Supplementary Materials Overview

The supplementary materials contain three sections: the first is devoted to demonstrating robustness of the main results, the second provides additional information on the local motion scores, and the third shows how the present method applies to the classic examples of non-Fourier motion introduced by Chubb and Sperling (1988).

The first section of the Supplement expands on the analyses carried out in the main text. Figures Figure S1 and Figure S2 (SM score), Figure S3 (PMO scores) and Figure S4 (RMO scores) show that the results in the main text are robust with respect to the size of the region used for discretization. Figures Figure S5 and Figure S6 show that the results in the main text, which analyzed motion in the YT plane, also hold for motion in the XT plane. Figures Figure S7 and Figure S8 show that the results are robust with respect to the binarization strategy: in the main text, the binarization cutpoint was the median for each movie segment; here we show that similar results are obtained when the threshold is set at the mid-level luminance value of the entire range.

Figures S9-11 show supplementary analyses related to covariation of motion signals. Specifically, Figure S9 shows that the pattern of covariation of motion signals, presented for Movie 1 in the main text, holds for other movies. Figure S10 presents a parallel analysis of a random spatiotemporal noise, showing that the correlations among the motion signals in popular movies arise from the structure of the movies themselves, not the analysis procedure. Finally, Figure S11 shows that the pattern of covariation of motion signals is robust with respect to the binarization strategy when the threshold is set at the mid-level luminance value of the entire range.

The second section of the Supplement provides mathematical background on the motion scores. In Figure S12, we compare the RMO and PMO scores to the output of the Reichardt model for Fourier (F) motion, and show that for RMO, the correspondence is precise. We then demonstrate some mathematical relationships between the scores: (a) that for the subtypes of glider (G) motion, the RMO scores must be identical (but the PMO scores need not be), and (b) that for the two kinds of pure non-Fourier (NF) motion, the PMO scores must be identical (but the RMO scores need not be).

The third section of the Supplement describes how the method applies to the non-Fourier motion stimuli of Chubb and Sperling (1988), and in so doing, discusses a technical point that is important when applying our analysis strategy to movies that are generated in the laboratory (but is not relevant analyzing digitized naturalistic movies). Specifically, it is crucial to recognize that the grid used to generate the stimulus and the grid used to analyze it are separate entities. All glider placements – not only those that are in register with the stimulus grid, but also the generic placements that are out of register – are relevant to the motion computation. If only the special, in-register placements are considered, the motion signal could be overlooked, but when this caveat is taken into account, classic non-Fourier motion signals are captured.

Section 1: Supplementary analyses of motion scores

Intermediate levels of discretization: SM score

In most analyses in the main text, each check of the 1x4x4 (or 4x1x4) ROI used to calculate motion scores corresponded to a single pixel of the movie in the original database. Here, we repeat the analysis with using other resolutions, downsampling an n x n block of pixels (n=2, 4, 8, and 16) to a single check. Downsampling was carried out by averaging the original movie over each n x n block.

Figures Figure S1 and Figure S2 show the SM score for these intermediate levels of discretization. Figure S1 shows the analysis based on raw luminance values, while Figure S2 shows the SM score following binarization at the median. In all cases, F scores are the lowest. G scores increasingly dominate as the region of discretization enlarges.
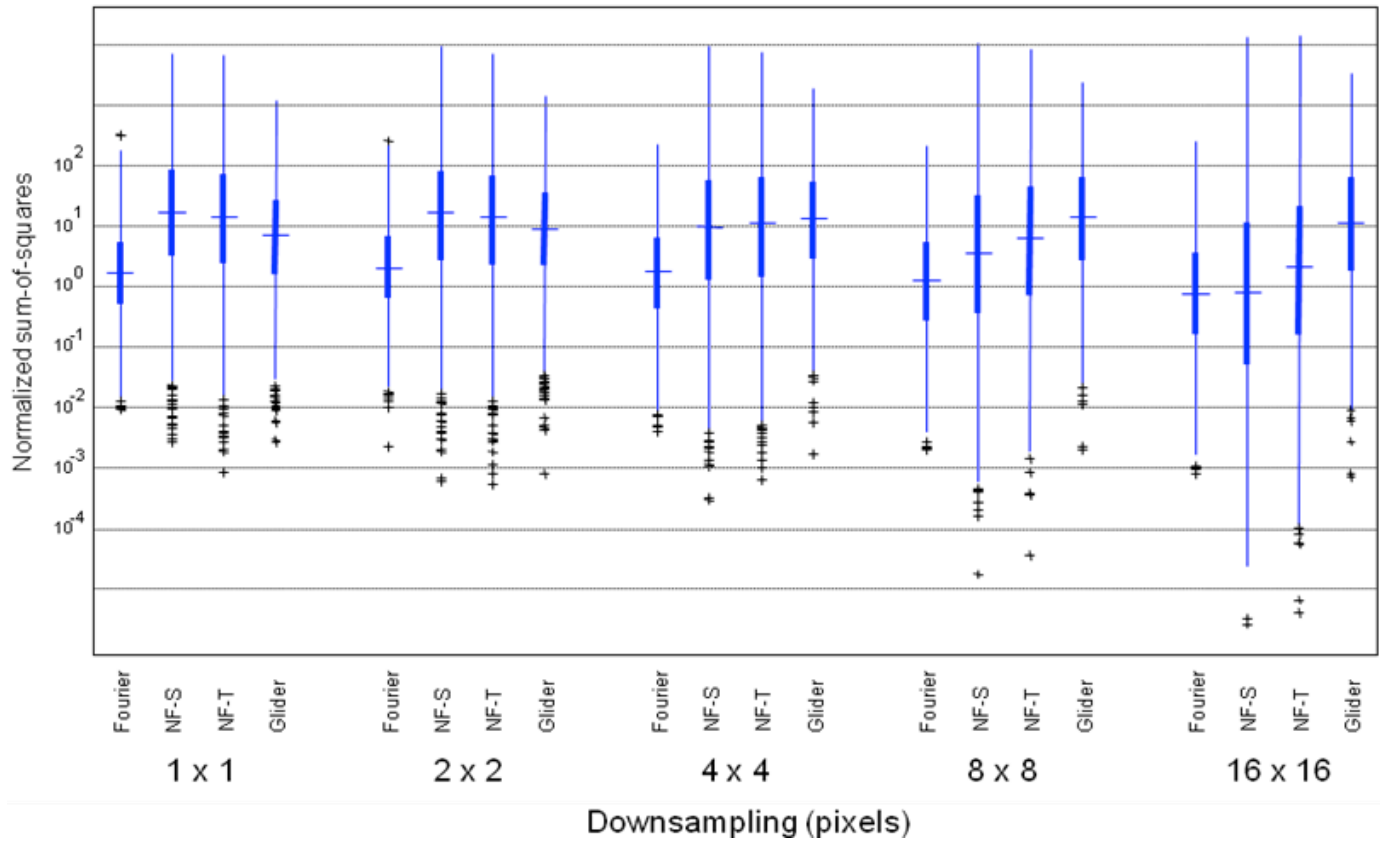
Figure S1: Dependence of motion signals on spatial discretization, for the different motion scores. Motion scores were calculated for each scene, and the distribution is summarized by the median (horizontal line), the interquartile range (heavy vertical line), the "whiskers" (thin vertical line, covering 4x the interquartile range, up to the closest data point) and the outliers (individual symbols, outside the range of the whiskers). Values are normalized by scores obtained from a movie of random pixels. A single movie (movie 4 the main text, "Mr. & Mrs. Smith") was re-analyzed at the original resolution and at four levels of downsampling, in which a single check in the ROI corresponded to an *n* x *n* block of pixels in the original movie. The level of downsampling ranged from 1x1 (no downsampling, corresponding to Figure 3 of the main text) to 16x16 (corresponding to Figure 4 of the main text). The ROI was oriented in the YT plane. Original movies' inputs were presented at 256 x 256 pixels per frame and 24 frames per second.
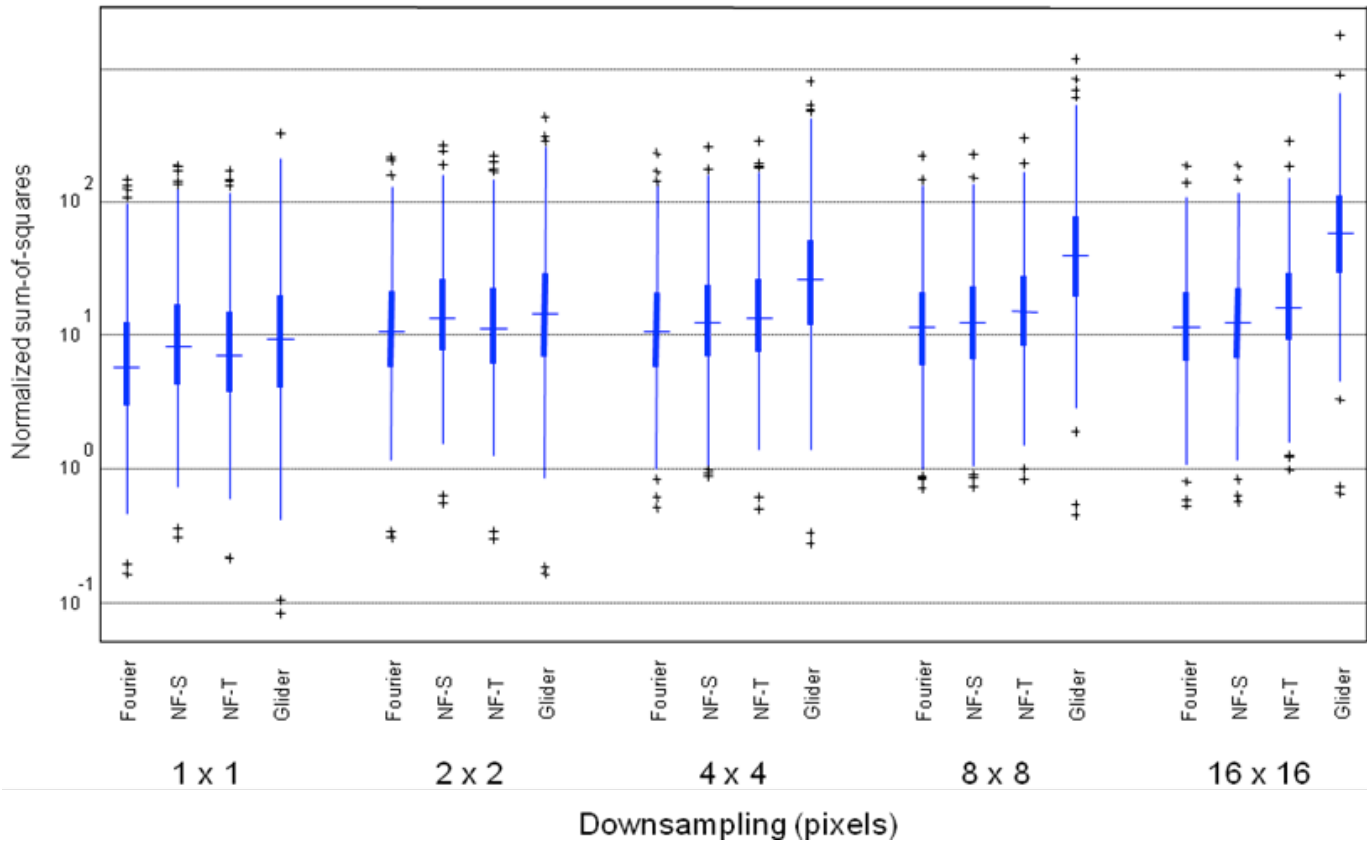
Figure S2: Dependence of motion signals on spatial discretization, for the different motion scores. The level of downsampling ranged from 1x1 (no downsampling, corresponding to Figure 6 of the main text) to 16x16 (corresponding to Figure 7 of the main text). Following downsampling, luminance values were binarized at the median for each shot. Other details are as presented in Figure S1.

Intermediate levels of discretization: PMO/RMO score

Figures Figure S3 (PMO) and Figure S4 (RMO) examine effects of intermediate levels of downsampling. Overall, the magnitude of each motion score decreases with progressively greater downsampling (i.e., with larger checks used for the analysis), and the ratios between motion scores become more restricted. Nevertheless, the results show that orderings between the scores for the different kinds of motions are preserved.
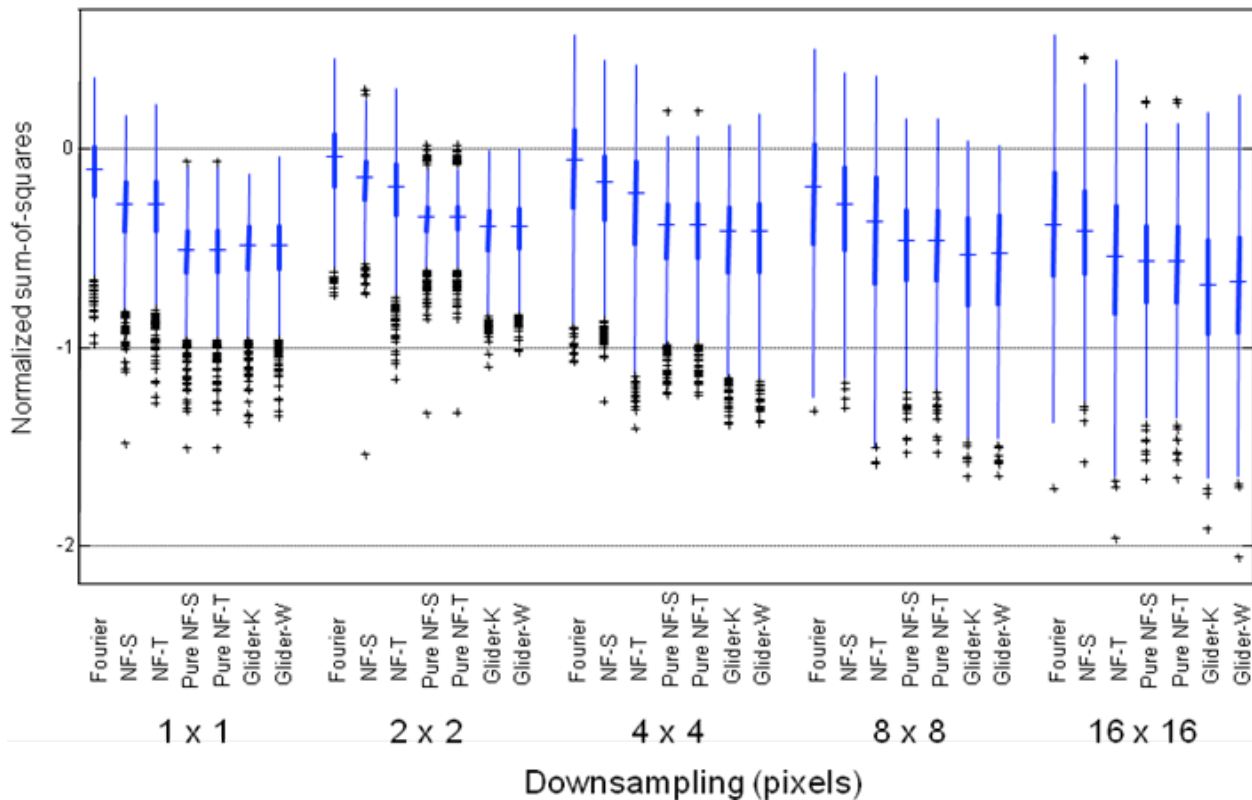

Figure S3: Dependence of motion signals on spatial discretization, for the PMO score. A single movie (movie 4 of the main text, similar movie as in Figure S1, "Mr. & Mrs. Smith") was re-analyzed at the original resolution and at four levels of downsampling, in which a single check in the ROI corresponded to an *n* x *n* block of pixels in the original movie. The level of downsampling ranged from 1x1 (no downsampling, corresponding to Figure 8 of the main text) to 16x16. The ROI was oriented in the YT plane. For other details see Figure S1.
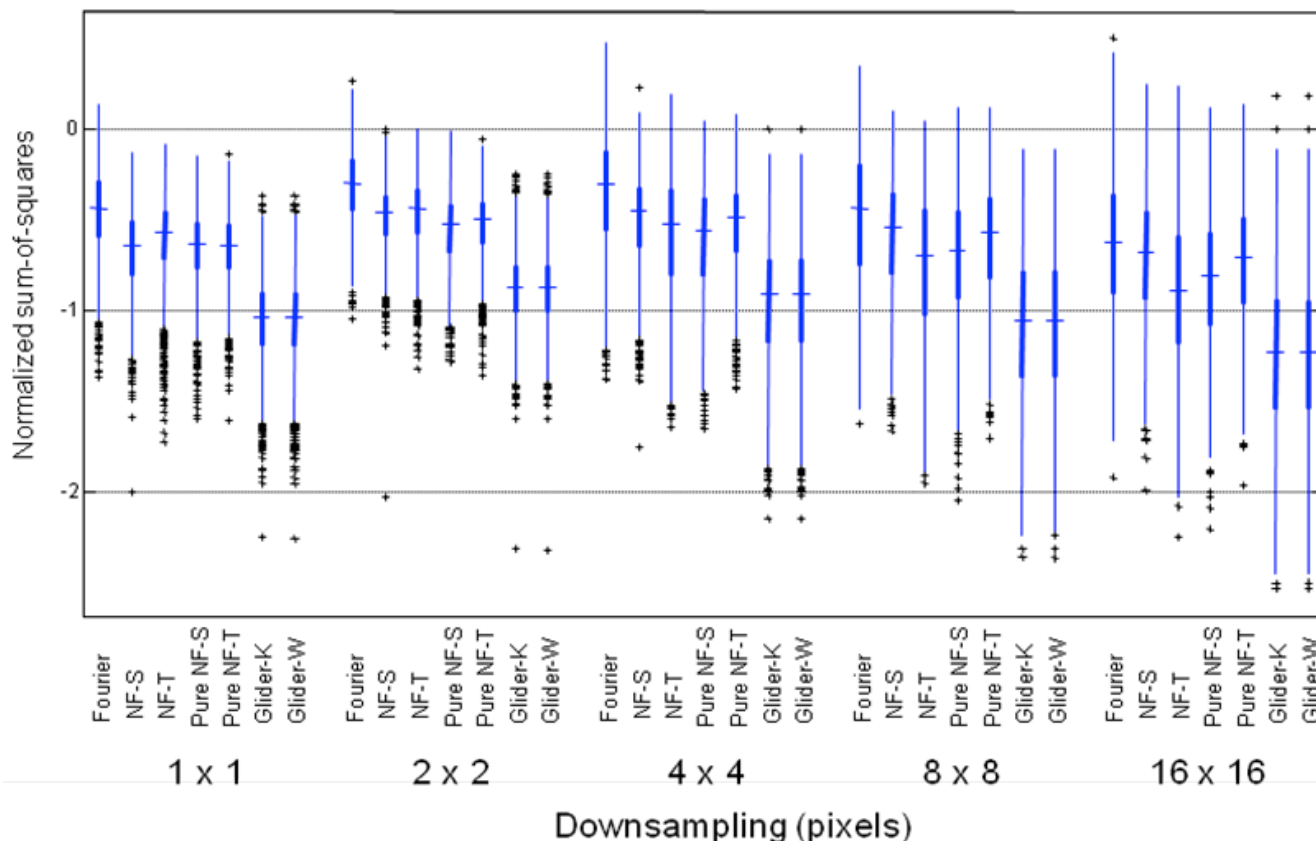
Figure S4: Dependence of motion signals on spatial discretization, for the RMO score. A single movie (movie 4 of the main text, similar movie as in Figure S1, "Mr. & Mrs. Smith") was re-analyzed at the original resolution and at four levels of downsampling, in which a single check in the ROI corresponded to an *n* x *n* block of pixels in the original movie.   The level of downsampling ranged from 1x1 (no downsampling, corresponding to Figure 9 of the main text) to 16x16. The ROI was oriented in the YT plane. For other details see Figure S1.

Orientation of motion direction

In the main text, we showed that there was a common pattern of motion scores for four movies; our analysis used an ROI (region of interest) set in the YT-plane and was therefore sensitive to horizontal motion (Figure 8 for PMO score, Figure 9 for RMO score). Figures Figure S5 and Figure S6 show when the ROI is set in the XT-plane (vertical motion), the pattern of motion scores is similar.
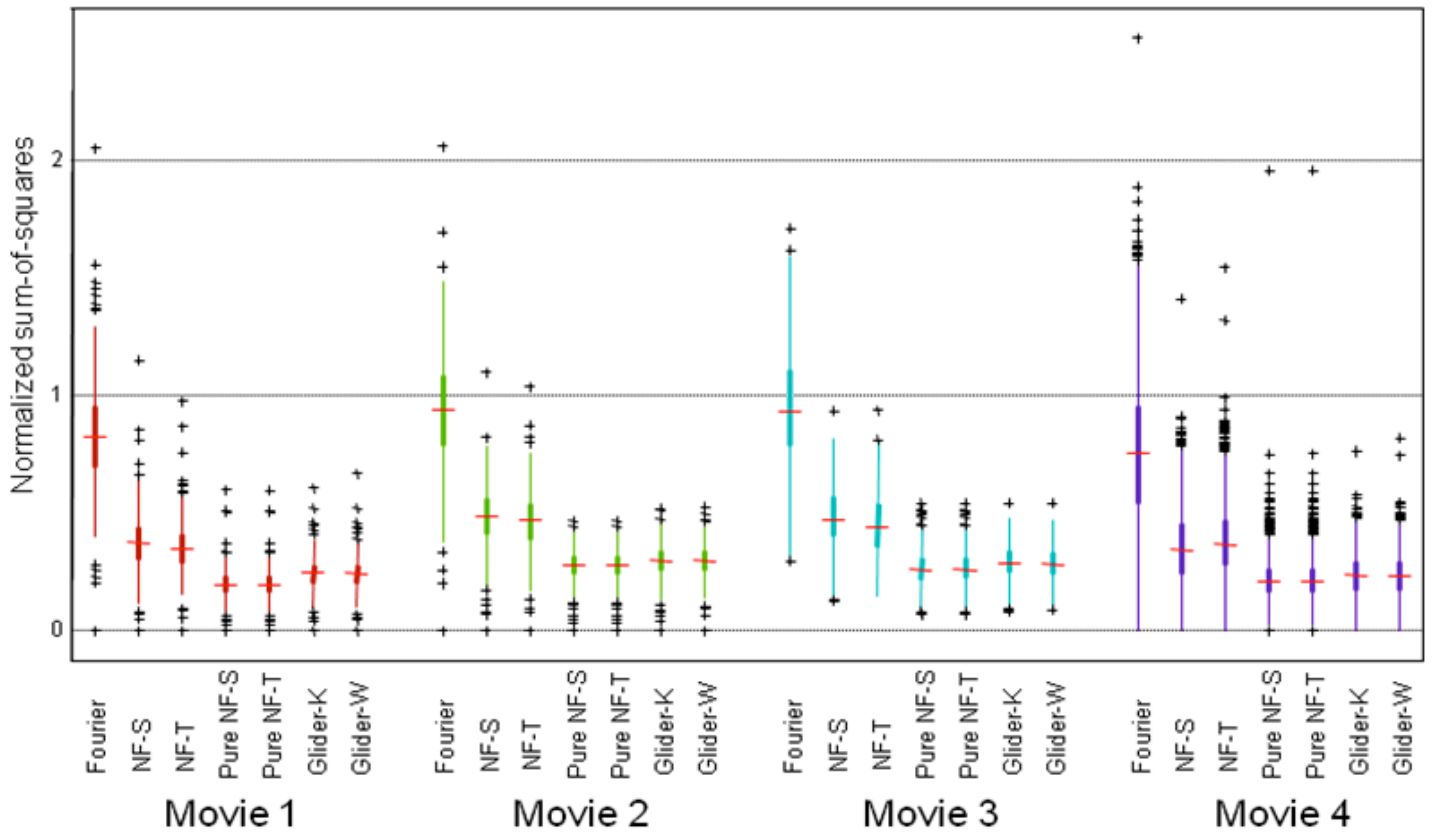
Figure S5: Prevalence of different kinds of motion signals is similar across movies, as measured by the PMO score in the XT-plane. Values are normalized by the PMO scores obtained from movie of random pixels. The region of interest (ROI) consisted of a 4x4 block of checks in the XT plane; each check corresponded to a single pixel in the discretization of the movie. Movies were (1) "The 39 Steps" (1935), (2) "Anna Karenina" (1935), (3) "A Night at the Opera" (1935) and (4) "Mr. & Mrs. Smith" (2005). For other details see Figure S1.
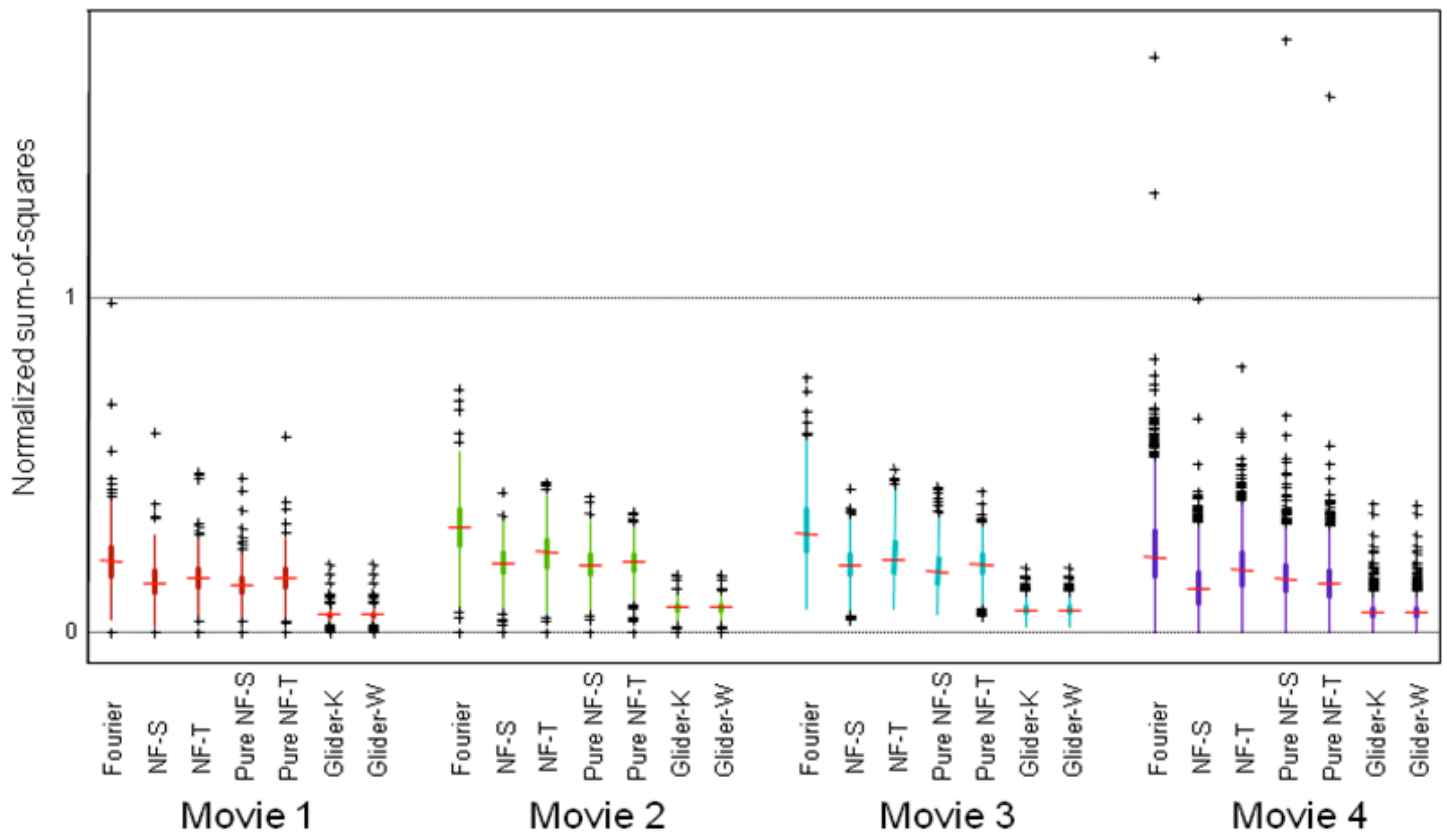
Figure S6: Prevalence of different kinds of motion signals is similar across movies, as measured by the RMO score in the XT-plane. For other details, see Figure S5.

Alternative binarization: PMO/RMO score

Here we show that the results in the main text hold when we use a binarization threshold equal to the midpoint of the global luminance range, rather than the median for each shot. Figures Figure S7 and Figure S8 shows that the results for the PMO and RMO scores respectively. The pattern of motion strengths is very similar to the results in the main text (Figures 8 and 9), obtained with binarization at the median of each shot..
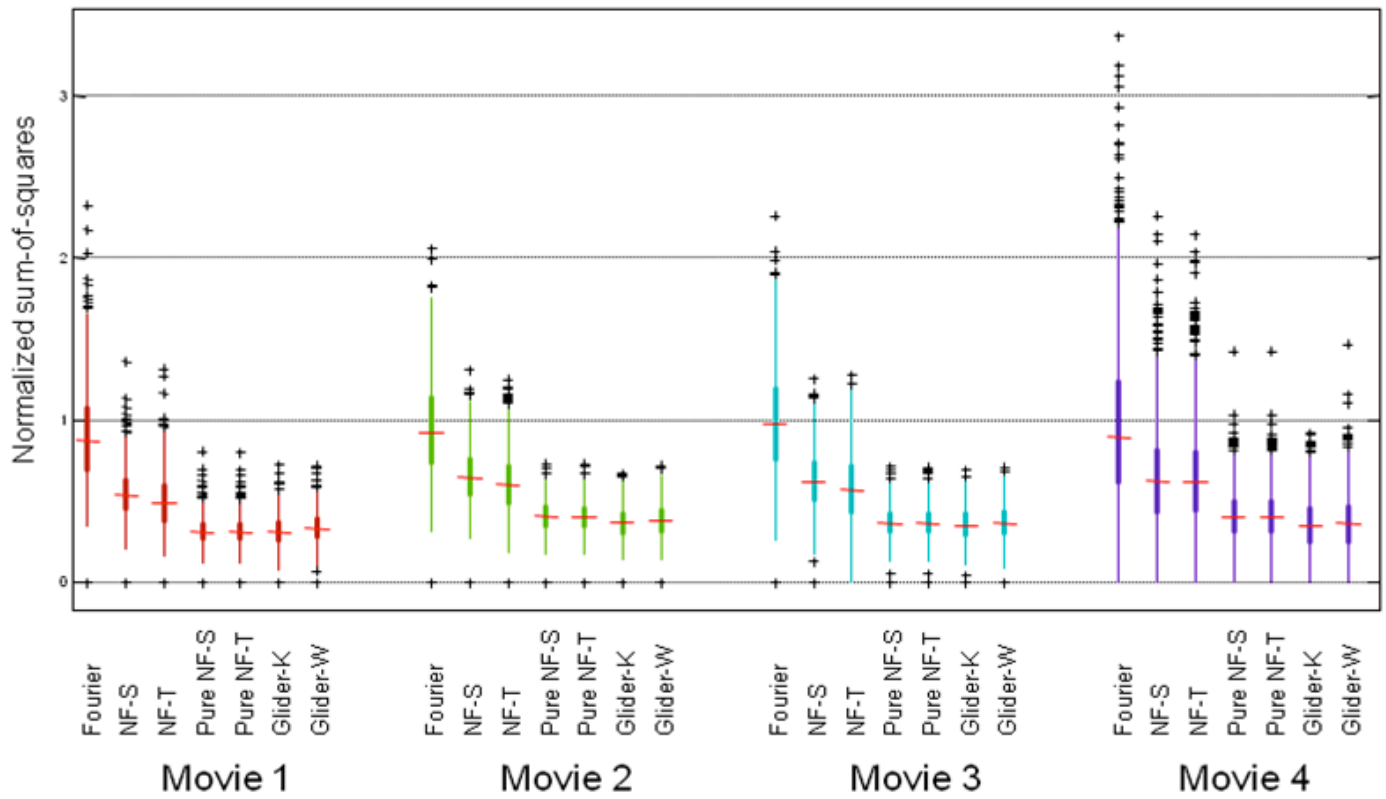
Figure S7: Prevalence of different kinds of motion signals is similar across movies, as measured by the PMO score. Pixels luminance values were thresholded at the midpoint of the luminance range (i.e 127 out of range of 0-255]). For other details, see Figure 8.
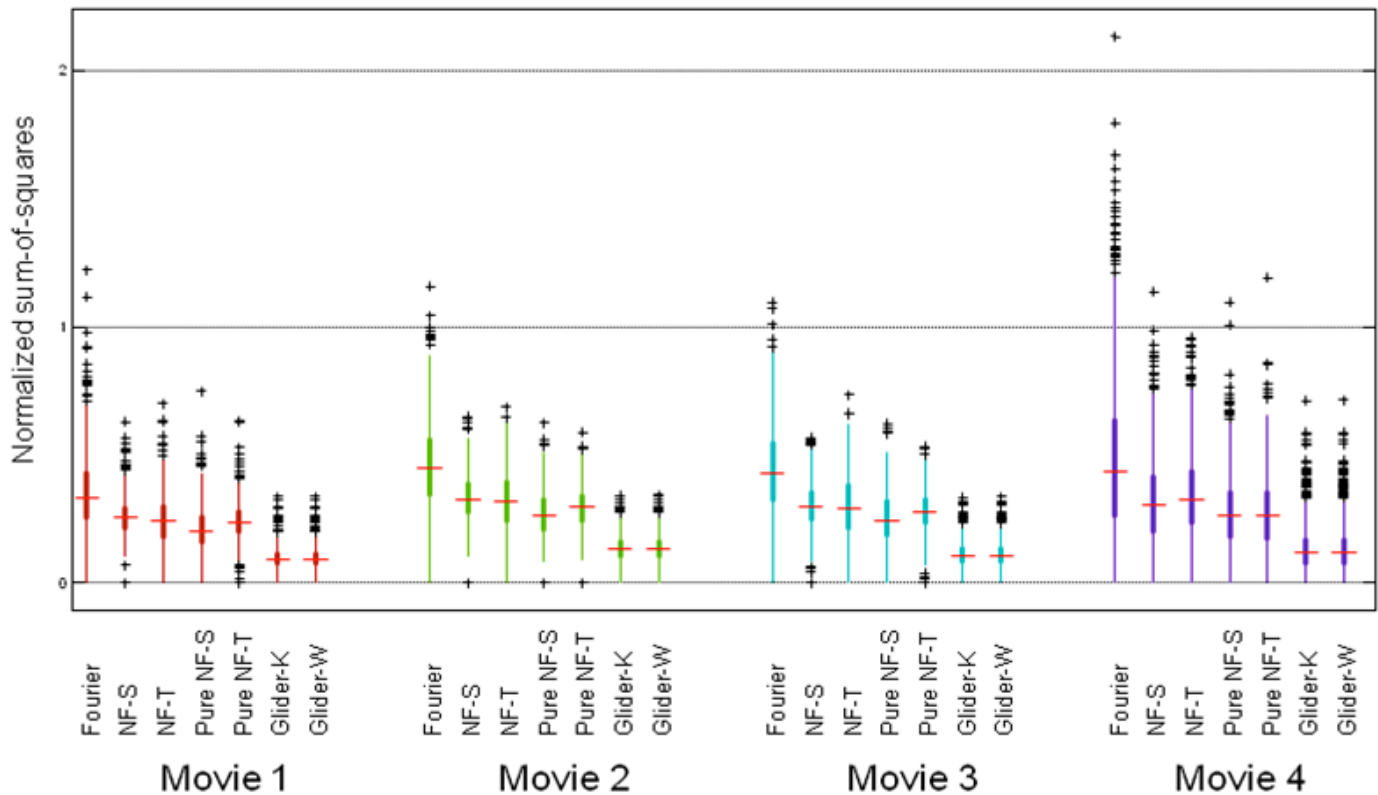
Figure S8: Prevalence of different kinds of motion signals is similar across movies, as measured by the RMO score. For other details, see Figure S7.

Covariance of motion signals: auxiliary analyses

Figure 10 of the main text showed how motion scores covary from scene to scene for movie 2 ("A Night at the Opera"); Figure S9 shows the parallel analysis of movie 4 ("Mr. & Mrs. Smith"). The main features of the plots are quite similar. For example, large values of the F motion RMO score tend to occur with large values of the G motion RMO scores, but the ratios between them (the bounding slopes of the scattergram) vary by approximately a factor of two.

In Figure S10 we show that the pattern of covariance of motion scores arises from the structure of the movies themselves, not from the way that the scores are calculated. That is, we carry out the same analysis on random movies in which each pixel is independently assigned to black or white. As is evident, there are largely uncorrelated among most pairs of motion scores. (There are, however, correlations between F and pure-NF scores, and, necessarily, between the pairs of scores that are mathematically identical, as described below.) Notably, the scatter of the motion scores are much smaller in random movies. This is because for a random movie, non-overlapping ROI's are independent – so that when local scores from a large number of ROI's are combined (approximately 256 x 256 ROI's per frame, and 500 frames per scene), the resulting sum has a very narrow fractional spread. Scatter is much larger for a naturalistic movie, since there are strong spatial and temporal correlations that correlate the ROI's within scenes.

Figure S11 shows that the covariation of the subtypes of the PMO and RMO scores is similar when we use a binarization threshold equal to the midpoint of the global luminance range, rather than the median luminance of each shot (as in Figure 10 and Figure S9).
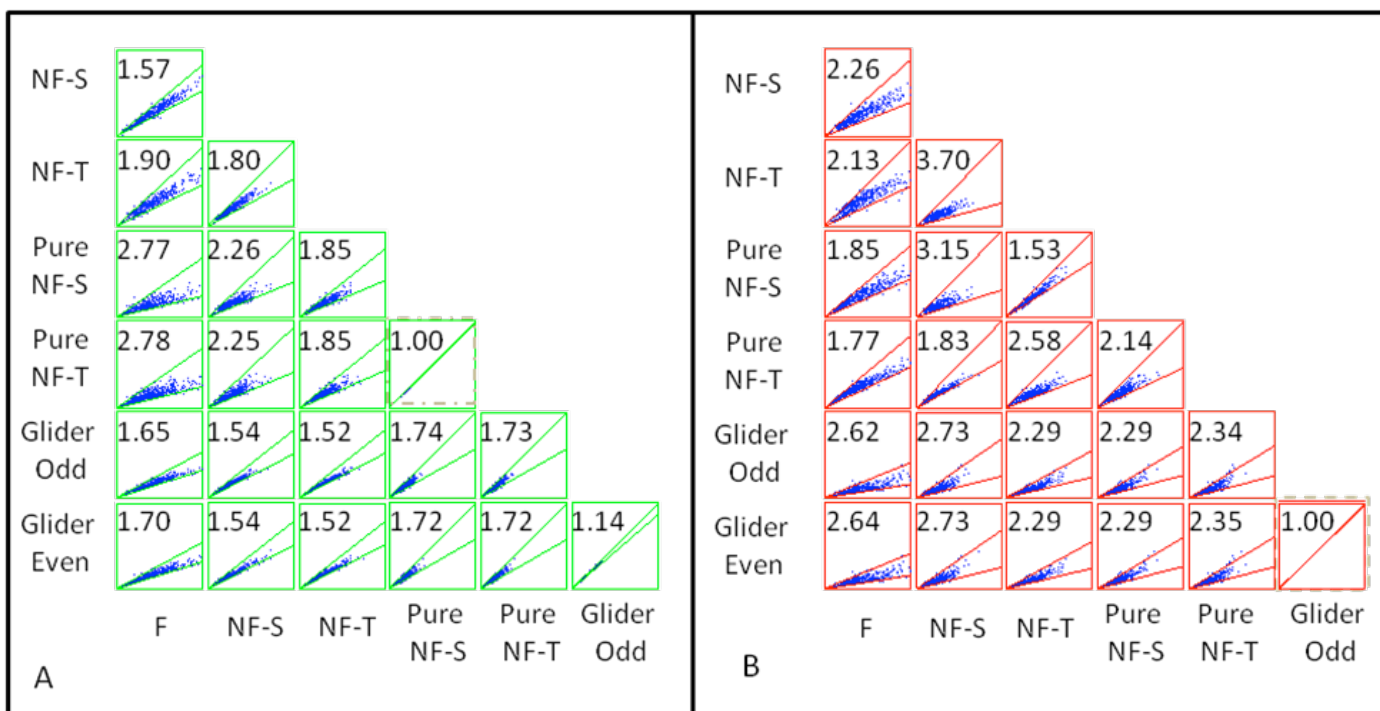
Figure S9: Covariance patterns of motion scores (panel A, PMO; panel B, RMO).   Within each scattergram, each point represents a pair of normalized motion scores determined from a single scene.   Axes range from 0 to 2 (PMO) and 0 to 1 (RMO). The number in each plot indicates the average ratio between the pair of motion scores; the lines indicate the wedge that contains 95% of the values. Large values of one motion score typically occur with large values of the other scores, but the ratios between a pair of scores can vary by up to a factor of 2. (Pure NF-S and NF-T PMO scores are identical, and the two G motion RMO scores are identical; see below.) Analysis was carried out in the YT plane, for movie 4 ("Mr. & Mrs. Smith").

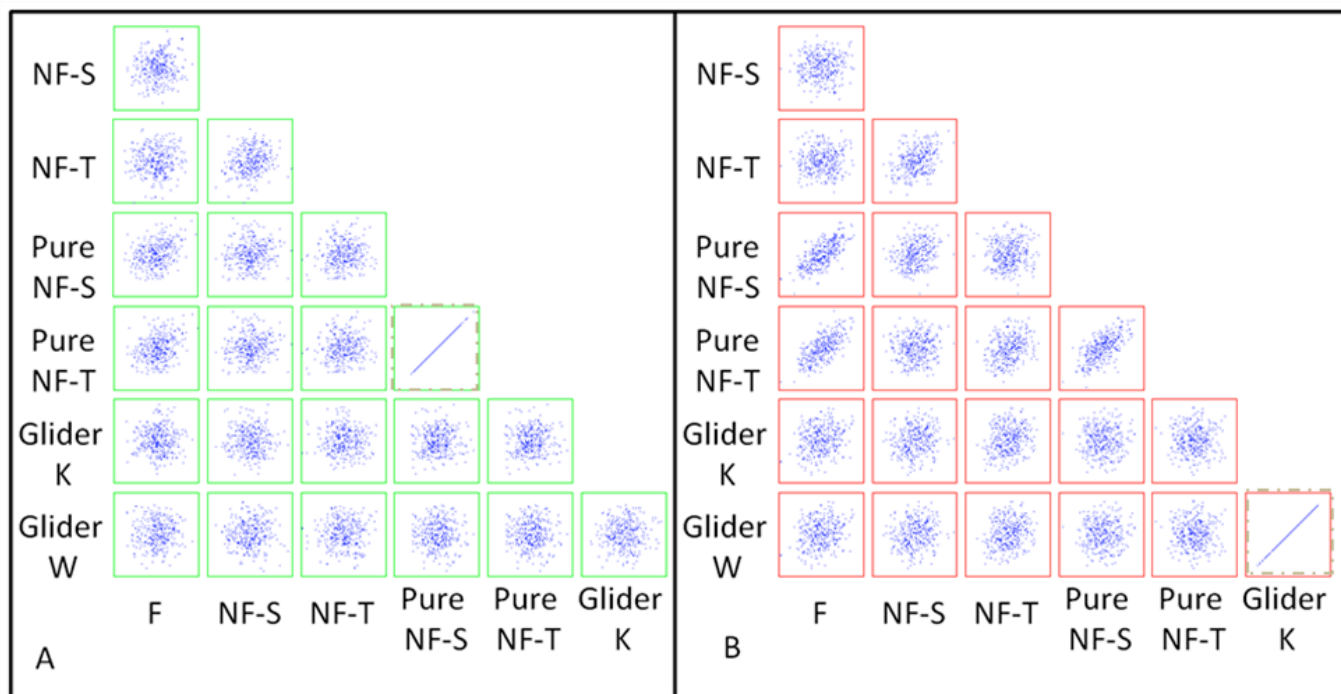Figure S10: Covariance patterns of motion scores (panel A, PMO; panel B, RMO) from 500-frame scenes drawn from a random movie. Within each scattergram, each point represents a pair of normalized motion scores determined from a single scene.   Axes range from 0.995 to 1.005 for RMO and PMO; the small scatter is due to the very large number of uncorrelated ROI's that make up each scene. Analysis was carried out in the YT plane.
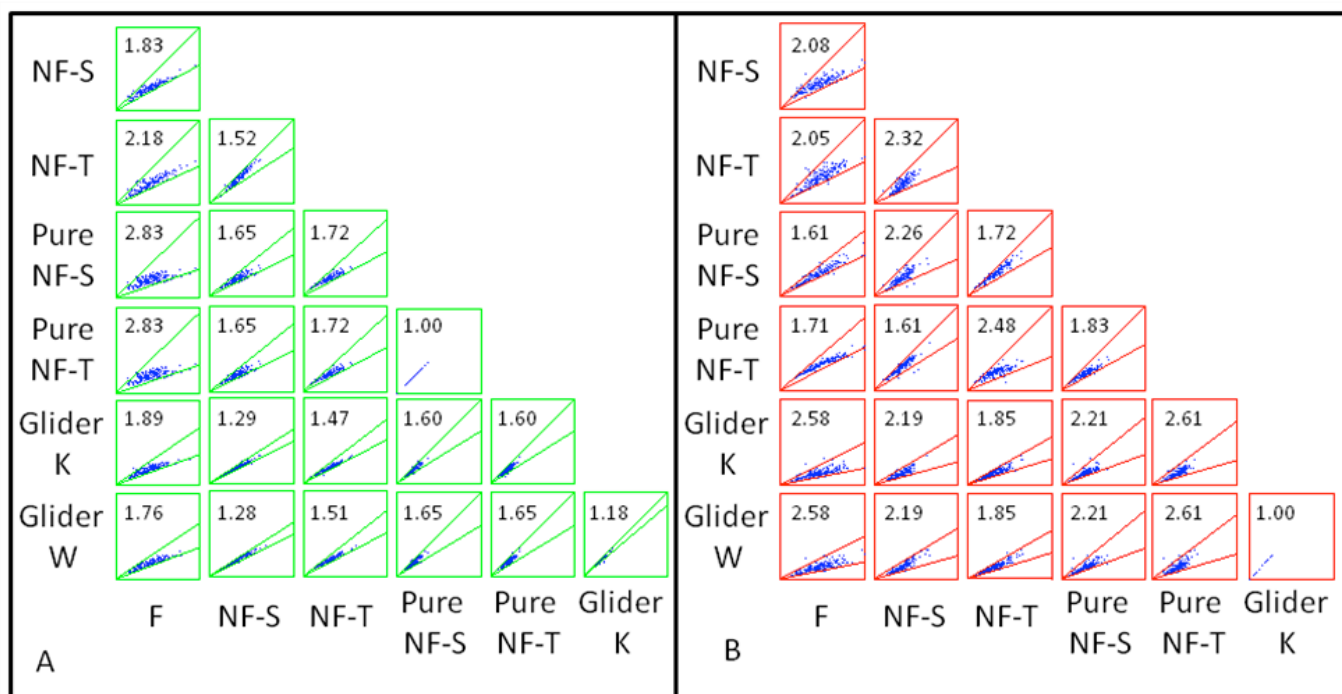
Figure S11: Covariance patterns of motion scores (panel A, PMO; panel B, RMO). Pixels luminance values were thresholded at the midpoint of the luminance range (i.e 127 out of range of 0-255]). Analysis was carried out in the YT plane, for movie 2 ("A Night At The Opera"). For other details, see Figure S9.

Section 2: Properties of the motion scores

Relationship to the Reichardt model

As mentioned in the main text, the motion scores developed here are intended to compare several kinds of local motion signals on an equal footing. We chose a computational structure for the scores that would enable a comparison between classical (F) motion signals and nonstandard ones (NF and G motion), for which there is no standard computational model. Here we make contact between this approach and the standard model-based approach for F motion, i.e., the Reichardt model. Specifically, we show that the RMO score for F motion coincides with the output of a Reichardt model for binary spatiotemporal movies. For non-binary movies and for the PMO score, the F motion scores correlate strongly with the output of the Reichardt model, but the correspondence is not exact.

To see why this correspondence is exact for binary movies and the RMO score, we note that for binary images, pairwise correlation, which amounts to multiplication, is equivalent to calculating parity. That is, taking black=1 and white=-1, the correlation across two checks is 1 if the parity of the number of black checks is even (i.e. two black checks or two white checks), and -1 if the parity is odd (i.e., one black check and one white check.) As a consequence, the initial stage of the Reichardt model [3], which calculates correlation by multiplication, corresponds precisely to determining whether the coloring of a two-check template follows the parity rule. Similarly, the second stage of the Reichardt model, at which opposing spatiotemporal correlations are subtracted, corresponds to comparing the number of times that the parity rule is followed, for templates in the two opposing directions. The latter is precisely what is computed by the opponent stage of the RMO procedure. Thus, the RMO score and the Reichardt detector output is perfectly correlated (Figure S12A).

The exact correspondence breaks down for the PMO procedure (Figure S12), since the PMO score in each direction is not simply the sum of the number of local rule violations, but rather, the number of checks whose state must be changed so that the entire ROI contains no parity-rule violations. While these quantities are distinct, they are nevertheless strongly correlated (Figure S12B). The precise correspondence also breaks down for grayscale images (Figure S12), since multiplication of grayscale values is not equivalent to a parity count on their binarized levels.
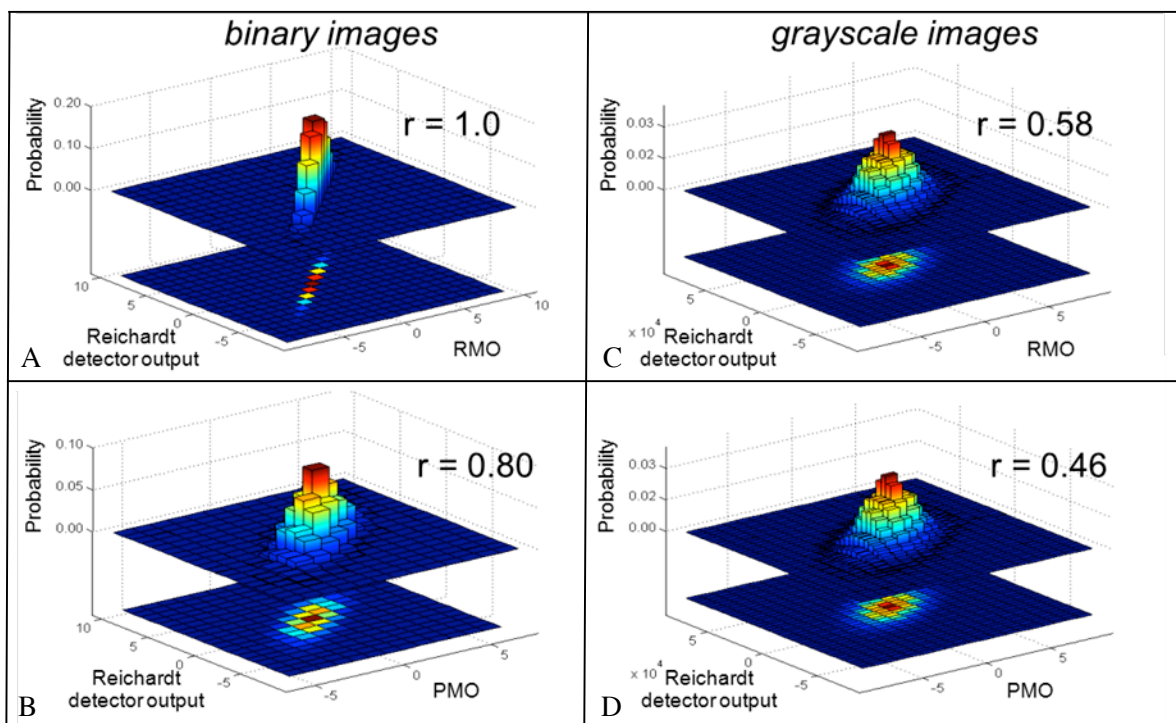
Figure S12: Correlations between RMO Fourier score, PMO Fourier score, and output of a Reichardt detector. Left, an exhaustive analysis of all binary images in a 4 x 4 ROI. The RMO score (A) correlates perfectly with the output of the output of a Reichardt detector; the PMO score (B) is strongly correlated to the Reichardt detector output, but not identical to it. Right, a parallel analysis of grayscale images. Gray levels are drawn independently from a uniform distribution. For each random example of a 4 X 4 grayscale image, the summed outputs of Reichardt detector are compared with the RMO (C) and PMO (D) Fourier scores of the corresponding binarized images. The scores are highly correlated with the Reichardt detector output; the correlation is stronger for the RMO score.

Exact relationships among some kinds of motion scores

Here we demonstrate some mathematical details concerning the relationships of the scores: (a) that for the two varieties of G motion, the RMO scores must be identical (but the PMO scores need not be), and (b) that for the two kinds of pure NF motion, the PMO scores must be identical (but the RMO scores need not be). These relationships are suggested by the data of Figures 10, Figure S9, and Figure S11; here we show that they are guaranteed mathematically.

(a) First, we consider the relationship of the G-W and G-K motion RMO scores. In both cases, the initial step of the calculation depends on counting the number of placements of a three-check template for which the colorings either follow (G-K) or violate (G-W) the parity rule. Denote the former by N_K and the latter by N_W. Since every coloring either follows or violates the parity rule, it follows that $N_K+N_W=P$, where P is the number of placements of a triangular glider template within the ROI. (For a 1 x 4 x 4 ROI, P=9). It therefore follows that $N_W=P-N_K$. Consequently, at the spatial opponent stage (Figure 4C of the main text), the value for the leftward-rightward difference signals will be opposite:

$$N_W^{L,forward} - N_W^{R,forward} = -(N_K^{L,forward} - N_K^{R,forward}).$$

Beyond this stage, there is one more subtraction (the forward vs. backward contrast), and the resulting values are then squared and summed to obtain a final RMO score. The forward vs. backward contrast maintains the equal-and-opposite relationship between the G-W and G-K scores, so that when values are squared and summed, the resulting scores are identical.

Note that this relationship does not hold for the PMO score; the latter is sensitive to *which* placements violate the rule, as switching the state of one check can affect the parity of more than one glider placement. Thus, the Hamming distances from the ROI to a perfect-match coloring need not be identical for G-W and G-K.

(b) Next, we consider the relationship between the PMO scores for spatial and temporal pure non-Fourier motion. The key step is to characterize the perfect-match ROI's, i.e., the ROI's in which every template placement follows the non-Fourier (NF) rule but does not follow the Fourier (F) rule. As we will show, the perfect-match ROI's are identical for the spatial and temporal cases. Since the PMO scores are determined by the distance from the ROI to the closest perfect match, it follows that PMO scores must be identical too.

To characterize the perfect-match ROI's for the spatial case, we start with Figure 2A, which enumerates all of the template colorings that follow the spatial non-Fourier rule (NF-S). Of these, the four on the left follow the Fourier rule; the pure NF-S library includes only the template colorings on the right. These have a simple description: in each case, the color assigned to a check at (y,t) is the opposite of the color assigned to the check at (y+1,t+1). Conversely, all such colorings are present in these four templates. This description is symmetric in space and time, and therefore, also applies to the pure temporal non-Fourier rule (pure NF-T). Thus, perfect-match ROI's for the spatial rule are also perfect-match ROI's for the temporal rule. Since PMO scores depend only on Hamming distances from these perfect matches, the pure NF-S and pure NF-T PMO scores must be identical.

Section 3: Analysis of synthetic non-Fourier motion stimuli

Perhaps the best-known examples of non-Fourier motion stimuli are the "drift-balanced" examples of Chubb and Sperling (1988). These are synthetic stimuli in which the moving object is distinguished from the background by the presence of random spatial or temporal contrast changes. The simplest example consists of a randomly flickering check moving across a gray background (Chubb and Sperling 1988, Fig. 4A). Fourier signals are eliminated by ensuring that the mean of the randomly flickering check is equal to that of the background. Here, we detail how our analysis captures the motion in this stimulus, and also highlight a technical point that is crucial in analyzing synthetic spatiotemporal sequences. We then show how the same reasoning applies to the other examples of Chubb and Sperling (1988), Figs. 4B-E.

Figure S13 shows how our analysis with the NF-S (panel A) and NF-T (panel B) templates applies to this stimulus. In the left column, the templates are placed in a way that is aligned with the stimulus grid. Note that each placement of the template always includes a check that is entirely contained within a gray region of the background (contrast=0). Hence, the products of the contrast values within the four checks is always 0, and no local motion signal is generated. In the second column, the template is displaced by a fraction of the discretization unit of the stimulus: in A (the spatial case), the displacement is a fraction of the check width; similarly, in B (the temporal case), the displacement is a fraction of the frame duration. The contrasts within these displaced templates are shown in the right column. In the spatial case (A), all of the checks of the NF-S template see a nonzero local contrast, and the products of these contrasts is always positive, producing a net NF-S local motion signals. Similarly, in the temporal case (B), an NF-T local motion signal is produced. Thus, generically positioned templates detect the NF motion, while templates that are aligned to the stimulus grid do not.

The reason for this is as follows. Formally, this drift-balanced stimulus – like the other NF stimuli of Chubb and Sperling (1988)– carry a motion signal because they contain many coherent correlations among four points in a slanted region of space-time. Consider movement in the $+Y$ direction, rendered in a stimulus of grid size Δy and frame duration Δt. For any two nearby points (y_1, t_1) and (y_2, t_2) that are inside the moving stimulus check within the same frame, then $(y_1 + \Delta y, t_1 + \Delta t)$ and $(y_2 + \Delta y, t_2 + \Delta t)$ will always be inside the same moving stimulus check on the next frame. This in turn means that for any two such points (y_1, t_1) and (y_2, t_2), the product $I(y_1, t_1) I(y_2, t_2) I(y_1 + \Delta y, t_1 + \Delta t) I(y_2 + \Delta y, t_2 + \Delta t)$ will always be positive, whether or not the check flickers as it moves. In contrast, the opposing-direction product $I(y_1, t_1) I(y_2, t_2) I(y_1 - \Delta y, t_1 + \Delta t) I(y_2 - \Delta y, t_2 + \Delta t)$ will always be zero. Each of these imbalances of fourth-order correlations represents a motion signal. Since the NF-S template computes such a product, the NF-S template will detect this signal – provided that two of its template checks are inside the same stimulus check (as shown in the middle and right columns of Figure S13A). This will generically be the case *unless the template is always positioned to have its spatial boundary coincident with an edge of the stimulus grid* (as shown in the left column of Figure S13A). Put another way, if the template is placed randomly on the movie (without regard to stimulus-check boundaries), then typical placements will yield a proper motion signal. Figure S13B shows that the same kind of phenomenon occurs in the time domain: the NF-T template will also detect the correlations that carry the motion, provided that its temporal boundaries occasionally occur during a frame. (Another way of looking at this is that the motion of the original Chubb and Sperling stimulus is defined by correlations of unsigned contrasts across space and time. This is captured by the fourth-order product $I(y_1, t_1)^2 I(y_1 + \Delta y, t_1 + \Delta t)^2$, which is in turn equivalent to the generic product $I(y_1, t_1) I(y_2, t_2) I(y_1 + \Delta y, t_1 + \Delta t) I(y_2 + \Delta y, t_2 + \Delta t)$ provided that (y_1, t_1) and (y_2, t_2) are within the same stimulus check.)

Note that this issue – the need to consider generic placement of the template with respect to the stimulus grid – does not arise in the analysis of digitized naturalistic movies, since the check and frame boundaries used for digitization bear no fixed relationship to the features in the original image. Thus, even if the grid used for the templates is the same as the grid of the movie digitization, the relationship of the templates to the objects in the original image will still be generic.
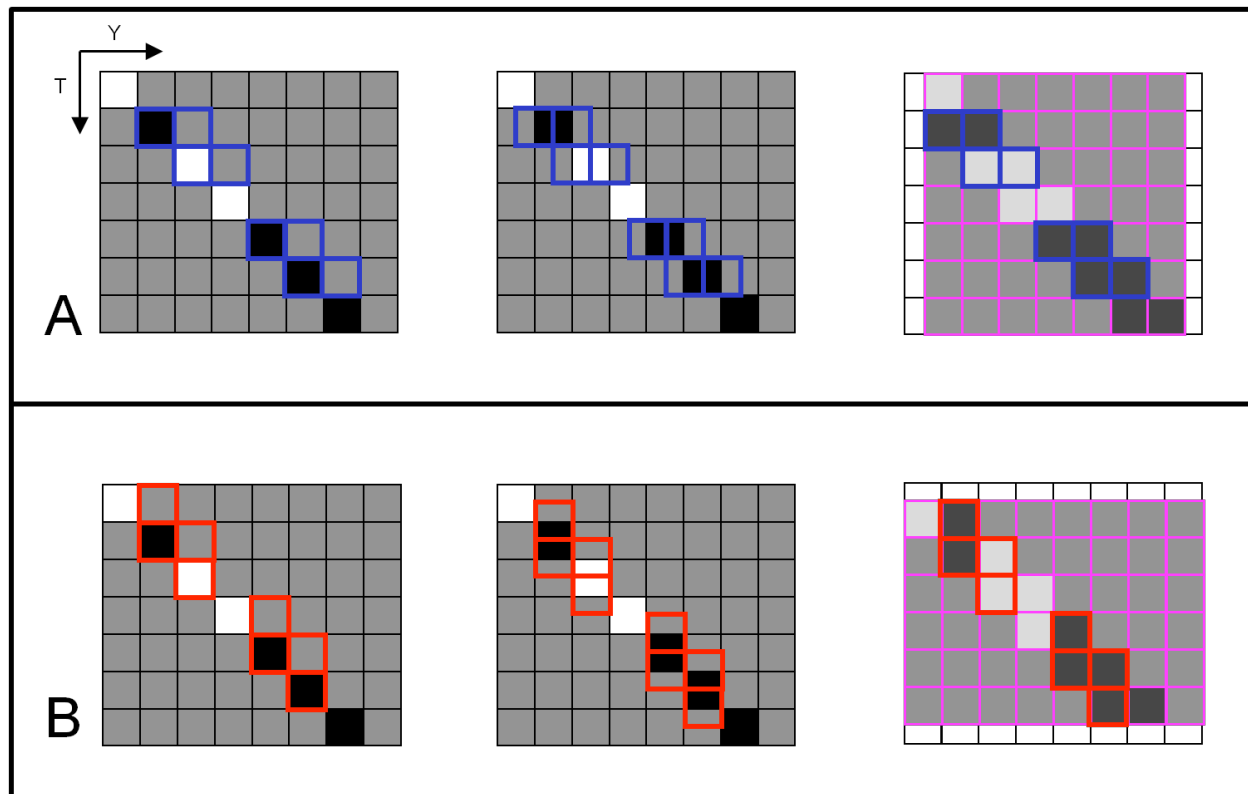
Figure S13: For analysis of artificial stimuli generated on a grid, consideration of generic placements of the template is crucial. This is shown for analysis via the NF-S template (panel A) and the NF-T template (panel B), as applied to the classic NF stimuli presented by Chubb and Sperling (1988). When the analysis template is aligned to the stimulus grid (left columns), at least one of its checks is entirely within the background. Thus, the product of the contrast values in the four template checks is zero, and no motion signal is detected. However, any misalignment (middle columns) allows all four template checks to be influenced by the stimulus, so the product of the four contrast values captures the motion (right columns).

With regard to the other examples of Chubb and Sperling (1988), the NF-T template also captures the motion signal. The basic reason is that the product of the luminances within the four checks of an NF-T template, $P = I(y,t)I(y,t + \Delta t)I(y + \Delta y, t + \Delta t)I(y + \Delta y, t + 2\Delta t)$, can be rewritten $P = U(y,t)U(y + \Delta y, t + \Delta t)$, where each term $U(y,t) = I(y,t)I(y,t + \Delta t)$ is a product of luminances at the same location, across two timesteps. The quantity U has a magnitude that is equal to the product of the unsigned contrasts at these two timesteps, and a sign that indicates whether there is flicker. Thus, four-term product P computed by the template indicates whether contrast and flicker at the location y and the time interval $[t, t + \Delta t]$ is correlated with contrast and flicker at an adjacent location $y + \Delta y$ and the later time interval $[t + \Delta t, t + 2\Delta t]$. Provided that the template is oriented parallel to the motion direction in each of the examples of Chubb and Sperling, this correlation is maximized. When the template is reversed, the correlation is incomplete. In their example C and D, this is because one of the U-terms can be positive while the other is negative; in their examples B and E, this is because one of the U-terms can be nonzero while the other is zero. In all cases, the difference between maximal correlation for one orientation of the template and incomplete correlation in the other orientation generates a motion signal via the opponent process – either at the level of the local motion signal, or, when the local motion signals are pooled within ROI's via the RMO and PMO calculations (Figure 2). Similar reasoning applies to the examples of Fleet and Langley (1994)'s Figure 11, but here, neither orientation of the template leads to a maximal correlation; instead, one orientation generates a larger correlation than the other.