Motion Models That Only Work Sometimes

BMVC 2012

C. Garcia Cifuentes, M. Sturzel, F. Jurie, G. J. Brostow

Concept

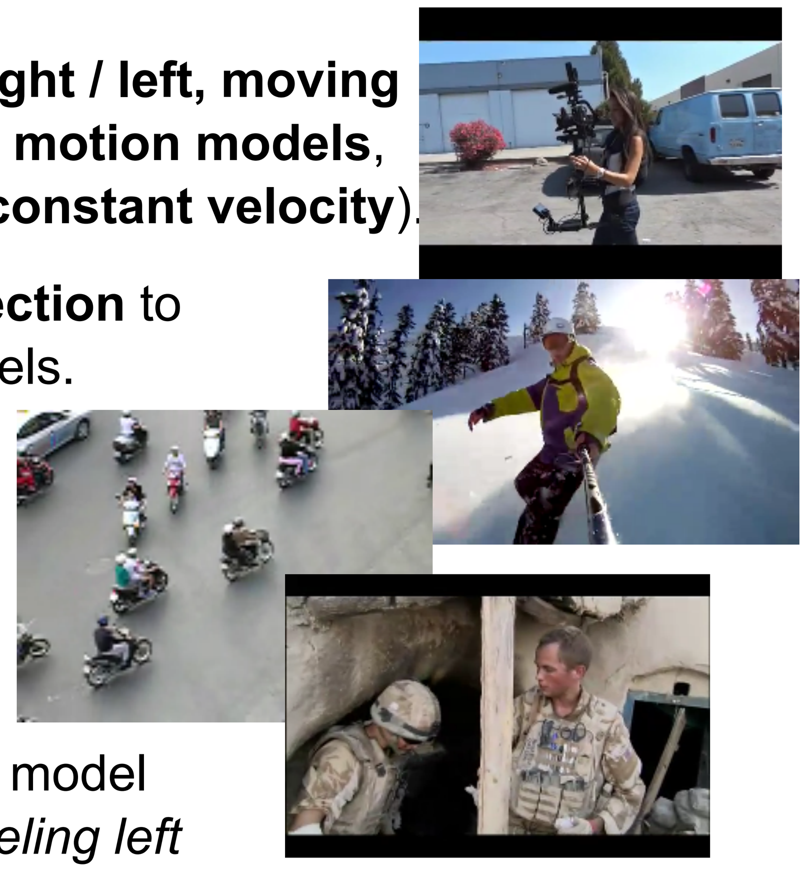
Tracking of 2D points can be improved by using an appropriate motion model in each situation. Can we predict such a suitable model for a query video given training videos in different scenes? Does the prediction improve tracking performance on average?

The motion model should account for **changes in speed and scale** in this scenario.

Method overview

TRAINING

1. Establish four rough categories of camera motion (**traveling right / left, moving forward / backward**). Design and implement four **specialized motion models**, plus two standard models for tracking (**Brownian** motion and **constant velocity**).
2. Collect a **dataset** of challenging videos. Assign them **by inspection** to one of the six categories, corresponding to the six motion models.
3. Extract local features, compute bag-of-words representations. Train an **SVM** using the labels obtained by inspection.



query clip

4. feature extraction [Wang et al. CVPR'11] → clip descriptor → 5. model predictor → motion model *e.g. traveling left*

6. tracking (particle filter)

$$\underbrace{p(\mathbf{x}_t|\mathbf{I}_{1:t})}_{\text{posterior}} = \underbrace{p(\mathbf{I}_t|\mathbf{x}_t)}_{\text{measurement model}} \int \underbrace{p(\mathbf{x}_t|\mathbf{x}_{t-1})}_{\text{dynamic model}} \underbrace{p(\mathbf{x}_{t-1}|\mathbf{I}_{1..t-1})}_{\text{prior}} d\mathbf{x}_{t-1}$$

7. robustness evaluation (manual landmarks)

$$robustness = \frac{1}{N}\sum_{i=1}^{N} \triangle_\theta(p_i, \hat{p}_i);$$

$$\triangle_\theta(p_i, \hat{p}_i) = \begin{cases} 1 & \text{if } |x_i - \hat{x}_i| < \theta \text{ and } |y_i - \hat{y}_i| < \theta \\ 0 & \text{otherwise.} \end{cases}$$

Results

Tracking performance of each motion model over all videos in the dataset.

Best-case results if model is selected by either a performance-based oracle, or our inspection-based labels.

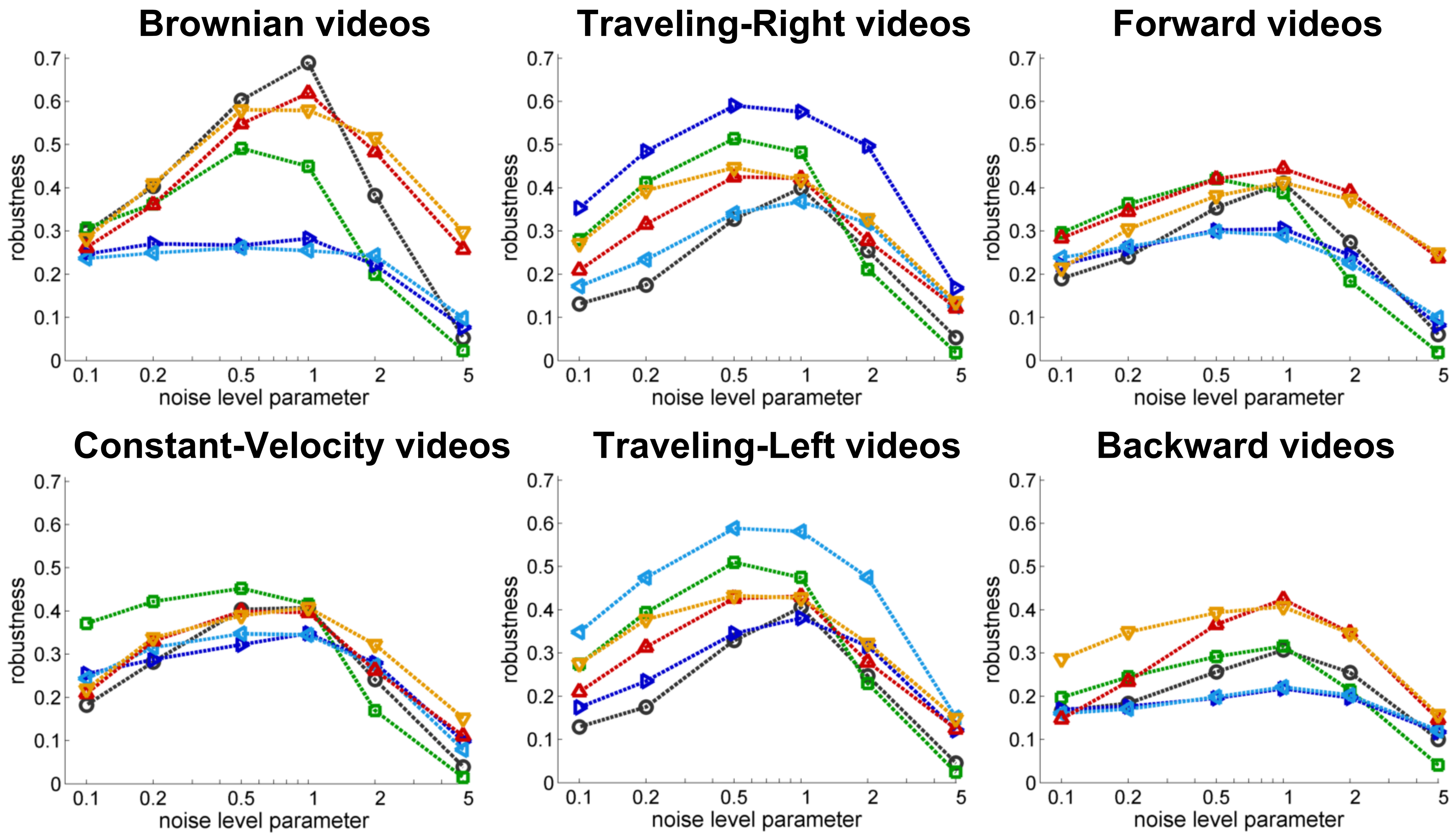
Tracking each video using the motion model predicted by our classifier, trained on inspection-based labels, evaluated leaving one video out.

	Fixed motion model						Ideal predictions			Our method
	Br	CVel	TRight	TLeft	Fwd	Bwd	best{Br, CVel}	best{all}	inspection labels	
tracking robustness ($\cdot 10^{-2}$)	42.3	43.2	37.9	37.2	44.7	43.7	49.3	56.1	52.6	51.9
$\pm$ std. dev. random runs	0.4	0.4	0.7	0.5	0.2	0.1	0.6	0.4	0.2	0.1
% best choice (± 2)	21	11	12	20	20	16	32	100	52	50

Bottom row: percentage of times a given strategy was the best choice among the six motion models.

Performance of each motion model on each group of videos.

Br TRight Fwd CVel TLeft Bwd



- Groups are partial to their respective motion models.
- Motion models differ in robustness to parameter settings.
- Noise parameter: constant factor applied to all standard deviations in the motion models (default: 1).

Motion models

Brownian :

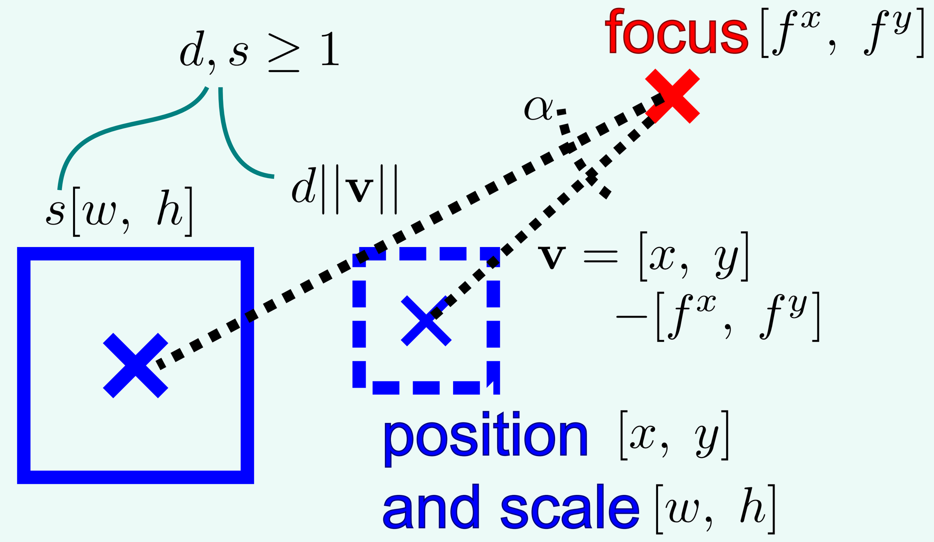
$$[x_{t+1},\ y_{t+1}] \sim \mathrm{Norm}\left([x_t,\ y_t], \mathrm{diag}(\sigma_x^2, \sigma_y^2)\right),$$
$$[w_{t+1},\ h_{t+1}] = s[w_t,\ h_t]; \quad s \sim \mathrm{Norm}(1, \sigma_s^2).$$

Constant Velocity,
Traveling Right, Traveling Left :

$$[\dot{x}_{t+1},\ \dot{y}_{t+1}] \sim \mathrm{Norm}\left([\dot{x}_t,\ \dot{y}_t], \mathrm{diag}(\sigma_x^2, \sigma_y^2)\right),$$
$$[x_{t+1},\ y_{t+1}] = [x_t,\ y_t] + [\dot{x}_{t+1},\ \dot{y}_{t+1}],$$
$$[w_{t+1},\ h_{t+1}] = s_{t+1}[w_t,\ h_t]; \quad s_{t+1} \sim \mathrm{Norm}(s_t, \sigma_s^2).$$

Forward, Backward :

$$[f^x_{t+1},\ f^y_{t+1}] \sim \mathrm{Norm}([f^x_t,\ f^y_t], \sigma_f^2\mathbf{I}),$$
$$d_{t+1} = d_t + \sigma_d\epsilon_1; \quad \epsilon_1 \sim \mathrm{Norm}(0, 1.0),$$
$$s_{t+1} = s_t + \sigma_s\epsilon_2; \quad \epsilon_2 \sim \mathrm{Norm}(\epsilon_1, 0.1),$$
$$\alpha \sim \mathrm{Norm}(0, \sigma_\alpha^2),$$
$$[x_{t+1},\ y_{t+1}]^T = d_{t+1}\mathbf{R}_\alpha \begin{bmatrix} x_t - f^x_{t+1} \\ y_t - f^y_{t+1} \end{bmatrix} + \begin{bmatrix} f^x_{t+1} \\ f^y_{t+1} \end{bmatrix},$$
$$[w_{t+1},\ h_{t+1}] = s_{t+1}[w_t,\ h_t].$$



UCL · université de Caen Basse-Normandie · EADS

http://visual.cs.ucl.ac.uk/pubs/MotionModelPrediction