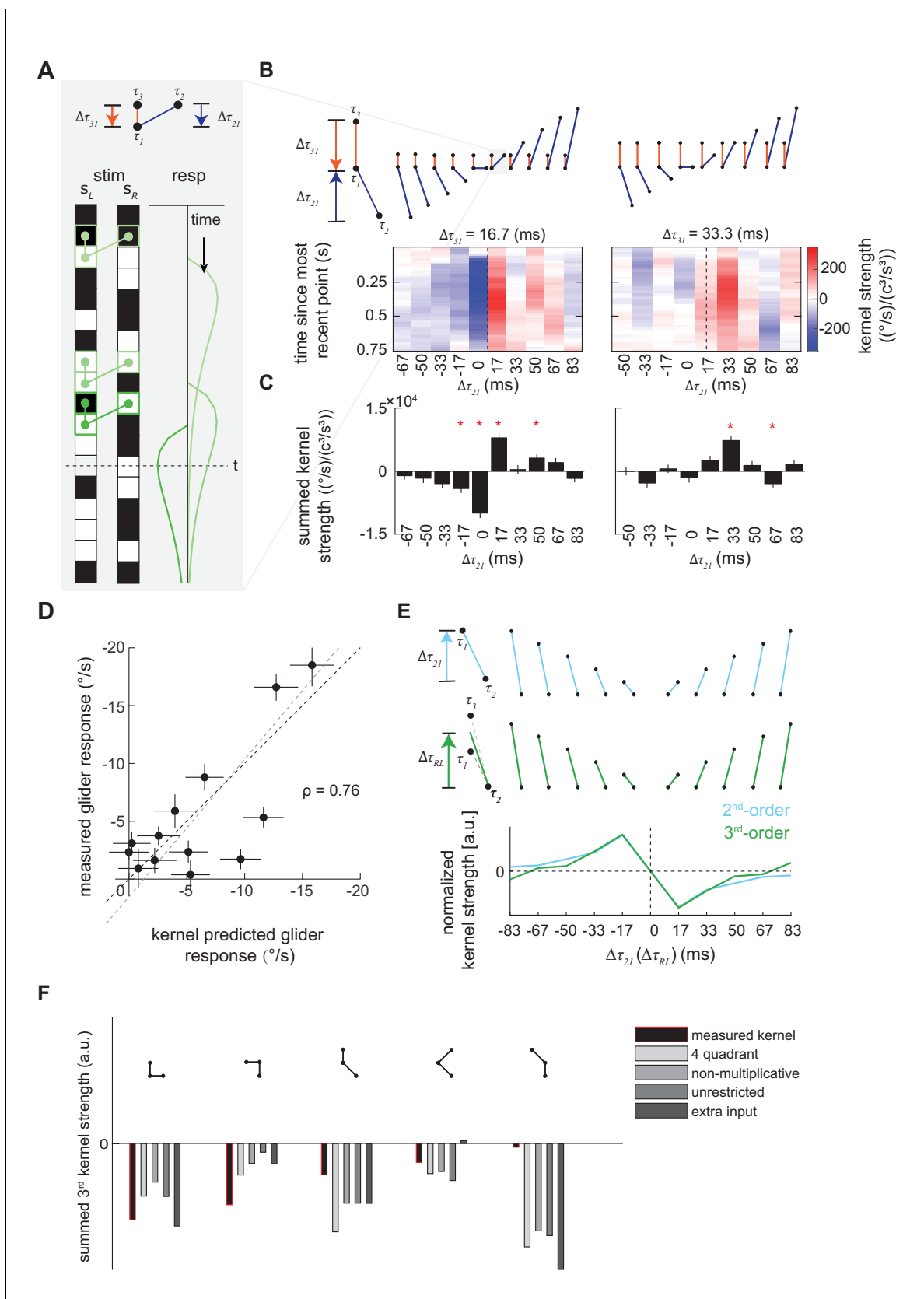
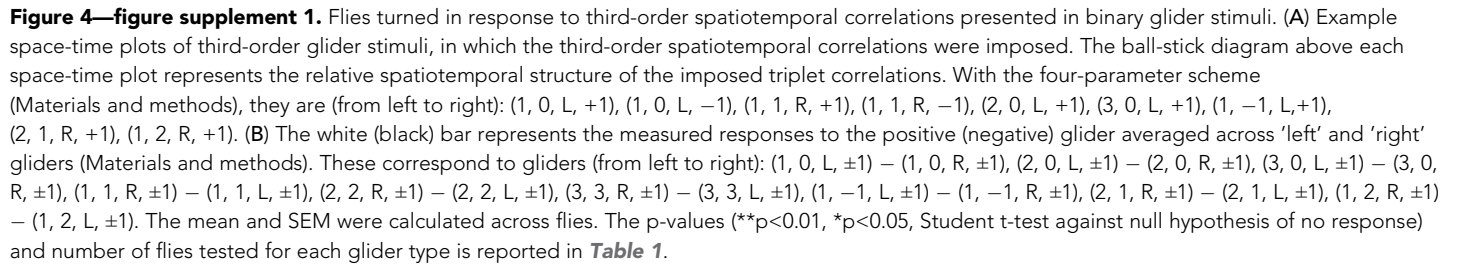


Figure 4. Characterization and validation of the measured third-order kernel. (A) Triplet impulse response description. Top: the ball-stick diagram represents the relative spatiotemporal position of three points in a triplet. The red line denotes the temporal distance between the two left points, $\Delta\tau_{31}$. Figure 4 continued on next page

Figure 4 continued

and the blue line denotes the temporal distance between the more recent point on the left and the sole right point, Δt_{21} . *Bottom*: Three specific example occurrences of the triplet elicit three impulse responses. The response at time t is the sum of the impulse responses to all previous occurrences of the triplet. The first triplet (*lightest green*) involves two black points and one white point, so their product is positive, and it elicits an impulse response with positive sign. The triplet occurs far from current time t , so its influence on the current response is small. The last triplet (*darkest green*) involves two white points and one black point, so the product is negative and it elicits an impulse response with flipped sign. It is close to current time t , and has a large influence on the current response. **(B)** Third-order kernel visualized using an impulse-response format (Materials and methods). *Top*: the ball-stick diagrams as in (A). *Bottom*: the color map plots the 'impulse response' to the corresponding triplets, and color represents the strength of the kernel. Different panels represent different Δt_{31} . In each color map, Δt_{31} is fixed, the columns represent Δt_{21} , and the rows represent the time since the most recent point in each triplet. The dashed lines indicate the place where the right point is in the middle of the two left points in time. **(C)** The summed strength of the third-order kernel along each column in A. Error bars represent SEM calculated across flies ($n = 72$), and significance was tested against the null kernel distribution ($*p < 0.05$, two-tailed z-test (Materials and methods)). **(D)** The scatter plot between the measured responses to third-order glider stimuli (Materials and methods, **Figure 4—figure supplement 1**) against responses predicted by the third-order kernel (Materials and methods). The correlation between the predicted and measured responses is 0.76. Black dashed line is unity; gray dashed line is the best linear fit. **(E)** The measured second- and third-order kernel share temporal structures. *Top*: the ball-stick diagrams represent the relative spatiotemporal positions of the two points in each pair (*blue*), and three points in each triplet (*green*). *Bottom*: The kernel strength of the second-order kernel (*blue*) and third-order kernel (*green*) summed across all elements sharing the same spatiotemporal structures, that is summed over rows in **Figure 4—figure supplement 2CD** (Materials and methods). **Figure 4—figure supplement 3**, **(F)** The extracted third-order kernels from four optimized motion detectors (Fitzgerald and Clark, 2015) compared to the measured kernel from the fly. The summed kernel strength is summed across all elements which shared the same spatiotemporal structures diagramed above.



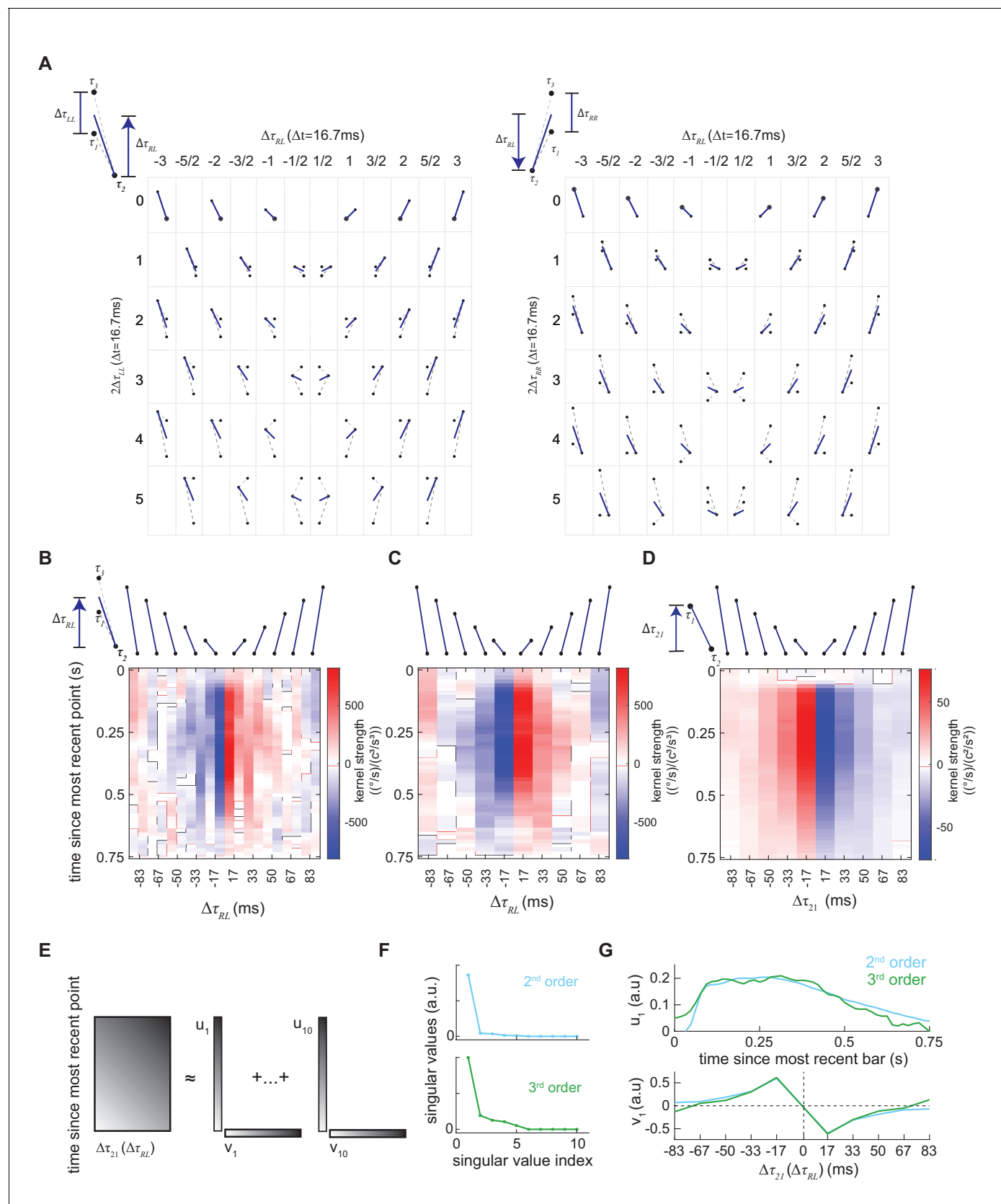


Figure 4—figure supplement 2. Rearranging the second- and third-order kernels and computing singular value decompositions on the rearranged kernels. (A) Diagrams to demonstrate how the third-order kernel was rearranged. The 'ball-stick' diagram represents the spatiotemporal structure of the Figure 4—figure supplement 2 continued on next page

Figure 4—figure supplement 2 continued

triplet correlations. *Left*: $\Delta\tau_{LR}$ is the average time difference between the left points and the right point, and $2\Delta\tau_{LL}$ is the temporal separation between two left points. In the grid, the rows represent $2\Delta\tau_{LL}$ in the unit of frame $\Delta t = 16.7$ ms, and the columns represent $-\Delta\tau_{LR}$. In the first row, the temporal distance between two left points is zero, so two left points overlap with each other, which is represented by a solid dot surrounded by an empty circle. The z-axis, which is not shown here, represents times since the most recent point. *Right*: Similar to *left* panel, but in this case, there are two points on the right and one point on the left. In the grid, the rows represent $2\Delta\tau_{RR}$, and the columns represent $\Delta\tau_{RL} = -\Delta\tau_{LR}$. Positive $\Delta\tau_{RL}$ means the left point is more recent than the right points (Materials and methods). (B) Two-dimensional representation of the third-order kernel by summing the rearranged third-order kernel elements in (A) and (B) over $\Delta\tau_{RR}$ and $\Delta\tau_{LL}$ (Materials and methods). In this new format, rows represent time since the most recent point, and columns describe the temporal distance between right and left points, with negative intervals meaning that the right point is more recent than the left point. (C) Two-dimensional representation of the third-order kernel created by summing neighboring columns in (B). (D) Second-order kernel visualized using an impulse-response format (Materials and methods). *Top*: the ball-stick diagrams represent the relative spatiotemporal structure of the pairwise spatiotemporal correlations. *Bottom*: the color map denotes the 'impulse response' to the corresponding pairs, and the color represents the kernel strength, the columns represent $\Delta\tau_{21}$, and the rows represent the time since the most recent point in the pair. (E) Diagram of the singular value decomposition of a matrix. The matrix can be approximated by a series of outer products of vectors. (F) The singular values of the second-order kernel (*blue*) and the third-order kernel (*green*). (G) The optomotor turning response of the fly can be modeled as a correlation-sensitive module that represents motion computation and a low-pass filter module that represents the dynamical change of the turning behavior (Clark et al., 2011). The left vector (*top*) could be interpreted as the dynamics in a behavioral module and the right vector (*bottom*) as a correlation-sensitivity module.

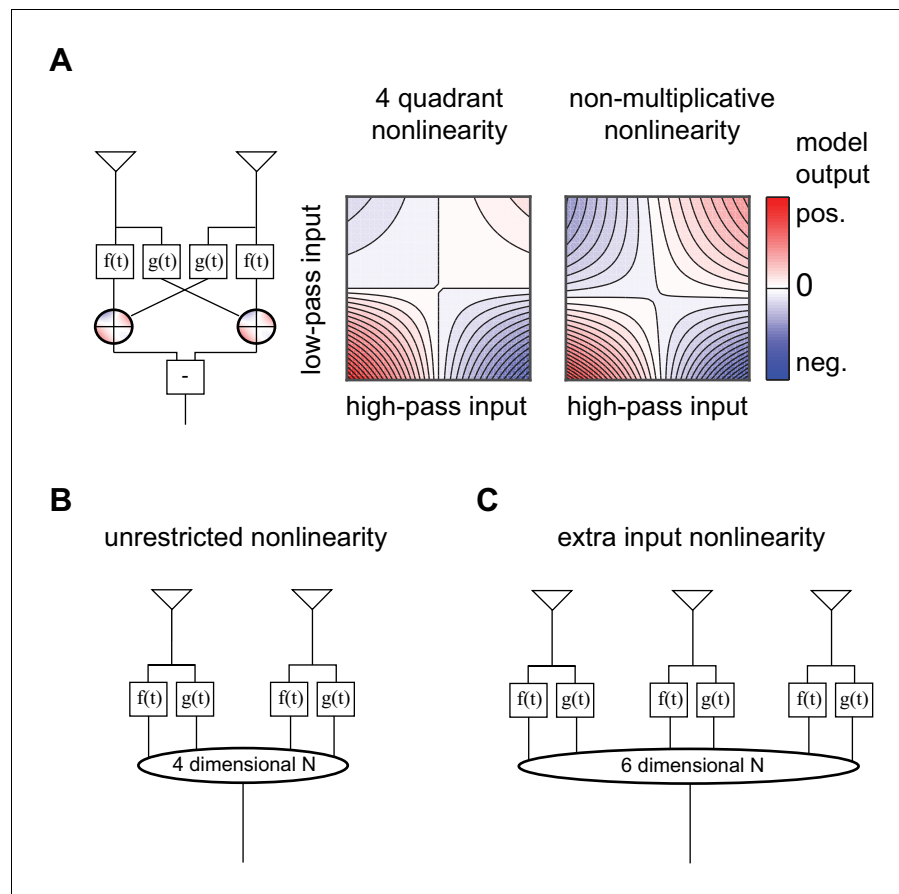


Figure 4—figure supplement 3. Four models optimized to estimate image velocities in natural scenes, adapted from *Fitzgerald and Clark (2015)*. (A) Four-quadrant and non-multiplicative model, these two models differ in the nonlinear interaction of the high- and low-pass inputs. (B) Unrestricted nonlinearity model, in which all interactions up to fourth-order were permitted between the four inputs from two points in space. (C) Extra input model, in which all interactions up to fourth-order were permitted between the six inputs from three points in space.

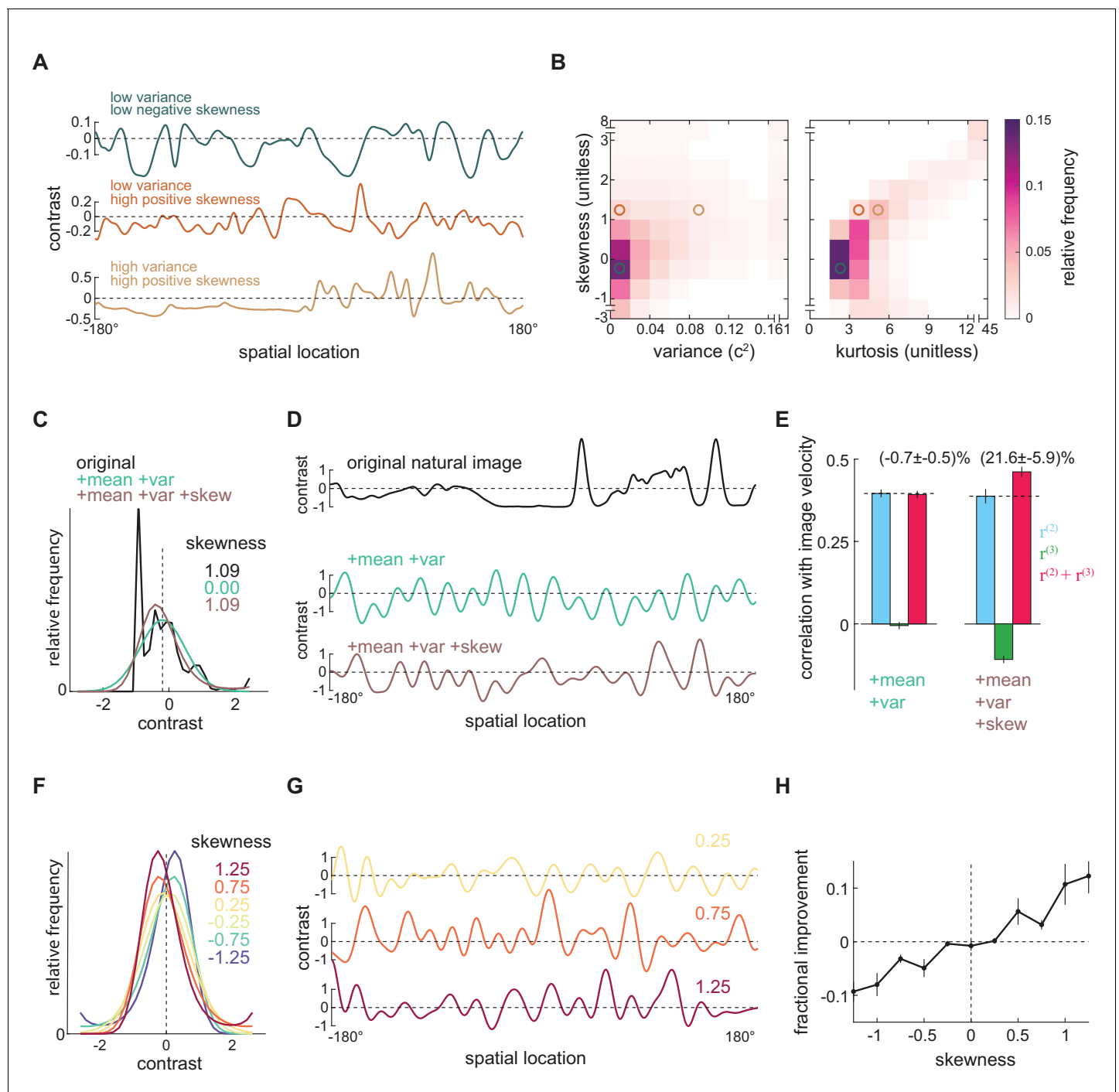


Figure 5. Positive skewness is sufficient for the third-order kernel to improve motion estimation. (A) Three example natural scenes with different degrees of variance and skewness. (B) Joint-density maps of individual image statistics over the ensemble of natural images, showing the relationship between skewness and variance (left) and skewness and kurtosis (right). (C) Contrast distributions of example images from the natural scene dataset and two synthetic image datasets. The natural image is shown (black), along with a maximum entropy distribution (MED) with matched mean and variance, denoted by +mean +var (green) and an MED with matched mean, variance, and skewness, denoted by +mean +var +skew (brown). (D) Example images from the natural scene dataset and two synthetic image datasets corresponding to three contrast distributions in (C). (E) The Pearson correlation coefficient between true image velocities and each kernel's responses in the two synthetic datasets +mean +var (green) and +mean +var +skew (brown). (F) Example of MEDs in six synthetic datasets, in which the image skewness ranged from -1.25 to 1.25. (G) Example of synthetic images in three synthetic datasets, corresponding to MEDs in (F) with constrained skewness of 0.25 (top), 0.75 (middle), and 1.25 (bottom). (H) Improvement added by the third-order response as a function of synthetic image skewness.

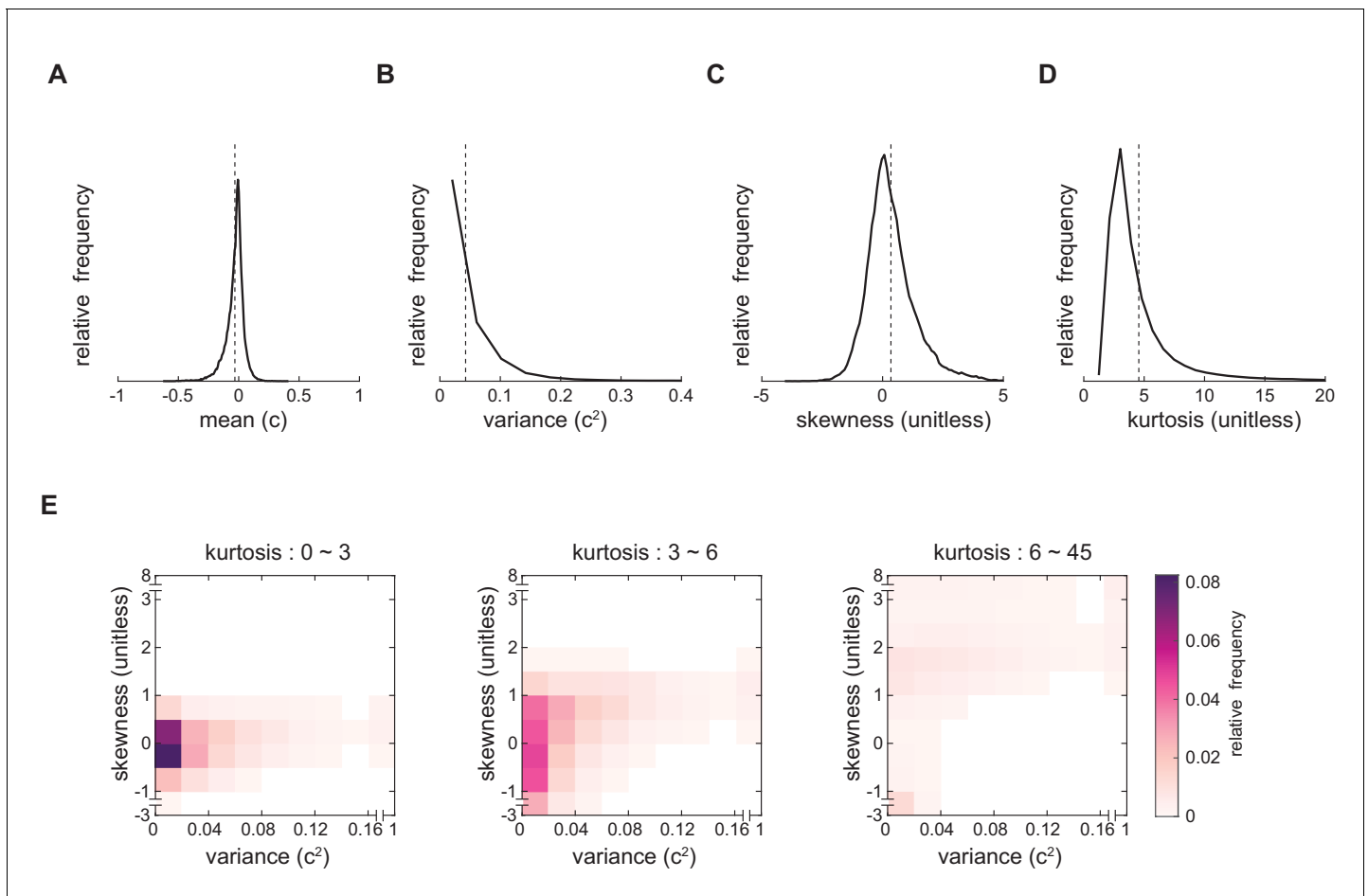


Figure 5—figure supplement 1. Natural scenes have heterogeneous contrast statistics. (A–D) Distribution of the contrast mean, variance, skewness and kurtosis of individual images within the entire natural scene ensemble. Mean is denoted by the vertical dashed line. (E) Image statistics are correlated with each other. Joint-density maps of variance and skewness with different ranges of kurtosis.

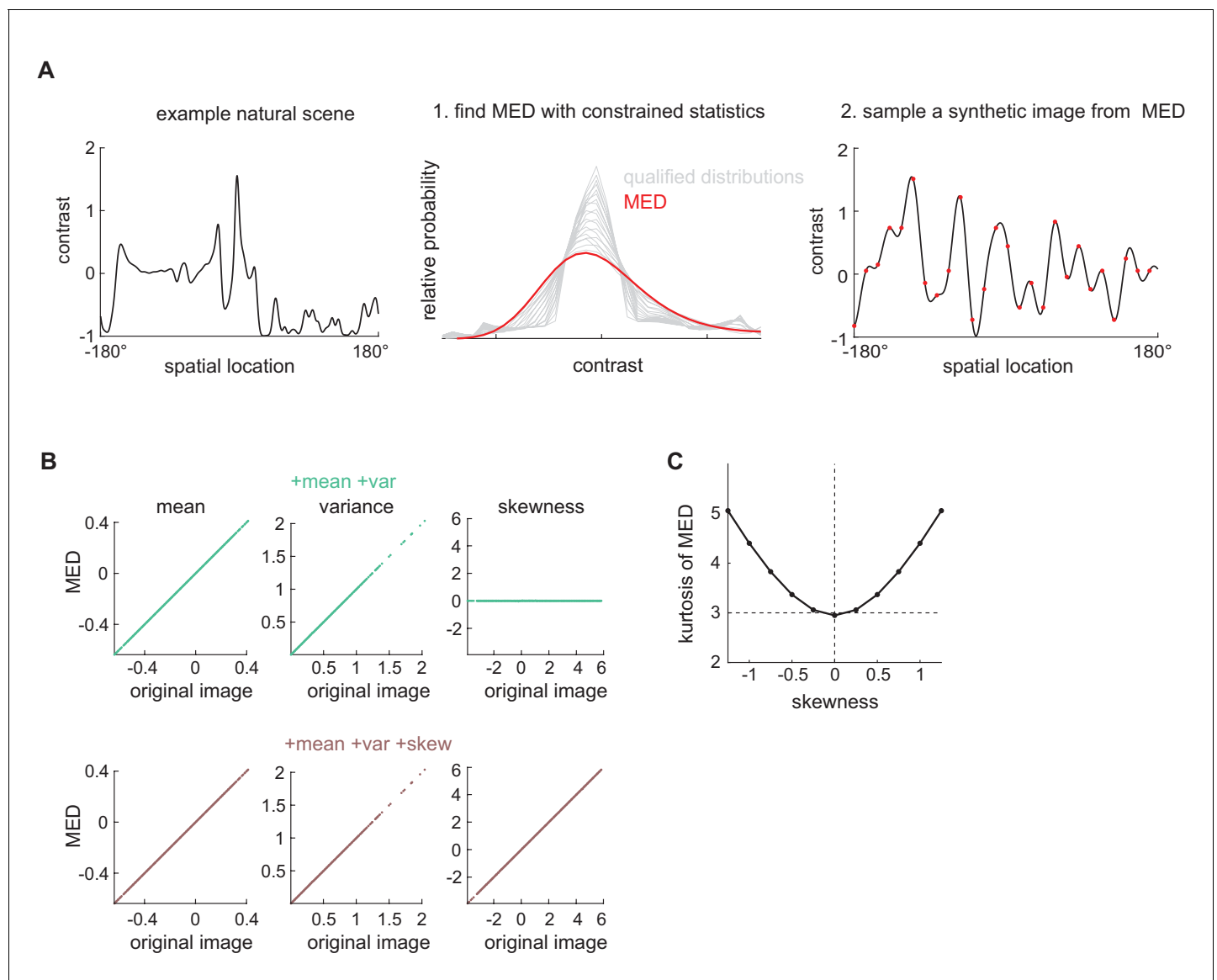


Figure 5—figure supplement 2. Using maximum entropy distributions (MEDs) to generate synthetic images with controlled image statistics. (A) Two steps to generate one synthetic image with desired contrast statistics (Materials and methods, Appendix 2). *Left:* Select a natural image and calculate the mean, variance, and skewness of its contrast. *Middle:* There is a family of contrast distributions (gray) that have the same mean, variance, and skewness as the original natural image. One may use an optimization algorithm to find the distribution with the maximum entropy (red). *Right:* Sample-independent contrast values from the solved MED and interpolate between them to induce local spatial correlations. (B) Validating that the MEDs possessed the constrained statistics. *Top row:* When mean and variance were constrained, the mean and variance of the MEDs matched those of the original images. The skewness of the MEDs was zero. *Bottom row:* When mean, variance, and skewness were constrained, the mean, variance, and skewness of the MEDs matched those of the original images. (C) The kurtosis of the MEDs (Figure 5F) with skewness constrained to be different values.

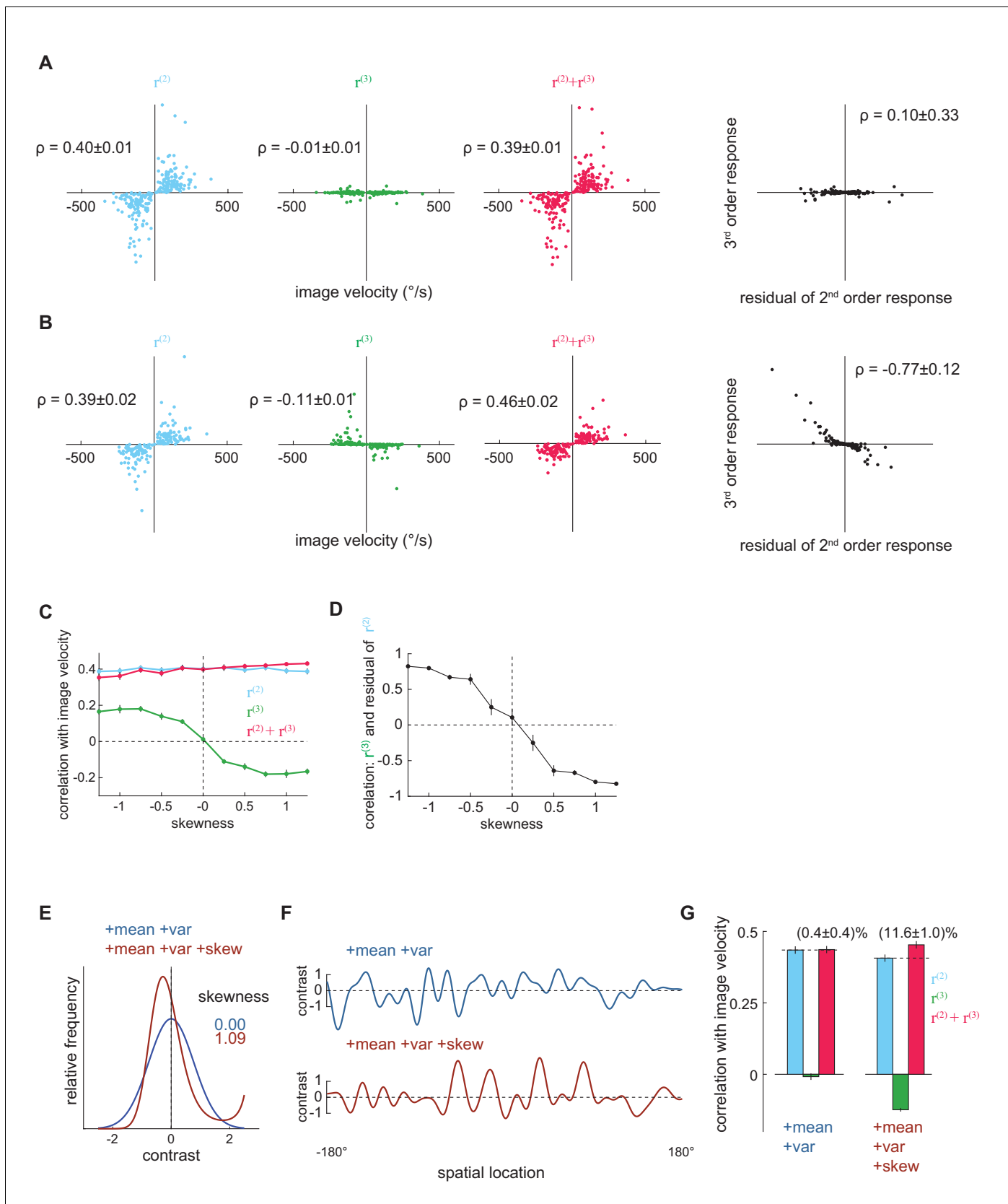
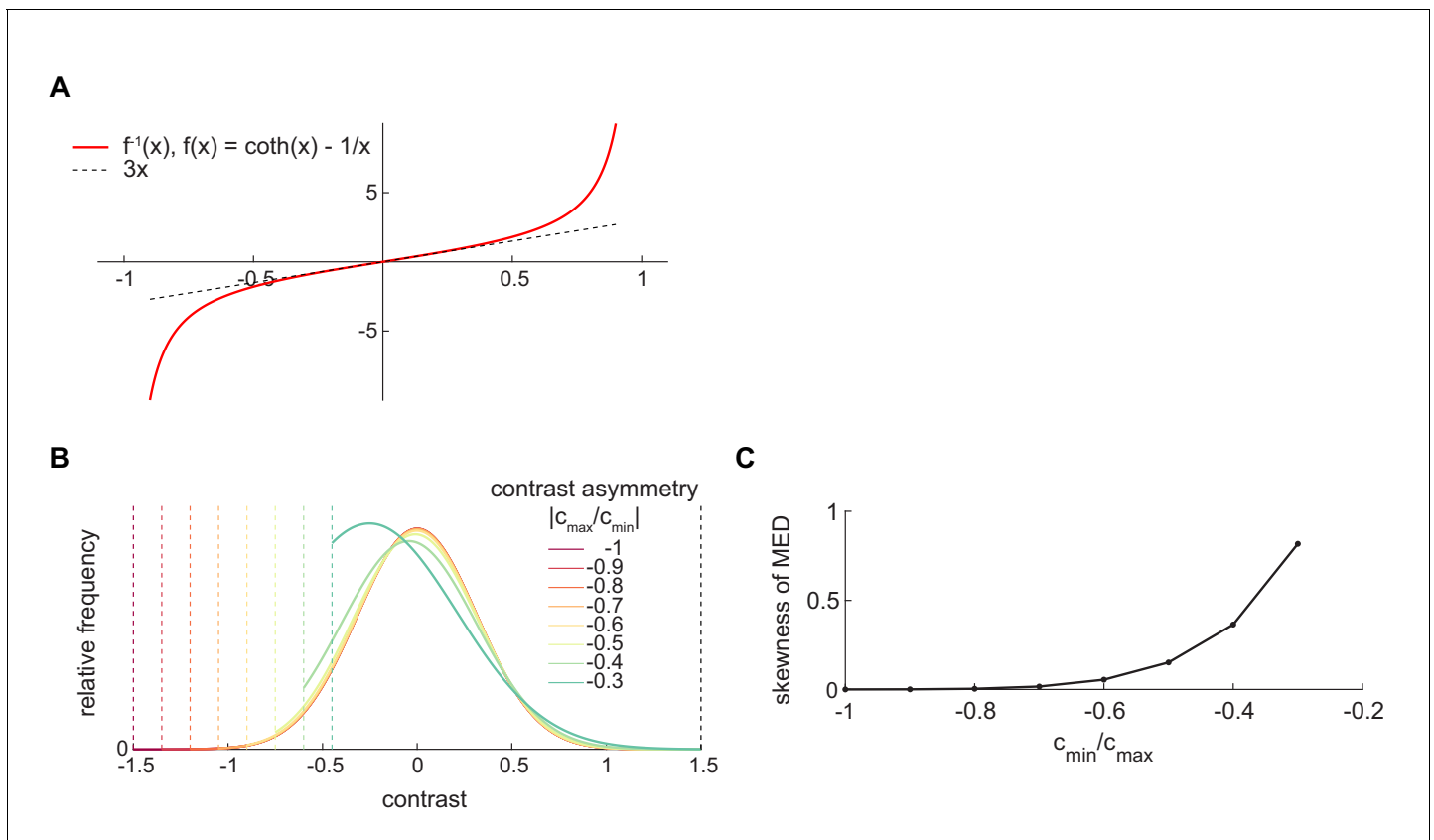


Figure 5—figure supplement 3. The performance of the measured kernels in various synthetic image datasets. (A) Scatter plots between the model responses and true image velocities when we constrained the synthetic images to match only the mean and variance of the natural images. (B) As in (A), Figure 5—figure supplement 3 continued on next page

Figure 5—figure supplement 3 continued

but in this case, we also constrained the synthetic images to match the skewness of the natural image. (C) The performance of the measured kernels as skewness varied in synthetic datasets. (D) The correlation between the third-order responses and the residual in the second-order responses as a function of the skewness in synthetic datasets. (E) In comparison to **Figure 5CD**, where the contrast range of the MEDs matched the natural images, the contrast range of the MEDs here were fixed to be $[-2.5, 2.5]$. (F) Example images from these two distributions. (G) The Pearson correlation coefficients between the true image velocities and each kernel's responses in the two synthetic datasets when the contrast range was fixed.



Appendix 2—figure 1. The shape of MED depends on the contrast range. (A) The shape of function $f^{-1}(x)$, where $f(x) = \coth(x) - 1/x$. (B) The mean-variance-constrained MED depends on the contrast range. The mean is constrained to be 0. The upper bound of the contrast range is fixed at 1.5, and the ratio of the lower-bound to upper bound changes from -1 to -0.2 . As the contrast range becomes more asymmetric, the MED becomes more asymmetric. (C) When the contrast range is asymmetric, the mean-variance-constrained MED has induced non-zero skewness.