CVSSP Centre for Vision, Speech and Signal Processing

EPSRC

UNIVERSITY OF SURREY

Exploring Causal Relationships in Visual Object Tracking

Karel Lebeda, Simon Hadfield, Richard Bowden

{K.Lebeda, S.Hadfield, R.Bowden}@Surrey.ac.uk

ABSTRACT

What is the problem?

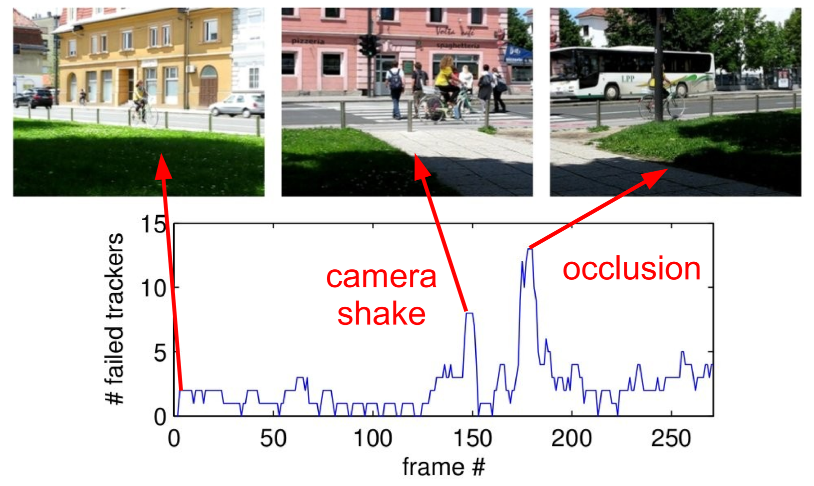
- In tracking, the **camera** and **object** motions are often **linked**.
- If the relationship between the camera motion and object motion can be **identified**, it could help **improve tracking**!

How do we solve it?

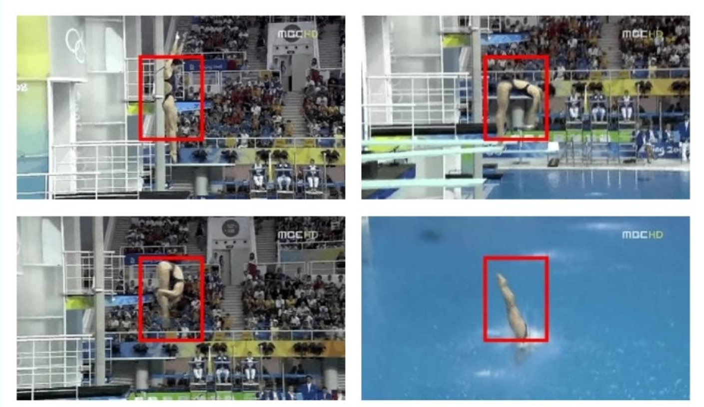
- We employ **transfer entropy** to **identify** and **model** the **relationship** between the camera and object.
- This can be used to **improve any tracking algorithm!**
- We demonstrate **62%** improvement in prediction. Tracking by **Struck** on the **VTB** dataset is improved by **7%** in accuracy and **22%** in robustness!

Where can it be useful?

- Common tracking failures due to camera motion or occlusion:



- In this sequence, where the cameraman follows the subject, tracking is reduced and a **fixed** bounding box provides **near-perfect** results:



Identifying and predicting these causal relationships automatically provides benefits to **any** tracking algorithm!

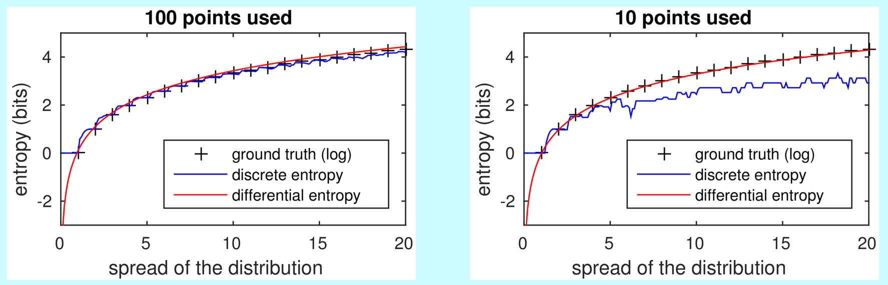
EXPLORING CAUSAL RELATIONSHIPS

How do we identify causal relationships?

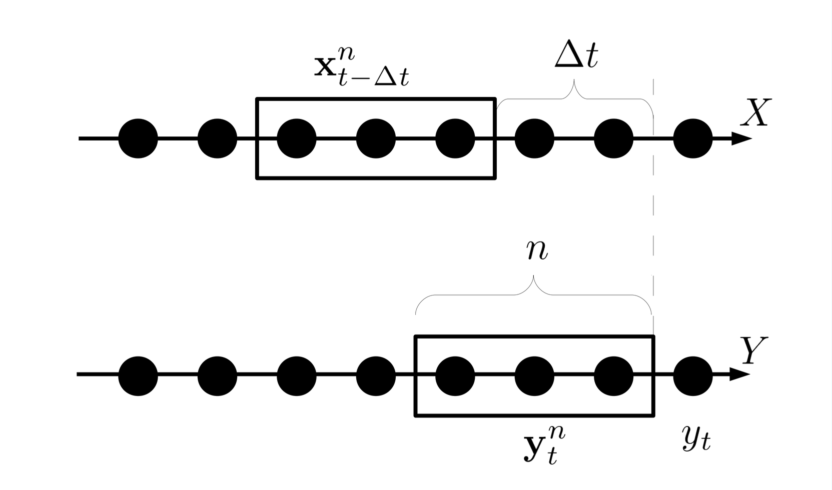
- Granger causality models only linear relationships!
- Our entropy-based method allows **complex non-linear** relationships to be modelled!

1. DIFFERENTIAL ENTROPY

$$H(X) = -\int_X p(x)\log p(x)\mathrm{d}x \quad \text{(approximated via KDE)}$$



2. TRANSFER ENTROPY



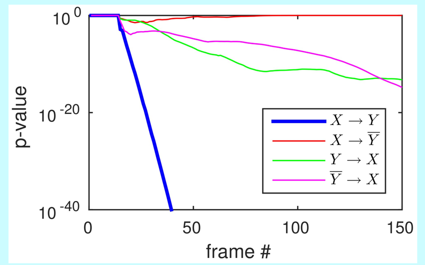
$$T_{X\to Y} = \iiint p(y_t, \mathbf{y}^n_t, \mathbf{x}^n_{t-\Delta t}) \log \frac{p(y_t|\mathbf{y}^n_t, \mathbf{x}^n_{t-\Delta t})}{p(y_t|\mathbf{y}^n_t)} \mathrm{d}y_t \mathrm{d}\mathbf{y}^n_t \mathrm{d}\mathbf{x}^n_{t-\Delta t}$$

$$T_{X\to Y} = \left(H(Y_t, \mathbf{Y}^n_t) - H(\mathbf{Y}^n_t)\right) - \left(H(Y_t, \mathbf{Y}^n_t, \mathbf{X}^n_{t-\Delta t}) - H(\mathbf{Y}^n_t, \mathbf{X}^n_{t-\Delta t})\right)$$

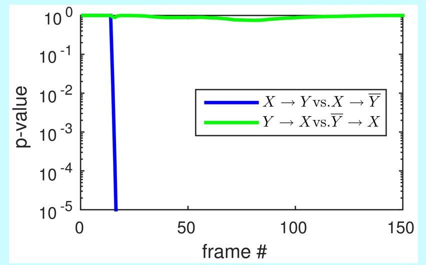
- If there is a **causal relationship** between X and Y, then **adding knowledge** about Y brings **more information** to a system which does not know X, than to one which does!

3. IDENTIFYING STATISTICALLY SIGNIFICANT RELATIONSHIPS

$\mathrm{E}(T_{X\to Y}) \leq 0$ vs. $\mathrm{E}(T_{X\to Y}) \leq \mathrm{E}(T_{X\to \tilde{Y}})$



vs.

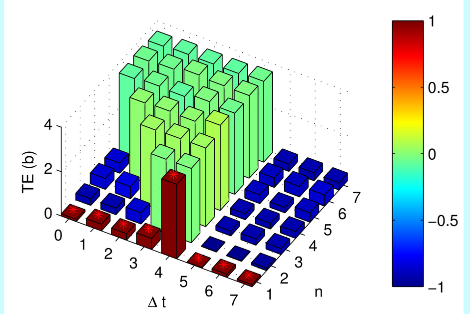


- Samples overlapping, not independent!
- Welch's test with shuffled data instead of simple t-test.

EXPLORING C.R.

4. OPTIMAL PARAMETERS

$$(\Delta t^*, n^*) = \underset{\Delta t, n}{\arg\max}\, T(\Delta t, n) \quad \text{s.t.} \quad f(\Delta t, n) > \theta_f$$



- Maximise TE (column height) such that relative improvement (column colour) is over some threshold.
- Provides **properties** of the causal relationship!

EXPLOITING CAUSAL RELATIONSHIPS

How do we use the causal relationships in tracking?

- Camera motion estimation is easier than object tracking.
- Using the identified causal relationships, we can **predict** the object pose in the next frame...

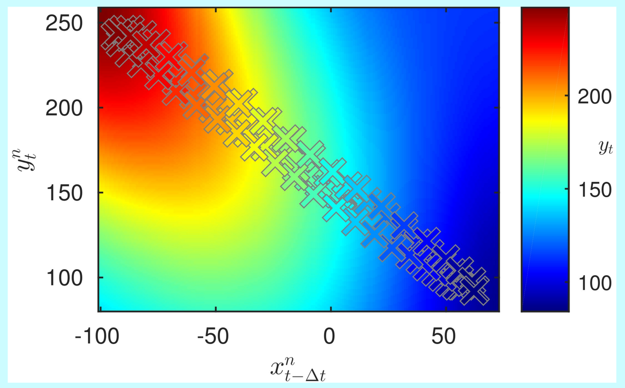
WINDOW-BASED CAUSAL PREDICTION

- Based on **autoregression** (prediction based on its own history)...

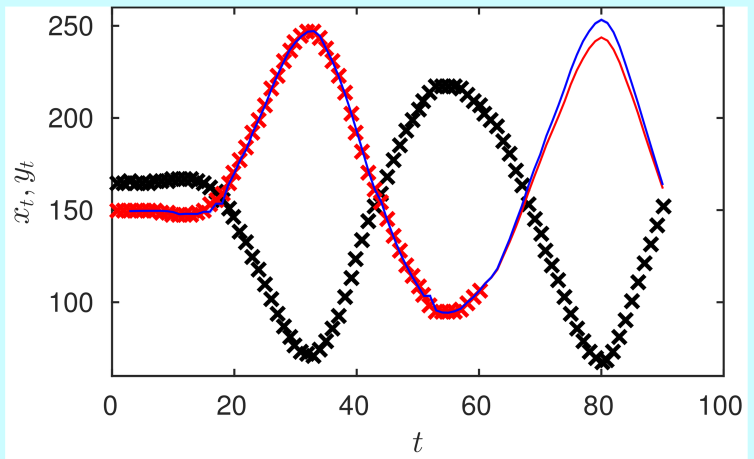
$$y_t = \phi_a(\mathbf{y}^n_t)$$

- The **causally-related** signal X can **aid** the prediction!

$$y_t = \phi_w(\mathbf{y}^{n^*}_t, \mathbf{x}^{n^*}_{t-\Delta t^*})$$



- Gaussian Process Regression with **concatenated features!**



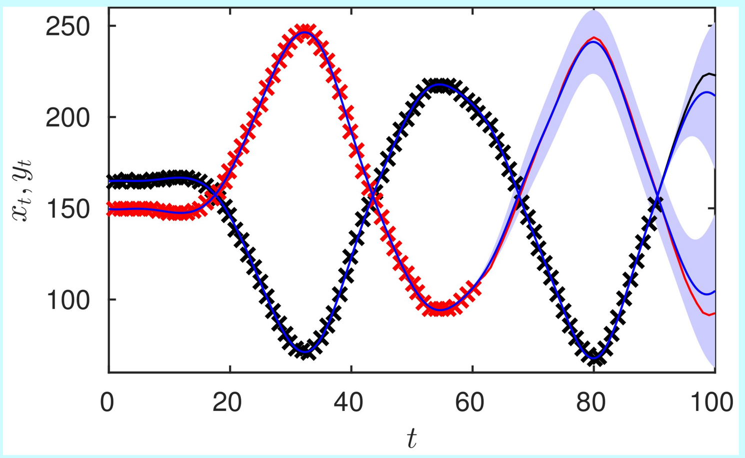
TIME-BASED CAUSAL PREDICTION

- Predicted variable as a **function of time**:

$$y_t = \phi_s(t \mid y_{1..t-1})$$

- **Encoded** causal relationship:

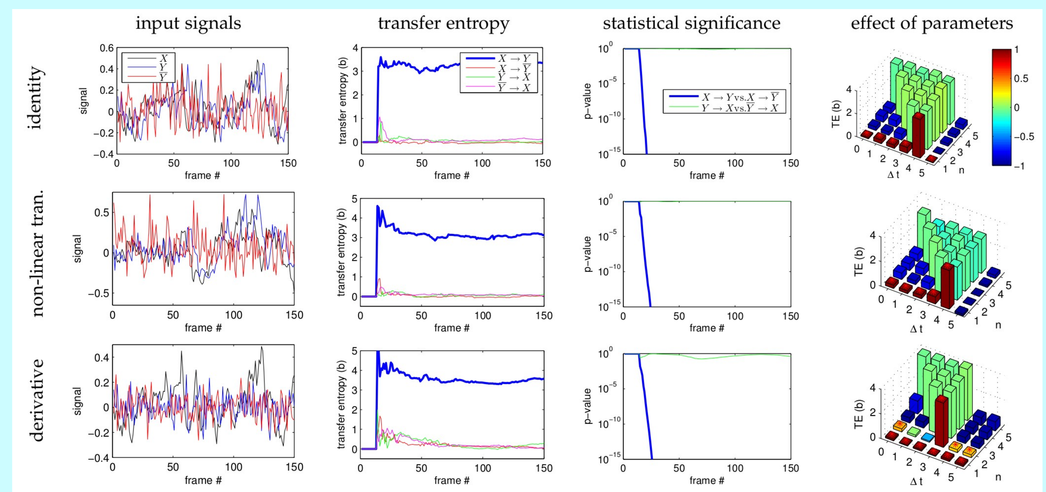
$$y_t = \phi_t(t \mid y_{1..t-1}, x_{1..t-\Delta t^*}, n^*)$$



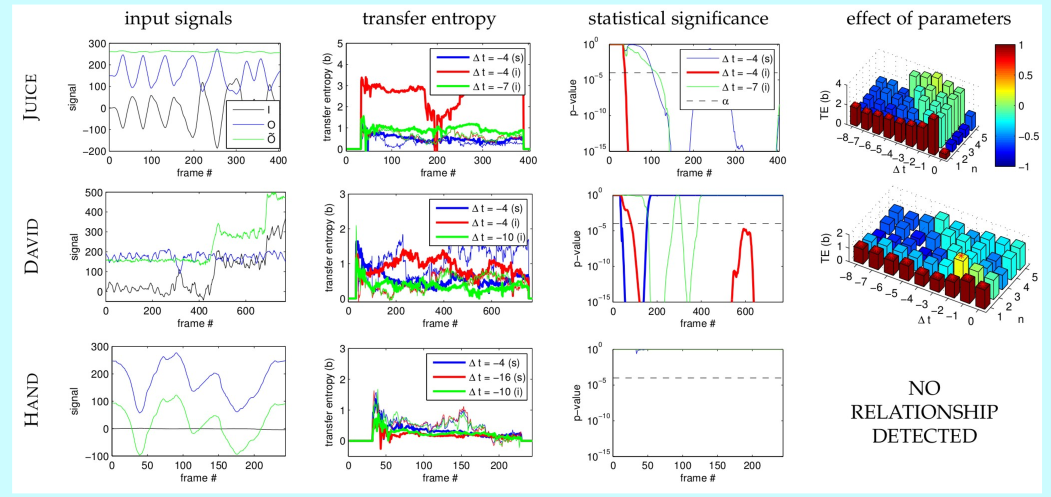
- Gaussian Process Regression with **coregionalisation**!

RESULTS: CAUSALITY

How well do we do?

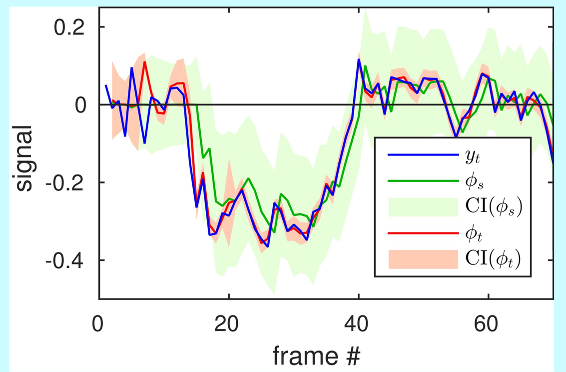


- Ground-truth parameters identified (Δt = 4, n = 1).

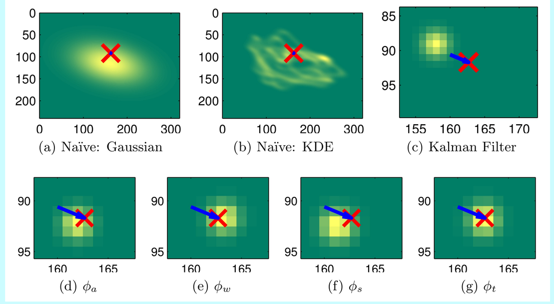


- No ground truth available, but intuitively correct!

RESULTS: PREDICTION



- Higher accuracy and confidence than autoregression when using causal prediction.



- Outperforming global predictors, Kalman filter, autoregression.

Category	Struck	Struck$_{\text{BMC}}$	Struck$_{\phi_t}$
BC	0.60/1.9	0.55/1.9	**0.61/1.7**
DEF	0.55/2.4	0.55/2.6	**0.60/2.2**
FM	0.53/3.2	0.51/3.3	**0.57/2.5**
IPR	0.53/2.6	0.50/3.0	0.55/**2.0**
IV	0.58/2.1	0.51/1.9	0.60/**1.6**
LR	0.51/1.4	0.48/1.1	0.56/**1.0**
MB	0.53/3.0	0.51/2.9	0.56/**2.0**
OCC	0.55/2.5	0.54/2.9	**0.59/2.0**
OPR	0.55/2.3	0.54/2.7	**0.59/1.9**
OV	0.55/3.0	0.58/2.7	0.55/**2.5**
SV	0.52/2.3	0.49/2.7	0.57/**1.9**
All	0.55/2.4	0.51/2.6	**0.59/1.9**

- VOT measures (accuracy/robustness) on the VTB50 dataset.
- Causal prediction aid performance, while simple background motion compensation fails.

CONCLUSION

What should you take away?

- We explore causal relationships between camera and object motion.
- These relationships (even complex non-linear) can be identified and measured.
- Camera motion estimation is more robust to errors than tracking → causal prediction.
- Causal prediction outperforms autoregression and Kalman filter (by over 60%)!
- SOTA tracker improved by 7&22% in benchmark.
- **Performance of ANY tracker can be improved!**

REFERENCES

[13] Granger. **Investigating causal relations by econometric models and cross-spectral methods**. *Econometrica*, 1969.

[23] Kaiser and Schreiber. **Information transfer in continuous processes**. *Physica D: NP*, 2002.

[25] Kristan et al. **The visual object tracking VOT2013 challenge results.** In *ICCV VOT*, 2013.

[27] Matas. **Visual tracking in the 21st century.** In *BMVC*, 2012.

[33] Schreiber. **Measuring information transfer.** *Physical Review Letters*, 2000.

This work was presented at the International Conference on Computer Vision, Santiago, Chile, on 13–16 December 2015. The authors were supported by the EPSRC grant EP/I011811/1 and the Rabin Ezra Scholarship.