Entropy Minimisation Framework for Event-based Vision Model Estimation - Supplementary Material

Urbano Miguel Nunes and Yiannis Demiris

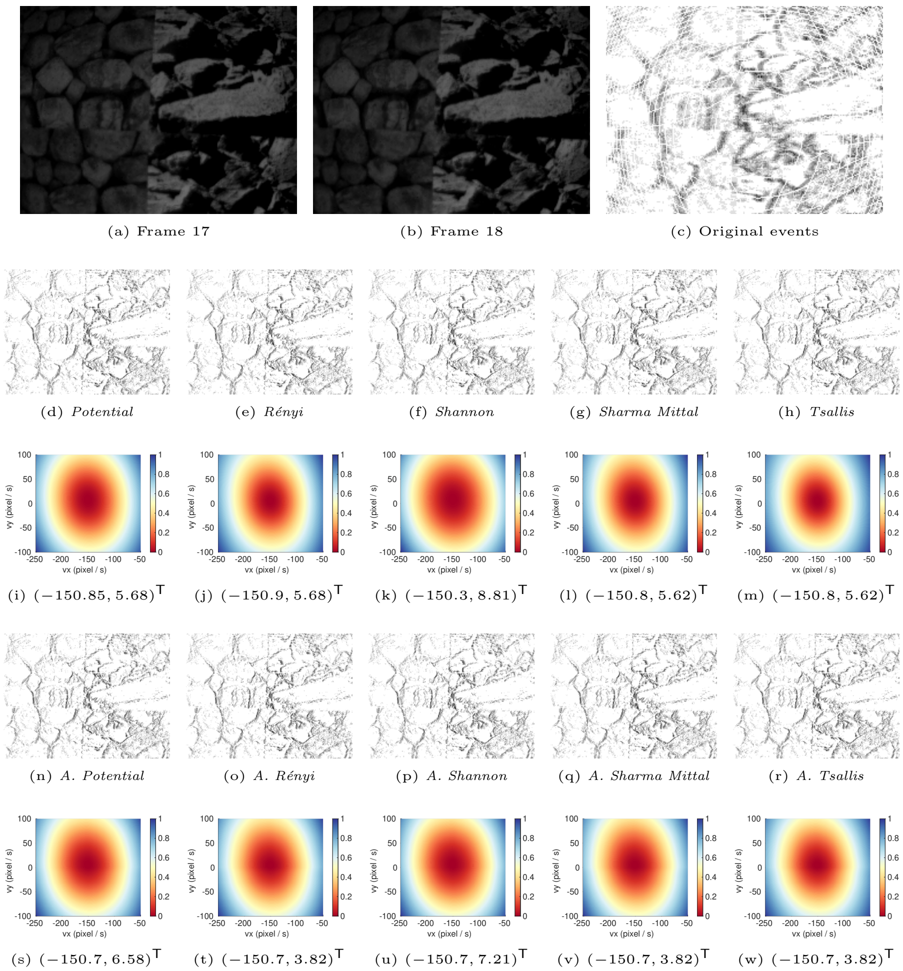
Personal Robotics Laboratory, Department of Electrical and Electronic Engineering, Imperial College London, London, UK {um.nunes, y.demiris}@imperial.ac.uk

1 Practical Considerations

The current implementation is designed to favour usability and scalability, rather than performance due to developmental reasons. Our framework converges on average in 4-5 line-searches for rotational motion estimation considering 20000 events. These values may be different by running the code provided since the initial conditions are far from the optimum. The time cost of evaluating the approximate entropy function and corresponding derivatives is $\sim$15ms and $\sim$25ms, respectively, on a single-threaded 3.7Ghz CPU. For comparison, our implementation of the Variance cost function [1] and corresponding derivatives takes $\sim$12ms and $\sim$21ms to be evaluated in the same settings, respectively. The framework is highly parallelisable and further speed-ups can be achieved. This includes the exact entropy, which can also achieve competitive computational costs for specific problems where a few $\sim$ 1000 of events are processed.

2 Additional Results on Optical Flow Estimation

Fig. 9 presents additional plots regarding optical flow estimation using the proposed framework. The optical flow is estimated between two consecutive frames of the `poster_translation` sequence [2]. In Fig. 9c, we show the original events, that were produced by the motion between the frames, projected onto the IWE without motion compensation. The resulting image is blurry exhibiting low contrast. By estimating the optical flow model parameters, if we project the modelled events onto the IWE, we get a sharper image with higher contrast. Similar results are obtained using the Contrast Maximisation (CMax) framework [1]. Distinctively, however, our framework does not directly optimise the contrast measure of the IWE. Instead, it minimises the dispersion of the modelled events in the feature space. Additionally, the measure profiles of the exact functions and respective approximations in function of the optical flow parameters are almost identical.



(a) Frame 17 (b) Frame 18 (c) Original events

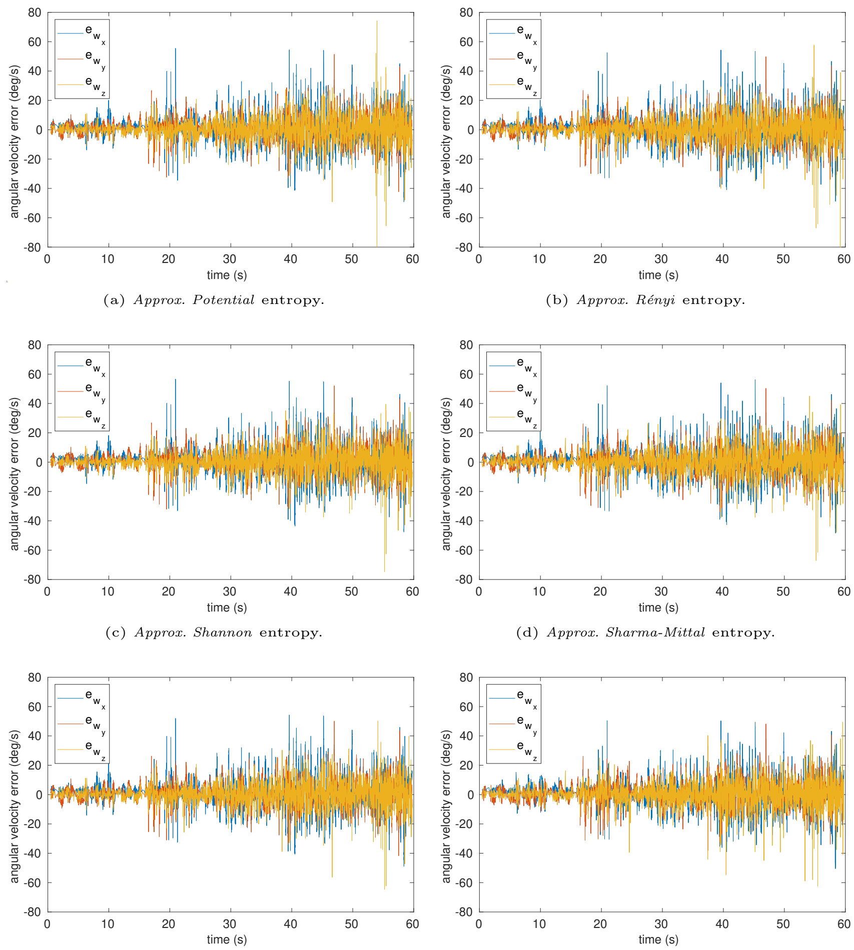
(d) *Potential* (e) *Rényi* (f) *Shannon* (g) *Sharma Mittal* (h) *Tsallis*

(i) $(-150.85, 5.68)^\mathsf{T}$ (j) $(-150.9, 5.68)^\mathsf{T}$ (k) $(-150.3, 8.81)^\mathsf{T}$ (l) $(-150.8, 5.62)^\mathsf{T}$ (m) $(-150.8, 5.62)^\mathsf{T}$

(n) *A. Potential* (o) *A. Rényi* (p) *A. Shannon* (q) *A. Sharma Mittal* (r) *A. Tsallis*

(s) $(-150.7, 6.58)^\mathsf{T}$ (t) $(-150.7, 3.82)^\mathsf{T}$ (u) $(-150.7, 7.21)^\mathsf{T}$ (v) $(-150.7, 3.82)^\mathsf{T}$ (w) $(-150.7, 3.82)^\mathsf{T}$

Fig. 9. Optical flow estimation between frames 17 and 18 of the `poster_translation` sequence [2]. The flow is dominated by the horizontal component: $\boldsymbol{\theta}^* \approx (-150, 0)^\mathsf{T}$. All methods tend to maximise the IWE contrast, which is reflected in the motion-compensated images (d)-(h) and (n)-(r). (i)-(m) Normalised profiles for the proposed *Potential* energy and exact entropy loss functions. (s)-(w) Normalised profiles for the proposed efficient approximate loss functions. The normalised profiles are plotted in function of the optical flow parameters. We can see that all of the proposed functions produce similar results, both in terms of contrast measures and model parameters estimated



(a) *Approx. Potential* entropy.

(b) *Approx. Rényi* entropy.

(c) *Approx. Shannon* entropy.

(d) *Approx. Sharma-Mittal* entropy.

(e) *Approx. Tsallis* entropy.

(f) Variance [1].

Fig. 10. Angular velocity estimation errors using the proposed approximate entropy functions and the variance [1] on the `poster_rotation` sequence [2], considering batches of 20000 events

3 Additional Results on Rotational Estimation

Fig. 10 shows additional angular velocity estimation errors using the Approximate Entropy Minimisation (AEMin) framework on the `poster_rotation` sequence [2], considering batches of 20000 events. The errors are small compared to the range of angular velocities undergone in the sequence. Table 4 provides additional quantitative results.

Table 4. Accuracy comparison of the proposed entropy loss functions on the `poster_rotation` sequence [2]. The angular velocity errors for each component ($e_{w_x}, e_{w_y}, e_{w_z}$), their standard deviation (σ_{e_w}) and RMS are presented in deg/s, w.r.t. IMU measurements, considering batches of 20000 events. The RMS error compared to the maximum excursions are also presented in percentage (RMS %), as well as the absolute and relative maximum errors. The best value per column is highlighted in bold

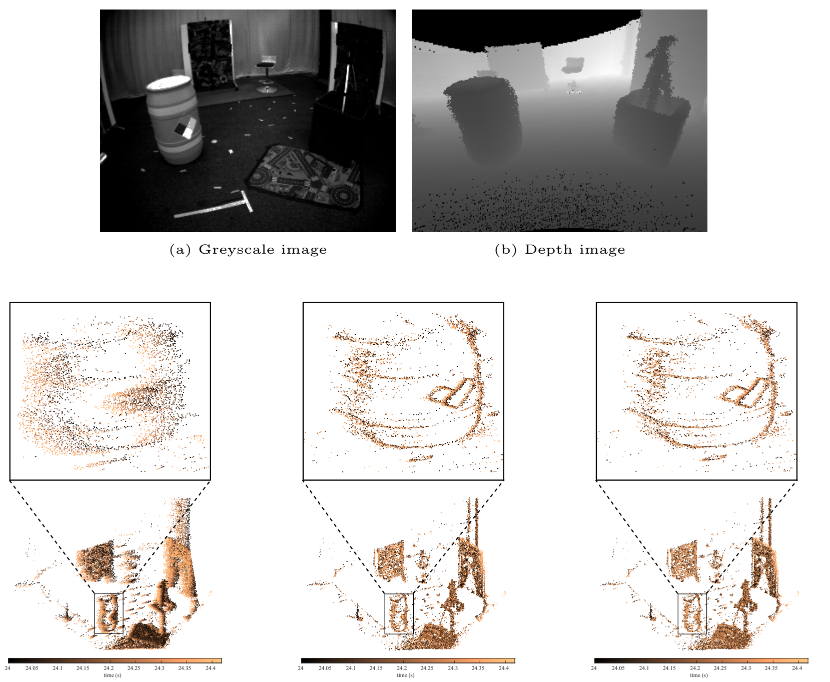
Function	e_{w_x}	e_{w_y}	e_{w_z}	σ_{e_w}	RMS	RMS %	max	max %
Approx. Potential	14.08	8.91	8.51	14.22	14.30	1.52	90.70	9.65
Approx. Rényi	14.31	9.16	8.34	14.24	14.30	1.52	88.16	9.38
Approx. Shannon	**13.92**	**8.78**	8.71	**14.07**	**14.17**	**1.51**	74.74	7.95
Approx. Sharma-Mittal	14.32	9.15	**8.32**	14.16	14.22	1.51	67.11	7.14
Approx. Tsallis	14.31	9.16	8.34	14.16	14.22	1.51	**64.68**	**6.88**

4 Additional Results on 6-DOF Estimation

In Fig. 11, we compare the estimated 6-DOF motion parameters and the resulting modelled 3D events according to the *Rényi* entropy and its approximate entropy, from events generated of the `indoor_flying1` sequence [3] (for illustration purposes, we consider the first 75000 events at the 24 second timestamp). The resulting modelled 3D events are almost identical (Figs. 11d and 11e) and the estimated 6-DOF parameters are also very similar. This corroborates that the proposed efficient approximate entropy functions provide a valid trade-off between accuracy and computational complexity. Table 5 provides additional quantitative results.

5 Entropy Minimisation Profiles

Up to this point, all the loss functions profiles in function of the parameter models seem to suggest that there is a close similarity between the exact loss functions and respective approximations profiles. In Fig. 12, we show the projected normalised loss functions profiles for a synthetic data sequence, which simulates the corners of a square moving perpendicularly to the image plane. We can observe that, even though all profiles exhibit a clear minimum at the



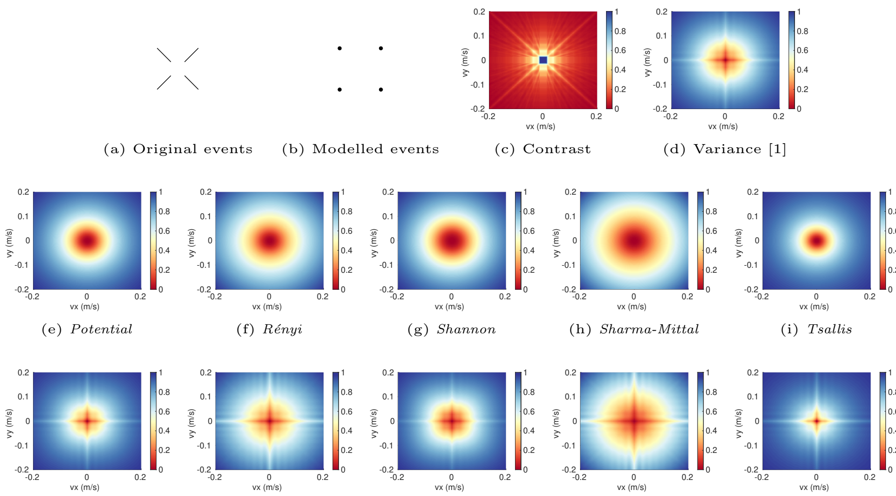
(a) Greyscale image (b) Depth image

(c) Original 3D events (d) 3D events modelled by *Rényi* (e) 3D events modelled by *A. Rényi*

Fig. 11. 6-DOF estimation from 75000 events in 3D space of the `indoor_flying1` sequence [3]. (a) Example greyscale image of the scene and (b) corresponding depth. (c) Original 3D events generated from the moving camera. (d)-(e) The events can be modelled directly in 3D to retrieve the 6-DOF motion parameters, as well as to recover the 3D shape of the objects. According to the *Rényi* entropy, the estimated 6-DOF parameters are: $\mathbf{w} = (0.003, -0.174, 0.027)^\intercal$, $\mathbf{v} = (-0.017, -0.102, 0.095)^\intercal$. Similar parameters are obtained according to its approximate entropy: $\mathbf{w} = (0.007, -0.172, 0.025)^\intercal$, $\mathbf{v} = (-0.023, -0.093, 0.096)^\intercal$

Table 5. Accuracy comparison of the proposed entropy loss functions on the `indoor_flying1` and `indoor_flying4` and `outdoor_driving_night1` sequences [3]. The average angular and linear velocity errors (e_w, e_v), their standard deviation ($\sigma_{e_w}, \sigma_{e_v}$) and RMS ($\mathrm{RMS}_{e_w}$, RMS_{e_v}) are presented in deg/s and m/s, respectively. The RMS errors compared to the maximum excursions are also presented in percentage (RMS_{e_w} %, RMS_{e_v} %). For each sequence, the best value per column is highlighted in bold

Sequence	**Function**	e_w	e_v	σ_{e_w}	σ_{e_v}	RMS_{e_w}	RMS_{e_v}	RMS_{e_w} %	RMS_{e_v} %
`indoor_flying1`	*Potential*	2.45	0.14	3.19	0.18	3.34	0.19	8.23	22.12
	Rényi	2.23	0.13	2.89	0.16	3.06	0.17	7.53	19.83
	Shannon	2.25	0.13	2.90	0.16	3.06	0.17	7.55	19.82
	Sharma-Mittal	2.25	0.13	2.90	0.16	3.06	0.17	7.53	19.97
	Tsallis	2.22	0.13	2.87	0.16	3.03	0.17	7.47	19.88
	Approx. Potential	2.20	0.10	2.79	0.12	3.02	0.13	7.44	15.31
	Approx. Rényi	2.16	**0.09**	2.71	**0.11**	2.93	**0.12**	7.22	**14.37**
	Approx. Shannon	2.24	0.10	2.80	0.12	3.04	0.13	7.50	15.60
	Approx. Sharma-Mittal	2.17	**0.09**	2.76	0.12	2.97	0.13	7.31	14.75
	Approx. Tsallis	**2.08**	0.10	**2.40**	0.12	**2.70**	0.12	**6.66**	14.80
`indoor_flying4`	*Potential*	4.38	0.28	5.50	0.34	5.52	0.35	22.11	18.79
	Rényi	4.28	0.25	5.31	0.30	5.30	0.30	21.25	16.37
	Shannon	4.18	**0.23**	5.24	0.29	5.21	**0.29**	20.89	**15.66**
	Sharma-Mittal	4.10	0.24	5.24	0.30	5.21	0.30	20.90	16.30
	Tsallis	**4.08**	0.25	**5.16**	0.30	**5.12**	0.30	**20.53**	16.38
	Approx. Potential	4.65	**0.23**	5.40	**0.28**	5.70	0.29	22.85	15.97
	Approx. Rényi	4.69	**0.23**	5.68	0.29	5.93	0.30	23.78	16.08
	Approx. Shannon	5.18	0.24	6.07	0.30	6.57	0.31	26.34	16.69
	Approx. Sharma-Mittal	4.67	**0.23**	5.42	**0.28**	5.82	0.29	23.30	15.86
	Approx. Tsallis	4.43	**0.23**	5.30	0.29	5.56	0.30	22.28	16.35
`outdoor_driving_night1`	*Potential*	4.75	1.86	15.18	1.88	15.30	2.42	4.10	23.85
	Rényi	4.20	**1.72**	14.75	1.70	14.87	**2.21**	3.98	**21.76**
	Shannon	**4.13**	1.73	**13.83**	**1.68**	**13.98**	2.21	**3.74**	21.80
	Sharma-Mittal	4.21	1.73	14.76	1.72	14.88	2.22	3.99	21.85
	Tsallis	4.18	1.73	14.73	1.71	14.85	2.22	3.98	21.85
	Approx. Potential	6.18	1.77	16.42	1.84	16.44	2.36	4.40	23.26
	Approx. Rényi	7.18	1.83	17.86	1.93	17.85	2.40	4.78	23.66
	Approx. Shannon	5.91	1.77	15.71	1.74	15.75	2.30	4.22	22.64
	Approx. Sharma-Mittal	7.29	1.93	17.79	2.02	17.79	2.54	4.76	25.04
	Approx. Tsallis	7.27	1.82	17.97	1.94	17.97	2.41	4.81	23.71



(a) Original events (b) Modelled events (c) Contrast (d) Variance [1]

(e) *Potential* (f) *Rényi* (g) *Shannon* (h) *Sharma-Mittal* (i) *Tsallis*

(j) *A. Potential* (k) *A. Rényi* (l) *A. Shannon* (m) *A. Sharma-Mittal* (n) *A. Tsallis*

Fig. 12. Projected normalised loss functions profiles at $\bar{v}_z = -1$ (m/s) and $\mathbf{n} = (0, 0, 1)^\mathsf{T}$ using the translational model on a synthetic data sequence. (a) Original events projected onto the IWE, which simulate the corners of a square moving perpendicularly to the image plane with $v_z = -1$ (m/s). (b) Modelled events projected onto the IWE, according to the optimal parameters $\bar{\mathbf{v}} = (0, 0, -1)^\mathsf{T}$ (m/s) and $\mathbf{n} = (0, 0, 1)^\mathsf{T}$. (c) IWE contrast profile, which exhibits a clear peak at $\bar{\mathbf{v}} = (0, 0, -1)^\mathsf{T}$ (m/s); however, due to the image discretisation, the profile is not smooth and not suitable to be used in the optimisation framework. In contrast, the rest of the profiles shown are smooth and exhibit a clear minimum at the optimal model parameters. We can also observe that the profiles of the proposed approximate functions are not *as smooth as* the profiles of the respective exact loss functions. Interestingly, the (d) variance profile [1] is similar to the profiles of the proposed approximate entropy functions, *e.g.* (k) *A. Rényi*

optimal model parameters, there is a clear distinction between the profiles of the exact loss functions and respective proposed approximations, *i.e.* the former exhibit a *smoother* landscape. Additionally, the variance loss function profile in Fig. 12d closely matches the proposed approximate entropy loss functions profiles, *e.g. A. Rényi* entropy function profile shown in Fig. 12k.

6 Additional Models

In this section, we provide additional models that are useful for common tasks in computer vision, which can be estimated using the proposed framework.

Isometry Transformation Estimation: The isometry transformation has 3-DOF and can be parameterised by the 2D linear velocity on the image plane $\mathbf{v} = (v_x, v_y)^\mathsf{T}$ and the angular velocity w, being expressed as

$$\mathbf{f}_k = \mathcal{M}(e_k; \boldsymbol{\theta}) \propto \mathcal{I}^{-1}(t_k; \boldsymbol{\theta}) \begin{pmatrix} \mathbf{x}_k \\ 1 \end{pmatrix}, \tag{27}$$

where $\boldsymbol{\theta} = (w, \mathbf{v}^\mathsf{T})^\mathsf{T}$ are the model parameters. The isometry matrix $\mathcal{I}$ can be written in function of the model parameters $\boldsymbol{\theta}$ as

$$\mathcal{I}(t_k; \boldsymbol{\theta}) = \begin{bmatrix} \mathcal{R}(t_k; w) & \Delta t_k \mathbf{v} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}, \quad \mathcal{R}(t_k; w) = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \in SO(2), \tag{28}$$

where $\phi = \Delta t_k w$.

Similarity Transformation Estimation: The similarity transformation is an isometry transformation with an additional isotropic scaling $\lambda = 1 + \Delta t_k s$, being expressed as

$$\mathbf{f}_k = \mathcal{M}(e_k; \boldsymbol{\theta}) \propto \mathcal{S}^{-1}(t_k; \boldsymbol{\theta}) \begin{pmatrix} \mathbf{x}_k \\ 1 \end{pmatrix}, \tag{29}$$

where $\boldsymbol{\theta} = (w, s, \mathbf{v}^\mathsf{T})^\mathsf{T}$ are the model parameters. The similarity matrix $\mathcal{S}$ can be written in function of the model parameters $\boldsymbol{\theta}$ as

$$\mathcal{S}(t_k; \boldsymbol{\theta}) = \begin{bmatrix} (1 + \Delta t_k s)\mathcal{R}(t_k; w) & \Delta t_k \mathbf{v} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}. \tag{30}$$

Affine Transformation Estimation: The affinity is a similarity transformation with non-isotropic scaling, which is accounted by two additional DOF, namely, the angle $\gamma = \Delta t_k w_\gamma$ that specifies the scaling direction and the ratio between the scaling parameters $\lambda_1/\lambda_2 = (1 + \Delta t_k s_1)/(1 + \Delta t_k s_2)$. It is expressed as

$$\mathbf{f}_k = \mathcal{M}(e_k; \boldsymbol{\theta}) \propto \mathcal{A}^{-1}(t_k; \boldsymbol{\theta}) \begin{pmatrix} \mathbf{x}_k \\ 1 \end{pmatrix}, \tag{31}$$

where $\boldsymbol{\theta} = (w, w_\gamma, s_1, s_2, \mathbf{v}^\mathsf{T})^\mathsf{T}$ are the model parameters. The affinity matrix $\mathcal{A}$ can be written in function of the model parameters $\boldsymbol{\theta}$ as

$$\mathcal{A}(t_k;\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{R}(t_k;w)\mathcal{R}(t_k;w_\gamma)^\mathsf{T}\mathbf{D}\mathcal{R}(t_k;w_\gamma) & \Delta t_k \mathbf{v} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}, \tag{32}$$

where $\mathbf{D}$ is a diagonal matrix, whose entries are the scaling parameters λ_1 and λ_2.

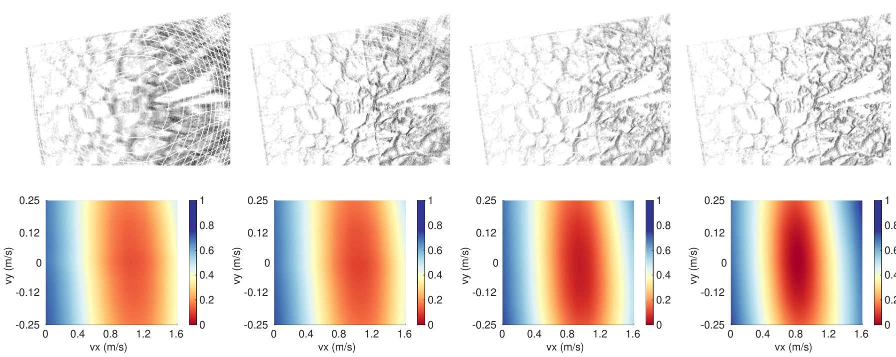
Translational Motion Estimation: By not considering the rotational part of the homography model (*i.e.* letting $\mathbf{w} = (0,0,0)^\mathsf{T}$), we can just consider estimating the translational motion. This model has 5-DOF and can be expressed as

$$\mathbf{f}_k = \mathcal{M}(e_k;\boldsymbol{\theta}) \propto \mathcal{T}^{-1}(t_k;\boldsymbol{\theta}) \begin{pmatrix} \mathbf{x}_k \\ 1 \end{pmatrix}, \tag{33}$$

where $\boldsymbol{\theta} = (\bar{\mathbf{v}}^\mathsf{T}, \mathbf{n}^\mathsf{T})^\mathsf{T}$ are the model parameters and the translation matrix $\mathcal{T}$ can be written in function of the model parameters $\boldsymbol{\theta}$ as

$$\mathcal{T}(t_k;\boldsymbol{\theta}) = \mathbf{I} - \Delta t_k \bar{\mathbf{v}}\mathbf{n}^\mathsf{T}, \tag{34}$$

where $\mathbf{I}$ corresponds to the identity matrix.



(e) $\bar{v}_z = 0$ (m/s) (f) $\bar{v}_z = -0.12$ (m/s) (g) $\bar{v}_z = -0.53$ (m/s) (h) $\bar{v}_z = -1.04$ (m/s)

Fig. 13. Example of iterations of the proposed framework optimisation procedure using the *A. Rényi* entropy. On the top row (from left to right), we show the evolution of the resultant IWE, according to the parameters being estimated. On the bottom row, we can see the projection of the normalised *A. Rényi* entropy profile at the value presented in the caption and $\mathbf{n} = (0,0,1)^\mathsf{T}$. The profile is smooth and has one clear minimum, which corresponds to the optimal parameters

In Fig. 13, we exemplify some iterations of the proposed framework optimisation procedure. By solving for the optimal model parameters, the framework minimises the modelled events' dispersion according to the entropy measure (bottom row). As presented, the *Approximate Rényi* entropy profile is also smooth and has one clear minimum, which corresponds to the optimal parameters.

7 Truncated Kernel Size

Table 6. Accuracy comparison of the proposed efficient *Approximate Tsallis* loss function on the `poster_rotation` sequence [2] in function of the truncated kernel size. The angular velocity errors for each component ($e_{w_x}, e_{w_y}, e_{w_z}$), their standard deviation (σ_{e_w}) and RMS are presented in deg/s, w.r.t. IMU measurements, considering batches of 20000 events. The RMS error compared to the maximum excursions are also presented in percentage (RMS %), as well as the absolute and relative maximum errors. The best value per column and per function is highlighted in bold

Size	e_{w_x}	e_{w_y}	e_{w_z}	σ_{e_w}	RMS	RMS %	max	max %
3	14.31	9.16	**8.34**	**14.16**	**14.22**	**1.51**	**64.68**	**6.88**
5	14.31	9.17	8.43	14.44	14.51	1.54	130.98	13.93
7	14.30	**9.15**	8.45	14.34	14.40	1.53	81.52	8.67
9	**14.29**	9.16	8.35	14.22	14.28	1.52	71.79	7.64

The truncated kernel size affects how each event's influence is spread across the discretised space. For example, a larger size means that each event can influence a larger region in space. Table 6 presents quantitative results regarding the performance of the proposed approximate measures, in function of the truncated kernel size. These results show that the truncated kernel size does not significantly influence the performance of the proposed optimisation framework using the proposed AEMin. Thus, fixing it to 3 lowers the computational complexity of the overall framework, without performance degradation (the framework's complexity using AEMin is $O(N_e \kappa^d)$, where κ is the truncated kernel size which may be fixed to 3).

8 Mathematical Utilities

8.1 Skew-symmetric matrices

A skew-symmetric matrix is a square matrix whose transpose equals its negative

$$\mathbf{A}^\mathsf{T} = -\mathbf{A}. \tag{35}$$

In particular, given a 3D vector $\mathbf{a} = (a_1, a_2, a_3)^\mathsf{T}$, we can construct a 3×3 skew-symmetric matrix as

$$\hat{\mathbf{a}} \coloneqq \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}. \tag{36}$$

8.2 Spherical Coordinates

Spherical coordinates are a system of coordinates that naturally describe positions on a sphere. These are defined by the longitude $\psi \in [0, 2\pi]$, latitude $\phi \in [0, \pi]$ and radius $r \geq 0$. The spherical coordinates (r, ψ, ϕ) are related to the Cartesian coordinates (x, y, z) by

$$\begin{cases} r = \sqrt{x^2 + y^2 + z^2} \\ \psi = \arctan \frac{y}{x} \\ \phi = \arccos \frac{z}{r} \end{cases} \Leftrightarrow \begin{cases} x = r \cos \psi \sin \phi \\ y = r \sin \psi \sin \phi \\ z = r \cos \phi \end{cases} . \tag{37}$$

References

1. Gallego, G., Rebecq, H., Scaramuzza, D.: A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3867–3876 (2018) 1, 3, 7
2. Mueggler, E., Rebecq, H., Gallego, G., Delbruck, T., Scaramuzza, D.: The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam. International Journal of Robotics Research **36**(2), 142–149 (2017) 1, 2, 3, 4, 10
3. Zhu, A.Z., Thakur, D., Özaslan, T., Pfrommer, B., Kumar, V., Daniilidis, K.: The multivehicle stereo event camera dataset: An event camera dataset for 3d perception. IEEE Robotics and Automation Letters **3**(3), 2032–2039 (2018) 4, 5, 6