Spatiotemporal Information Flows

Anonymous Author(s)
Affiliation
Address
email

Abstract

We develop a novel flow extraction framework, *spatiotemporal information flow* to capture, in an unbiased manner, salient causal relationships between pixels over space and time. Real spatiotemporal dynamical systems such as cellular morphodynamics are complex, nonlinear and evolve over time in response to feedbacks. This makes it highly challenging to model, simulate or fit phenomena from first principle physics. More critically, we often do not know *a priori* and desire to discover the salient variables to include and the key relationships to model. As such causal measures to identify relationships direct from observational multivariate timeseries have been developed. These measures however have largely only been studied in 1D. Here, spatiotemporal information flows present a general, multiscale method to extend 1D causal measures to 2D + time video. Applying spatiotemporal information flows we discover the salient pixel-to-pixel information transfer highways in videos of diverse phenomena from traffic and crowd flow, to collision physics, to fish swarming, to moving camouflaged animals, to human action, embryo development, cell division and cell migration.

1 Introduction

In this work, we consider the problem of extracting salient causal dynamic patterns from video. Capturing semantically meaningful salient video dynamics over time is challenging. In general video dynamics represent real spatiotemporal dynamical systems and are complex, nonlinear and ever evolving and cannot in general be described from first principles physical differential equations. Over short-time intervals, $[t, t+\delta t]$, frame-to-frame changes, optical flow is the most popular approach and many methods of estimation have been proposed including block based matching Liu and Delbruck [2018], gradient based methods (Farnebäck [2003]) and learning based methods Dosovitskiy et al. [2015]. For longer-time intervals with deep learning, the short time dynamics captured by optical flow have been successfully exploited as a supplementary signal in applications such as action recognition, segmentation and video relationships Sun et al. [2018]. However deep learning training is expensive and subject to the availability of appropriate training data. In addition, for novel scientific discovery in fields such as biology, labelled data can be expensive and unavailable or the concepts too abstract and non-amenable for annotatation, optical flow and statistical modelling remains an important tool for applications of cellular tracking and cell-cell interaction networks Zhou et al. [2019].

Classically, optical flow seeks to compute the physical displacement $(\Delta x, \Delta y)$ between two video frames taken at times t and $t + dt$ at every pixel position assuming conservation of intensity, $I(x, y) \approx I(x + \Delta x, y + \Delta y)$. However, physical motion describes only one potential aspect of 'causal information flow' within a sequence of video images. In particular, motion captures only the egocentric motion of single actors and misses potential communal interactions between actors within a video such as, the directional changes of pedestrians and cars on highways, the physical collision of snooker balls, the influence of neighboring cells in confluent tissue, and the cooperation of individual body parts to execute a stereotypic action - all of which can be established over many frames.

Submitted to *A causal view on dynamical systems* workshop at NeurIPS 2022. Do not distribute.

We consider whether complex salient video dynamics could be more comprehensively captured by treating individual image pixels and their local neighborhood as a multivariate 1D time series in the context of dynamical systems and measuring the pixel-to-pixel causal information flow. Causal measures for quantification of information transfer between 1-dimensional (1D) timeseries have been extensively studied in scientific disciplines such as econometrics Granger [1969], ecology Sugihara et al. [2012], Detto et al. [2012] and neuroscienceSeth et al. [2015]. The resulting methods can be generally classified into two categories: functional and effective connectivity. Functional connectivity is defined by observable statistical dependencies among timeseries and includes families of methods such as correlation, coherence, Granger causality, or transfer entropy Bastos and Schoffelen [2016]. On the other hand, effective connectivity is defined by the influence on activity, and depends on an explicit parametric model of the interaction. Examples include structural causal models and dynamic differential covariance Valdes-Sosa et al. [2011] Chen et al. [2022].

Here, we develop a general, multiscale framework termed 'spatiotemporal information flows' inspired by classic optical flow approaches to operationalise these 1D functional and effective connectivity measures for real 2D videos. We then apply spatiotemporal information flows to real world datasets to extract salient pixel information highways. First, we uncover dynamic patterns in a crowd flow segmentation dataset (Ali and Shah [2007]) to qualitatively and quantitatively compare and highlight differences to mean optical flow. Second, we demonstrate the generality of information flows with diverse examples of video dynamical systems from action recognition (Soomro et al. [2012]), collision physics (Yi et al. [2018]), developmental biology (Tomer et al. [2012]), and cell migration (Ulman et al. [2017]). Finally, we compared the ability of optical and information flows on the challenging task of segmenting moving camouflaged animals in the wild (Lamdouar et al. [2020]).

2 Related Works

Causal information flow as an alternative to optical flow. Yamashita et al. [2012] introduced an initial notion of causal flow for video and compared their results to Horn-Schunk's version of optical flow. However, their work strictly implements one particular variant of Granger causality, namely PCCA, operated on a single spatial scale and only demonstrated their method on crowd flows.

Causal inference of object relationships in videos. Literature exists related to inferring the causal relationships in videos, typically amongst precomputed keypoints or other object-level representations instead of individual pixels, often using a framework grounded in Granger causality Narayan and Ramakrishnan [2014], Prabhakar et al. [2010], Swears et al. [2014]. As such these methods tend to identify causal links between defined actors, instead of general salient 'information highways' uncovered in our analysis and require *a priori* domain knowledge.

Functional connectivity networks Notions of functional connectivity have been used in biology, including in the fields of neuroscience and cell biology. As in the machine learning literature, most works in this area focus on establishing functional connectivity networks between discrete, predefined entities such as physical partitions of the brain Bastos and Schoffelen [2016], Noble et al. [2019] or between segmented cells Zamir et al. [2022].

3 Methods

3.1 Spatiotemporal information flow

Given a grayscale input video, $X \in \mathbb{R}^{T \times H \times W}$ with T total frames, and $H \times W$ image frame pixel size, let $X_{i,j,t}$ be the intensity timeseries at the pixel position $(i, j) \in [0, H-1] \times [0, W-1]$ over the time window $[t, t + \Delta t]$. Consider all pixel timeseries within a $P \times P$ pixel neighborhood centered around (i, j), $\{X_{(i+\delta_i, j+\delta_j, t)}\}$ where $\delta_i = -P/2, ..., P/2, \delta_j = -P/2, ..., P/2$ denote the integer pixel offsets, and denote the information transfer as quantified by causal effect between the pixel timeseries $X_{i,j}$ to a neighbor pixel timeseries, $X_{i+1,j}$ as $I_{X_{i,j} \to X_{i+1,j}}$ then we define the spatiotemporal information flow, ΔI following Yamashita et al. [2012] as the mean vector summarising the net causal effect spatial gradient over the central (i, j) pixel; the total sum of causal

effects multiplied by an orientation vector, $O_{i,j}$ with the following $x-$, $y-$ components

$$O_{(i,j),x} = (-1) \sum_{\delta_j=1}^{P/2} I_{X_{i,j} \to X_{i,j-\delta_j}} + \sum_{\delta_j=1}^{P/2} I_{X_{i,j} \to X_{i,j+\delta_j}} \tag{1}$$

$$O_{(i,j),y} = (-1) \sum_{\delta_i=1}^{P/2} I_{X_{i,j} \to X_{i-\delta_i,j}} + \sum_{\delta_i=1}^{P/2} I_{X_{i,j} \to X_{i+\delta_i,j}} \tag{2}$$

$$\nabla I_{i,j} = \sum_{\delta_i,\delta_j} X_{i+\delta_i,j+\delta_j} \sqrt{O^2_{(i,j),x} + O^2_{(i,j),y}} \tag{3}$$

To reduce computation, we extract all non-overlapping 3x3 neighborhoods (i.e. $P = 3$), compute $\nabla I_{i,j}$ and linearly interpolate the resulting vector field to compute the spatiotemporal information flow ∇I_t over the time interval with the same image size as the video frame, $\nabla I \in \mathbb{R}^{H \times W \times 2}$. Color RGB videos were first converted to grayscale videos.

3.2 Implemented causal measures as information flows

We extend and implement four literature causal information measures as spatiotemporal information flows as described below:

Maximum Cross-Correlation (max. CC). Cross-correlation measures the similarity between two timeseries over different temporal lags. The peak position offset of the maximum spatial cross-correlation between a pixel neighborhood at time t and the same neighborhood in the successive frame at time $t + \delta t$ is commonly used to extract dense optical flow in fluid mechanics and biology where it is called particle image velocimetry Adrian [2005]. We extend this concept to convert correlation into a pixel timeseries causal measure. Consider all pixel timeseries in a $P \times P$ neighborhood around (i, j), $\mathbf{X}_{i,j,t} = \{X_{(i+\delta_i,j+\delta_j,t)}\}$ over a time interval $[t, t+\Delta t]$, and construct a m-frame lagged signal $\mathbf{X}_{i,j,t-m} = \{X_{(i+\delta_i,j+\delta_j,t)}\}$, where $t \in [t_0, t_0 + \Delta t - m]$ as a 'previous' signal and a m-frame leading signal $\mathbf{X}_{i,j,t} = \{X_{(i+\delta_i,j+\delta_j,t)}\}$, where $t \in [t_0 + m, t_0 + \Delta t]$ as a 'current' signal where $m > 0$. Compute the spatiotemporal 3D correlation,

$$CC(t) = \sum_{\delta'_i,\delta'_j,\delta'_t} \mathbf{X}_{i+\delta i',j+\delta j',t-m+\delta t'} \mathbf{X}_{i,j,t}, \in \mathbb{R}^{P \times P \times (\Delta t - m)} \tag{4}$$

over spatial and temporal lags using the same padding. Finally take the maximum value over time, $CC = \max_t CC(t) \in \mathbb{R}^{P \times P}$ to compute the matrix of causal effect of the central pixel timeseries, $X_{i,j}$ on each neighbor pixel timeseries, $X_{i+\delta_i,j+\delta_j}$ such that $I_{X_{i,j} \to X_{i+1,j}} = CC_{\delta_i=+1,\delta_j=0}$.

Conditional Granger causality (cGC). Granger causality (GC) establishes that a variable, x can be said to be 'causal' to another variable y if the additional inclusion of the history of x reduces the uncertainty in the prediction of future values of y. Initially conceptualized by Wiener [1956] GC was operationalized using vector autoregressive (VAR) models by Granger [1969]. Later, Geweke [1984] extended the VAR method to include a third variable z in the regression models which x and y may depend upon. The modified method is commonly referred as conditional Granger causality (cGC). Due to the conceptual simplicity, GC variants are perhaps the most popular class of tests for causality based on temporal precedence for 1D timeseries (Bressler and Seth [2011]). To adapt the cGC setup to pixel timeseries, we jointly consider all pixel timeseries in a $P \times P$ neighborhood around (i, j), $\mathbf{X}_{i,j} = \{X_{(i+\delta_i,j+\delta_j,t)}\}$ over a time interval $[t, t + \Delta t]$ and test the effect of including the present time vs its exclusion. Let $\mathbf{X}_{i,j,t-m} = \{X_{(i+\delta_i,j+\delta_j,t-m)}\}$ denote the $m > 0$-frame lagged collection of signals, where $t \in [t_0 + m_{\max} - m, t_0 - m + \Delta t]$ and $\mathbf{X}_{i,j,t} = \{X_{(i+\delta_i,j+\delta_j,t)}\}$, the 'current' collection of signals which we wish to predict, and $t \in [t_0 + m_{max}, t_0 + \Delta t]$ as the 'current' signal, and $m_{\max}$ is the maximum specified lag, then construct the pair of restricted and

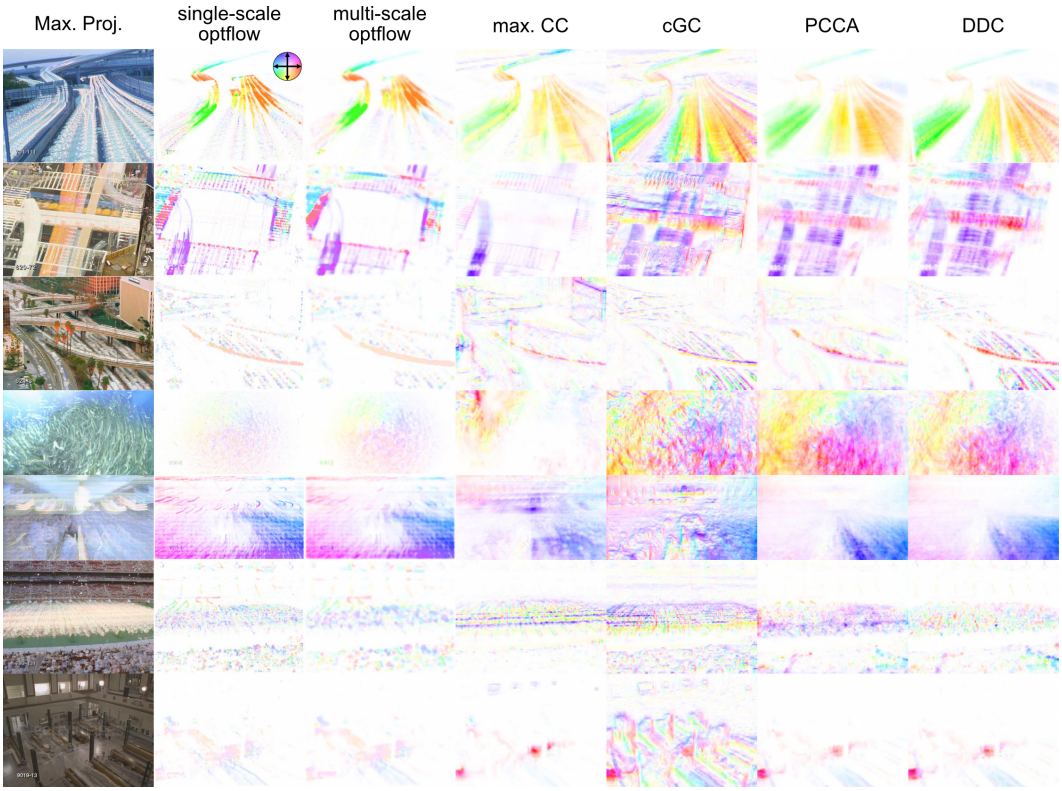

Figure 1: Comparison of the mean single- and multi-scale optical flow with four spatiotemporal information flows. **Row 1:** highway traffic. **Row 2:** pedestrian and motor traffic at a 4-way junction. **Row 3:** Multi-storey, multi-directional highway traffic. **Row 4:** Anti-clockwise rotating fish swarm. **Row 5:** two way crowd movement indoors. **Row 6:** cheerleading performance in a sports arena. **Row 7:** visitors in an art gallery. Direction of flow is colored per the color wheel

unrestricted regression models

$$\mathbf{X}_{i,j,t} = \mathbf{B}_{m=0}\mathbf{X}_{i,j,t} + \sum_{m=1,\ldots,m_{\max}} \mathbf{A}_m \mathbf{X}_{i,j,t-m} + \mathbf{E}^{full}_{i,j,t} \tag{5}$$

$$\mathbf{X}_{i,j,t} = \sum_{m=1,\ldots,m_{\max}} \mathbf{C}_m \mathbf{X}_{i,j,t-m} + \mathbf{E}^{reduced}_{i,j,t} \tag{6}$$

The causal effect, $cGC \in \mathbb{R}^{P\times P}$ is computed following Geweke [1984] as

$$cGC = \ln \frac{\text{var}(\mathbf{E}^{reduced}_{i,j,t})}{\text{var}(\mathbf{E}^{full}_{i,j,t})} \tag{7}$$

where the variance operation is taken over time. The information transfer, $I_{X_{i,j}\rightarrow X_{i+1,j}} = cGC_{\delta_i=+1,\delta_j=0}$.

Probabilistic Canonical Correlation Analysis (PCCA) Granger causality. Instead of considering a single scalar intensity timeseries per pixel, Yamashita et al. [2012] extended conditional Granger causality to the case when a pixel can be described by a multivariate timeseries. Their method is based on probabilistic canonical correlation analysis (PCCA) which corrects for the additional intercorrelations between consituents of a multivariate series (Fujita et al. [2009]). This allowed them to consider not just a single pixel intensity timeseries for pixel position (i, j) but a $P \times P$ pixel block of intensity timeseries centered around (i, j). Analogously, instead of neighbor pixels, neighboring $P \times P$ pixel blocks are defined. Crucially, the information flow computation remains as desribed in Eqn.3. We refer to Yamashita et al. [2012] for the mathematical details. In

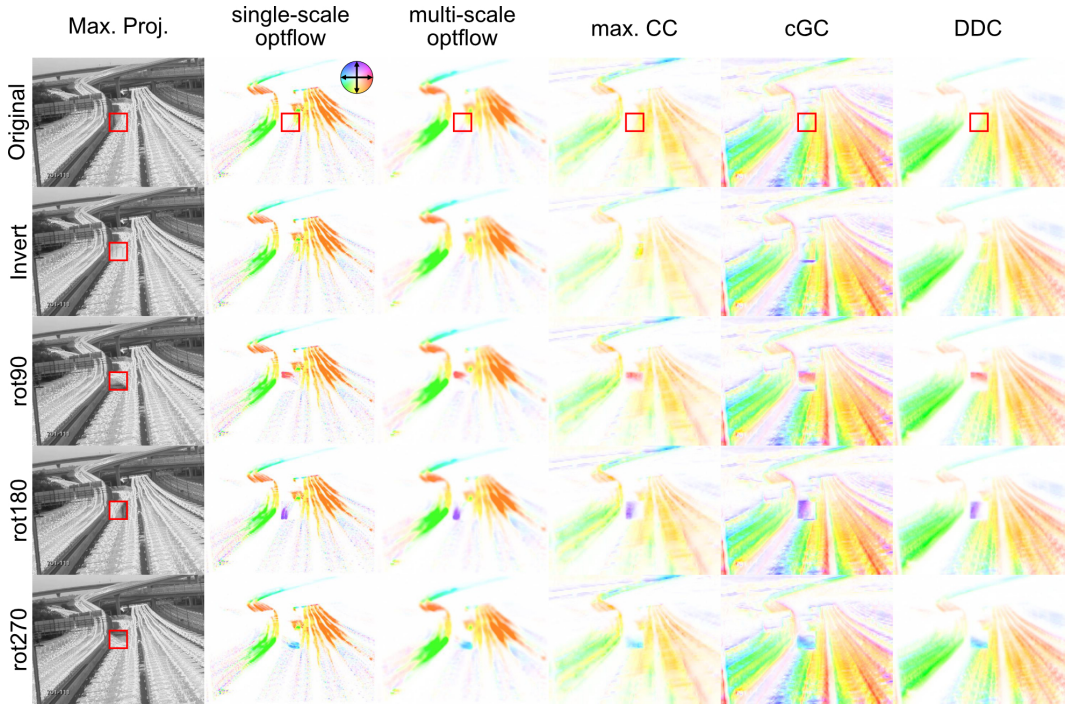

Figure 2: Perturbations of a traffic flow by inverting and rotating the pixel intensity within a fixed region of interest (red box), 50x50 pixels in size and the resulting computed single- and multi-scale optical flow and three different spatiotemporal flows. Direction of flow is colored per the color wheel

theory, the resulting multi-block Granger causality is more robust to noise that may affect single pixels.

Dynamic differential covariance (DDC). The VAR models that operationalise Granger causality testing based on fitting residuals are equivalent to using state-space dynamical systems to model pixel intensity dynamics (Barnett and Seth [2015]). Chen et al. [2022] instead considers an linear ordinary differential equation (ODEs) dynamical system to model pixel intensity dynamics. They show that the maximum least estimator of the assumed linear model coefficients are a measure of causal information flow which they term dynamic differential covariance (DDC). We adapt DDC to jointly consider all pixel timeseries in a $P \times P$ neighborhood around (i, j), $\mathbf{X}_{i,j,t} = \{X_{(i+\delta_i, j+\delta_j, t)}\}$ over a time interval $[t, t + \Delta t]$. Let $d\mathbf{X}_{i,j,t}/dt = \{dX_{(i+\delta_i, j+\delta_j, t)}/dt\}$ denote the time derivative of the signals, then

$$\frac{d\mathbf{X}_{i,j,t}}{dt} = \mathbf{W}\mathbf{X}_{i,j} \tag{8}$$

$$DDC = \mathbf{W} \tag{9}$$

where $\mathbf{W}$ is robustly estimated using ridge regression. $DDC \in \mathbb{R}^{P \times P}$ and the matrix entries are the information transfer such that $I_{X_{i,j} \to X_{i+1,j}} = DDC_{\delta_i=+1, \delta_j=0}$.

3.3 Multiscale implementation

We develop a multiscale information flow based on Gaussian image pyramids and average pooling. Given a grayscale input video sequence, $X \in \mathbb{R}^{T \times H \times W}$ with T total frames, and $H \times W$ image frame pixel size, and an array of N desired downsampling factors, $s = \{s_1, s_2, ..., s_N\}$ in increasing magnitude, compute the dowsampled video sequences with linear interpolation, $\mathbf{X}_s = \{\sigma_1(X_{s_1}) \in \mathbb{R}^{T \times H/s_1 \times W/s_1}, \sigma_2(X_{s_2}) \in \mathbb{R}^{T \times H/s_2 \times W/s_2}, ..., \sigma_N(X_{s_N}) \in \mathbb{R}^{T \times H/s_N \times W/s_N}\}$ where $\sigma(\cdot)$ denotes an optional isotropic Gaussian smoothing with the specified smoothing σ. For each video sequence compute the respective information flow fields, $\nabla I = \{\nabla I_{s_1} \in \mathbb{R}^{H/s_1 \times W/s_1 \times 2}, \nabla I_{s_2} \in \mathbb{R}^{H/s_2 \times W/s_2 \times 2}, ..., \nabla I_{s_N} \in \mathbb{R}^{H/s_N \times W/s_N \times 2}\}$. The multiscale informa-

Table 1: Comparison of optical and information flows to detect perturbed video regions on 4 different perturbations of the crowd flow segmentation dataset. Bold numbers indicate the best value in column.

	Invert		**rot90**		**rot180**		**rot270**	
Flow	mAP_{25}	IoU_{25}	mAP_{25}	IoU_{25}	mAP_{25}	IoU_{25}	mAP_{25}	IoU_{25}
Single-scale optflow	36.9	35.3	93.3	**70.4**	93.3	**68.6**	93.3	**70.4**
Multi-scale optflow	37.5	29.2	**100.0**	64.2	**96.7**	62.2	**100.0**	63.8
max. CC	**74.8**	**46.4**	84.7	51.3	90.6	51.8	84.7	52.1
cGC	16.3	12.8	30.1	22.6	22.6	19.8	27.7	21.1
DDC	47.8	22.6	93.3	55.3	93.3	54.5	93.3	55.0

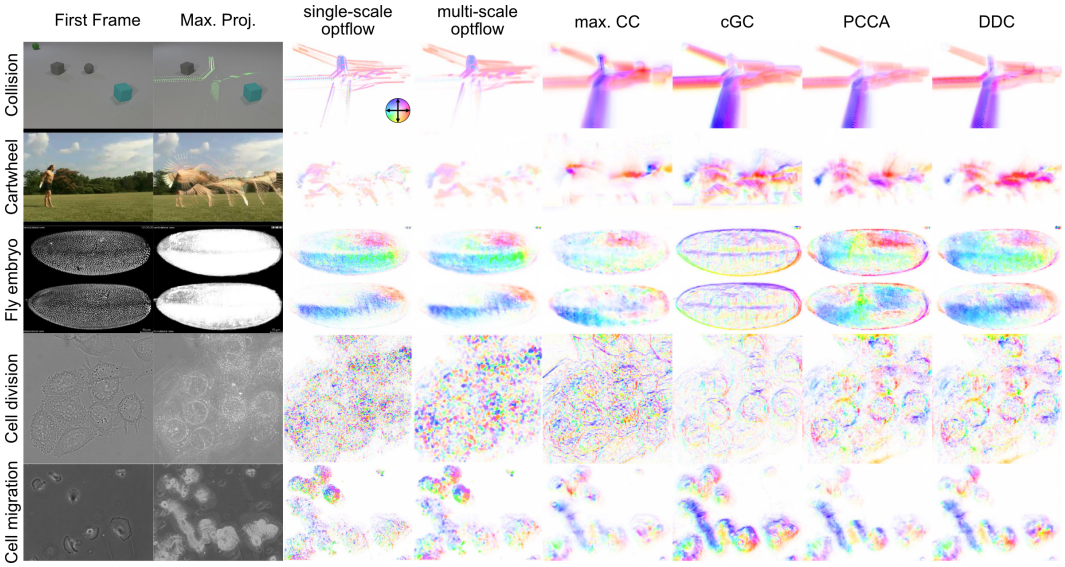

Figure 3: Comparison of the mean single- and multi-scale optical flow with four spatiotemporal information flows. **Row 1:** collision scence. **Row 2:** cartwheel scene. **Row 3:** *Drosophila* embryo. **Row 4:** cell divsion. **Row 5:** cell migration. Direction of flow is colored per the color wheel

tion flow, $\nabla I_{ms} = \text{mean}\{\nabla I\}$ is the mean flow field after linearly upsampling all flows to the image dimensions of the largest scale, s_1.

4 Experiments

We conduct qualitative and quantitative experiments on various video datasets to verify our implementation and demonstrate the utility of our proposed spatiotemporal information flows in comparison to classic optical flow. Optical flow was computed with the algorithm of Farnebäck [2003] at both single and multiple scales.

4.1 Crowd flow segmentation dataset

4.1.1 Crowd flow information flows

Crowd flow segmentation refers to the task of extracting salient motion patterns of high density moving objects where individual objects cannot be independently identified and segmented. Following Yamashita et al. [2012] we computed optical and information flows for the dataset of Ali and Shah [2007] comprising 30 videos of diverse crowd flow, including highway traffic, pedestrian crossing, fish swarms and cheerleading. For a given grayscale video, $X \in \mathbb{R}^{T \times H \times W}$ we consider the intensity pixel time series $X_{i,j,t}$ at the pixel position $(i, j) \in [0, H-1] \times [0, W-1]$ over the full video duration $[0, T]$. Figure 1 shows that mean optical flow (first two columns) oversmooths motion patterns over

Table 2: Comparison of performance of optical and information flows to segment moving camouflaged animals over a range of IoU cutoffs. Bold highlights the best value per row.

Cutoffs	**Single-scale optflow**	**Multi-scale optflow**	**max. CC**	**cGC**	**DDC**
mAP_5	19.9	19.4	10.9	17.6	**22.8**
IoU_5	4.0	2.1	2.2	**4.7**	3.3
mAP_{15}	6.6	8.0	8.5	13.5	**13.6**
IoU_{15}	3.8	2.1	2.2	**4.6**	3.1
mAP_{25}	3.8	3.8	5.6	**9.7**	7.6
IoU_{25}	3.4	1.8	1.9	**4.2**	2.7

time. Transiently occurring patterns or those patterns that induce less pixel displacement becoming weaker and less intense, the longer the video duration. In contast, our spatiotemporal information flows better attends to the distinct motion patterns, giving more equal intensities to and retaining distinct patterns over time as evidenced by its ability to recover 'information highways'. Notably, due to differences in the implicit and explicit modeling assumptions of individual causal measures, the different information flow implementations exhibit differences in their behavior. Maximum cross-correlation (max. CC) appears similar to mean optical flow with more homogeneous representation of dynamical patterns however it seems sensitive to motion artifact. In the multi-storey example (Row 3) max.CC over enhances the static buildings and for fish swarming (Row 4) it fails to find the dominant anti-clockwise swirling in the second half of the video. On the other hand, conditional Granger causality (cGC) is highly sensitive to find all information sources, as seen by imprintings of the zebra crossing (Row 2) and individual people (Row 5) in the final flow. Consequently it appears cGC is less able to 'weight' the relative intensity of all motion patterns which results in blurring or overemphasis of information highways in cases where the information is a composition of transient dynamical patterns. Notably, under cGC the full stadium (Row 6) and art gallery (Row 7) is highlighted. Lastly PCCA and DDC share similar outputs and appear to strike a balance between max.CC and cGC, as they are able to find dominant flow patterns whilst suppressing less interesting sources of motion variation. This is nicely seen in the case of the stadium where both highlight the walkway between the seated spectators instead of highlighting everything (optical flow and cGC) or nothing (max.CC).

4.1.2 Perturbed crowd flow information flows

To understand quantitatively the behavior between the information flows, we generated perturbed versions of the crowd flow segmentation dataset by inverting the intensities within or rotating 90, 180, 270 degrees the central 50x50 pixel region of each video. Figure 2 shows an example and the resulting computed optical and information flows. Visually all flows clearly detect a difference in the flow direction after region rotation. However flow differences was less apparent for intensity inversion. We thus posed the detection of the perturbed region as a bounding box detection problem and measured the detection performance of each flow based on thresholding the difference between unperturbed vs perturbed flow. As PCCA appeared qualitatively similar to DDC but was significantly slower to compute we did not include it for comparison. Table 1 summarises the detection results. As expected, when rotated, optical flow optimized for motion flow performs best however both max. CC and DDC perform competitively. Notably when pixel intensity was inverted such that the motion flow remained the same both max. CC and DDC were superior.

4.2 Information flows on diverse motion videos

We sought to next extract spatiotemporal information flow from videos exhibiting diverse dynamics and generated by different modalities to test generality, (Figure 3, see Appendix B for links to the datasets). On physics-engine simulated collisions of objects (Yi et al. [2018]) information flows accurately extracts the collision flow irrespective of whether the objects were initially present or subsequently left the frame whereas optical flow only tracked the corners of the objects, (Row 1). On complex human actions such as performing a cartwheel, information flow was capable of capturing body part motion to recreate a flow of key snapshot postures (Row 2). *Drosophila*, fruitfly embryos undergo characteristic, motion patterns during development for correct gene patterning (Alberts [2017]). These patterns occur in sequential order, transient, comprised of small and dramatic motions,

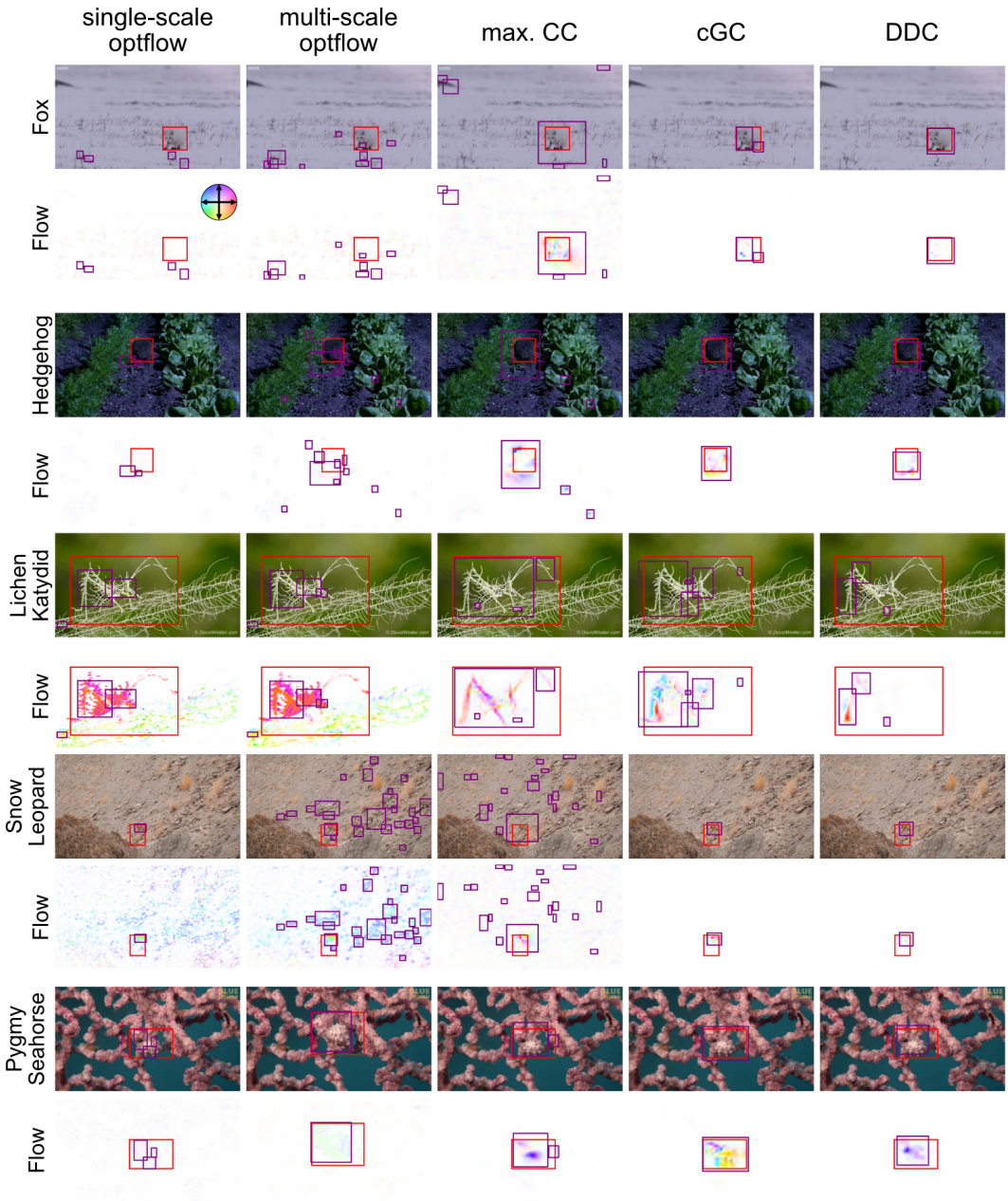

Figure 4: Comparison of the mean single- and multi-scale optical flow with spatiotemporal information flow detection of different camouflaged moving animals in video frames. Red box = ground truth. Purple box = detection by binary thresholding of flow and running connected components. Direction of flow is colored per the color wheel.

separated by periods of stationarity (Zhou [2017]). Mean optical flow and max. CC capture the largest motion patterns but PCCA and DDC are able to capture a superposition of all occurring patterns (Row 3). In cell division, where the majority of motion is due to stochastic morphological fluctuations and only a brief moment of physical division, only dynamic model based information flows (cGC, PCCA and DDC) recovered the circular division patterns that drive the salient biological motion. Lastly, we tested cell migration where individual cells actively change shape whilst they move (Row 5). Mean optical flow was unable to retain memory of the dynamics resulting in patchy flow. In contrast information flow is better able to extract spatiotemporally consistent histories of both the subcellular morphodynamics and the mean cell migration trajectory. In conclusion, information flows

are generally applicable and demonstrate the ability to retain history and uncover salient dynamical patterns even over long video durations and in the presence of complex nonlinear dynamics.

4.3 Moving camouflaged animals (MoCA) dataset

Last we compared the ability of optical and information flow to capture subtle movements in camouflaged animals. The moving camouflaged animals dataset is a comprehensive dataset comprising 141 videos of animals with bounding box annotations per 5 frames. We used 11 frames, 5 either side of the annotated frame and binary thresholding of the computed flow only followed by morphological closing to generate bounding boxes with no non-max bounding box suppression to test detection. The majority of these videos exhibit significant camera motion which requires sophisticated registration. To make it a fair test, we computed flow on a subset of 39 videos which we identified to have the least visible camera action. Table 2 shows the quantitative detection results at different IoU cutoffs. Notably cGC and DDC information flows significantly outperform and are more robust than optical flow with DDC reaching a mean average precision of 22.8 at IoU cutoff of 5. To investigate the low IoU cutoffs we additionally visualised the detections (purple boxes) relative to the ground truth bounding box (red box), Figure 4. We found that the low cutoff were justified due to the observation that information flows targetly aims to isolate the precise motion sources and sinks as evidenced by flow magnitude being localised and near to the moving arms, legs and heads of the animals. We note from the fox and hedgehog examples that cGC and DDC tended to compute less noisy flows even when motion was minimal and to track the bounding box with less background noise. In the lichen katydid example, computed information flows precisely exclude extraneous background motion to focus on the organism's moving appendages in all three spatiotemporal flows. A similar pattern is found in the snow lepard case. Lastly, all methods will perform reasonably well when there is clear relative motion of the animal to its background such as in the pygmy seahorse case.

5 Conclusion

In conclusion, we have introduced a formal framework notion of spatiotemporal information flows in this paper which generalised 1D causal measures to discover causal information relationships in 2D + time video. We demonstrate spatiotemporal flows possess an ability to capture community structure in complex dynamical systems, providing a potentially more robust and informative alternative to standard optical flow approaches. We implemented here our multiscale framework based on Gaussian image and pyramids and average pooling. Future work will investigate more optimum methods to combine information flows across individual scales and time and better adapt to the spatiotemporal complexity of the video e.g. to better handle intervals of minimal movement. In particular we hope to incorporate more complex dynamic models such as nonlinear dynamical systems.

References

Ronald J Adrian. Twenty years of particle image velocimetry. *Experiments in fluids*, 39(2):159–169, 2005.

Bruce Alberts. *Molecular biology of the cell*. WW Norton & Company, 2017.

Saad Ali and Mubarak Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6. IEEE, 2007.

Lionel Barnett and Anil K Seth. Granger causality for state-space models. *Physical Review E*, 91(4): 040101, 2015.

André M Bastos and Jan-Mathijs Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, 9:175, 2016.

Steven L Bressler and Anil K Seth. Wiener–granger causality: a well established methodology. *Neuroimage*, 58(2):323–329, 2011.

Yusi Chen, Burke Q Rosen, and Terrence J Sejnowski. Dynamical differential covariance recovers directional network structure in multiscale neural systems. *Proceedings of the National Academy of Sciences*, 119(24):e2117234119, 2022.

Matteo Detto, Annalisa Molini, Gabriel Katul, Paul Stoy, Sari Palmroth, and Dennis Baldocchi. Causality and persistence in ecological systems: a nonparametric spectral granger causality approach. *The American Naturalist*, 179(4):524–535, 2012.

Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2758–2766, 2015.

Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, pages 363–370. Springer, 2003.

André Fujita, Joao Ricardo Sato, Kaname Kojima, Luciana Rodrigues Gomes, Masao Nagasaki, Mari Cleide Sogayar, and Satoru Miyano. Identification and quantification of granger causality be-tween gene sets. *arXiv preprint arXiv:0911.1159*, 2009.

John F Geweke. Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association*, 79(388):907–915, 1984.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. HMDB: a large video database for human motion recognition. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2011.

Hala Lamdouar, Charig Yang, Weidi Xie, and Andrew Zisserman. Betrayed by motion: Camouflaged object discovery via motion segmentation. In *Proceedings of the Asian Conference on Computer Vision*, 2020.

Min Liu and Tobi Delbruck. Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors. 2018.

Sanath Narayan and Kalpathi R Ramakrishnan. A cause and effect analysis of motion trajectories for modeling actions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2633–2640, 2014.

Stephanie Noble, Dustin Scheinost, and R. Todd Constable. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, 203:116157, 2019. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2019.116157.

Karthir Prabhakar, Sangmin Oh, Ping Wang, Gregory D Abowd, and James M Rehg. Temporal causality for the analysis of visual events. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1967–1974. IEEE, 2010.

Anil K Seth, Adam B Barrett, and Lionel Barnett. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–3297, 2015.

Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. Ucf101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*, 2012.

George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *science*, 338(6106):496–500, 2012.

Shuyang Sun, Zhanghui Kuang, Lu Sheng, Wanli Ouyang, and Wei Zhang. Optical flow guided feature: A fast and robust motion representation for video action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1390–1399, 2018.

Eran Swears, Anthony Hoogs, Qiang Ji, and Kim Boyer. Complex activity recognition using granger constrained dbn (gcdbn) in sports and surveillance video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

Raju Tomer, Khaled Khairy, Fernando Amat, and Philipp J Keller. Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. *Nature methods*, 9(7):755–763, 2012.

Vladimír Ulman, Martin Maška, Klas EG Magnusson, Olaf Ronneberger, Carsten Haubold, Nathalie Harder, Pavel Matula, Petr Matula, David Svoboda, Miroslav Radojevic, et al. An objective comparison of cell-tracking algorithms. *Nature methods*, 14(12):1141–1152, 2017.

Pedro A Valdes-Sosa, Alard Roebroeck, Jean Daunizeau, and Karl Friston. Effective connectivity: influence, causality and biophysical modeling. *Neuroimage*, 58(2):339–361, 2011.

Norbert Wiener. The theory of prediction. *Modern mathematics for engineers*, 1956.

Yuya Yamashita, Tatsuya Harada, and Yasuo Kuniyoshi. Causal flow. *IEEE Transactions on multimedia*, 14(3):619–629, 2012.

Kexin Yi, Jiajun Wu, Chuang Gan, Antonio Torralba, Pushmeet Kohli, and Josh Tenenbaum. Neural-symbolic vqa: Disentangling reasoning from vision and language understanding. *Advances in neural information processing systems*, 31, 2018.

Amos Zamir, Guanyu Li, Katelyn Chase, Robert Moskovitch, Bo Sun, and Assaf Zaritsky. Emergence of synchronized multicellular mechanosensing from spatiotemporal integration of heterogeneous single-cell information transfer. *Cell Systems*, 13(9):711–723.e7, 2022. ISSN 2405-4712. doi: https://doi.org/10.1016/j.cels.2022.07.002.

Felix Zhou. *Phenotyping cellular motion*. PhD thesis, University of Oxford, 2017.

Felix Y Zhou, Carlos Ruiz-Puig, Richard P Owen, Michael J White, Jens Rittscher, and Xin Lu. Motion sensing superpixels (moses) is a systematic computational framework to quantify and discover cellular motion phenotypes. *elife*, 8:e40162, 2019.

A Information flow at different scales

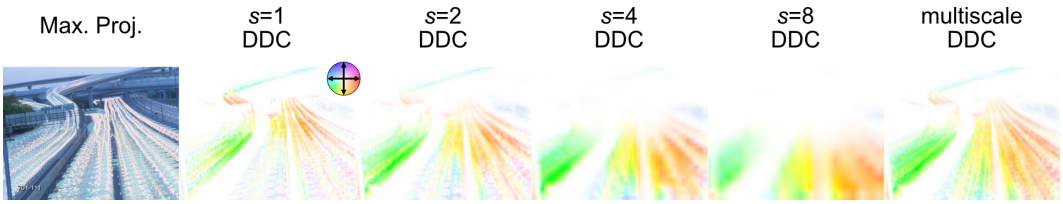

Figure 5: Visualisation of the individual information flow at each of the 4 dowsampling factors compared to the multiscale flow, the average of the 4 for dynamic differential covariance flow. Direction and magntiude of flow is colored per the color wheel.

B Datasets used

Below are links to the datasets used in this paper.

- Crowd flow segmentation dataset (Ali and Shah [2007]), https://www.crcv.ucf.edu/research/data-sets/crowd-segmentation.
- CLEVRER: CoLlision Events for Video REpresentation and Reasoning dataset for videos of collisions, (Yi et al. [2018]). http://clevrer.csail.mit.edu/.
- HMDB: a large human motion database for cartwheel example, (Kuehne et al. [2011]). https://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/#Downloads.
- Tomer et al. [2012] supplementary movie 3 for *Drosophila* fruitfly embryo development example.
- HeLa cells on a flat glass (DIC-C2DH-HeLa.zip) for cell divison and Glioblastoma-astrocytoma U373 cells on a polyacrylamide substrate (PhC-C2DH-U373.zip) from the cell tracking challenge, (Ulman et al. [2017]). http://celltrackingchallenge.net/2d-datasets/
- MoCA : Moving camouflaged animals dataset, (Lamdouar et al. [2020]). https://www.robots.ox.ac.uk/ vgg/data/MoCA/.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]