Keeping Track: Online Learning for Robust Visual Tracking

Imperial College London

Graduate School Research Symposium 2012

Stephan Liwicki – Department of Computing

1. Introduction

The need for robust tracking is manifold. Its applications include, but are not limited to: traffic monitoring [1], human computer interaction [2], video retrieval [3], surveillance [4] and face analysis [5]. However, it remains unsolved, as many systems fail under varying pose and lighting, occlusions and appearance changes.

Our contributions are

- a seldom found combination of both efficiency *and* robustness,
- a measure which suppresses noise and outliers in image data,
- the formulation of an accurate incremental tracking algorithm,
- adaptive state-of-the-art tracking in realistic environments.

2. Unknown, Robust Feature Space

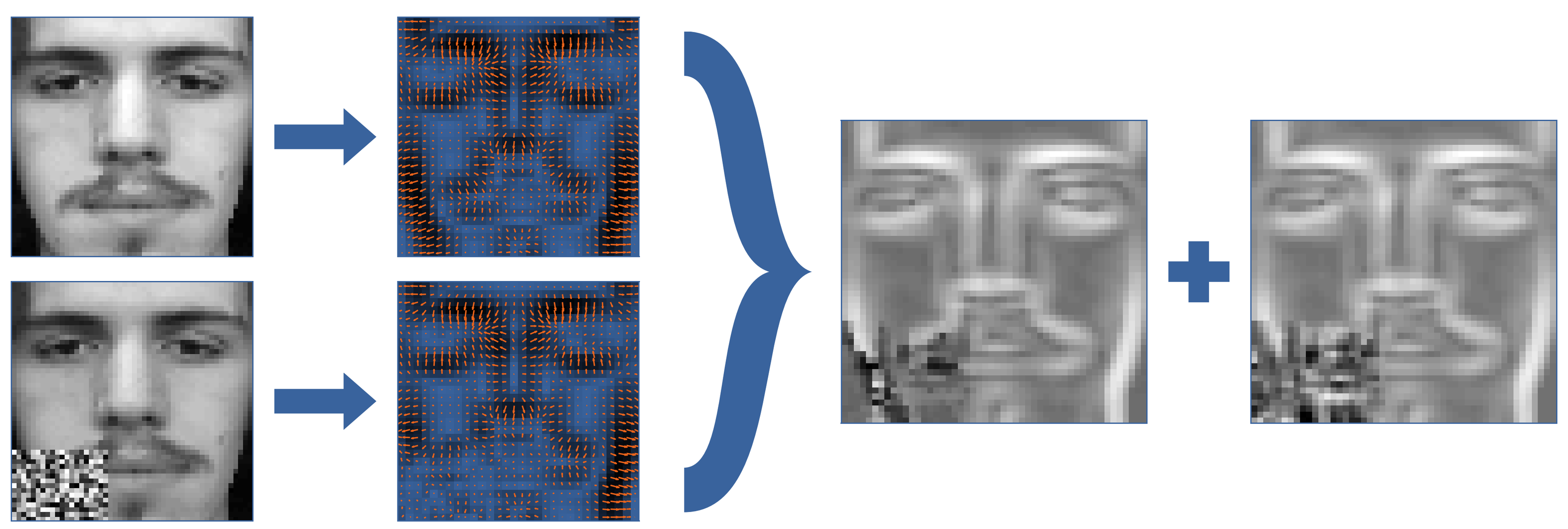

Figure 1: Visualisation of our kernel when corruption is present.

We utilise an unknown mapping, $\phi : \mathbb{C}^d \to \mathcal{K}$, which moves the vectorised input data into a non-linear representation, the so-called Krein space. A known kernel, $k : \mathbb{C}^d \times \mathbb{C}^d \to \mathbb{C}$, implicitly defines this mapping, as it is equivalent to the dot-product between two sample vectors in Krein space, *i.e.* $k(\mathbf{x}_p, \mathbf{x}_q) = \phi(\mathbf{x}_p)^H \phi(\mathbf{x}_q)$.

> Our robust edge gradient-based kernel, $k(\mathbf{x}_p, \mathbf{x}_q)$, is given by
>
> $$\frac{\sum_{c=1}^{d} \mathbf{R}_p(c)\,(\cos(\Delta\theta(c)) - j\sin(\Delta\theta(c)))}{2\sqrt{\sum_{c=1}^{d} \mathbf{R}_p^2(c)d}} + \frac{\sum_{c=1}^{d} \mathbf{R}_q(c)\,(\cos(\Delta\theta(c)) - j\sin(\Delta\theta(c)))}{2\sqrt{\sum_{c=1}^{d} \mathbf{R}_q^2(c)d}}$$
>
> with gradient magnitude $\mathbf{R}$, angle difference $\Delta\theta$, dimensionality d.
>
> **The robust properties of our kernel are derived from**
>
> - the edge gradient orientation features, rather than pixel intensity,
> - the split into two terms, in contrast to standard multiplication,
> - the cosine and sine on the difference of gradient orientations [6].

In the following, we introduce efficiency as we find and utilise two known mappings whose dot-product represents the kernel, *i.e.* $k(\mathbf{x}_p, \mathbf{x}_q) = a(\mathbf{x}_p)^H b(\mathbf{x}_q)$. Note, however, $a(\mathbf{x}_p) \neq b(\mathbf{x}_p)$.

3. Adaptive Tracking Algorithm

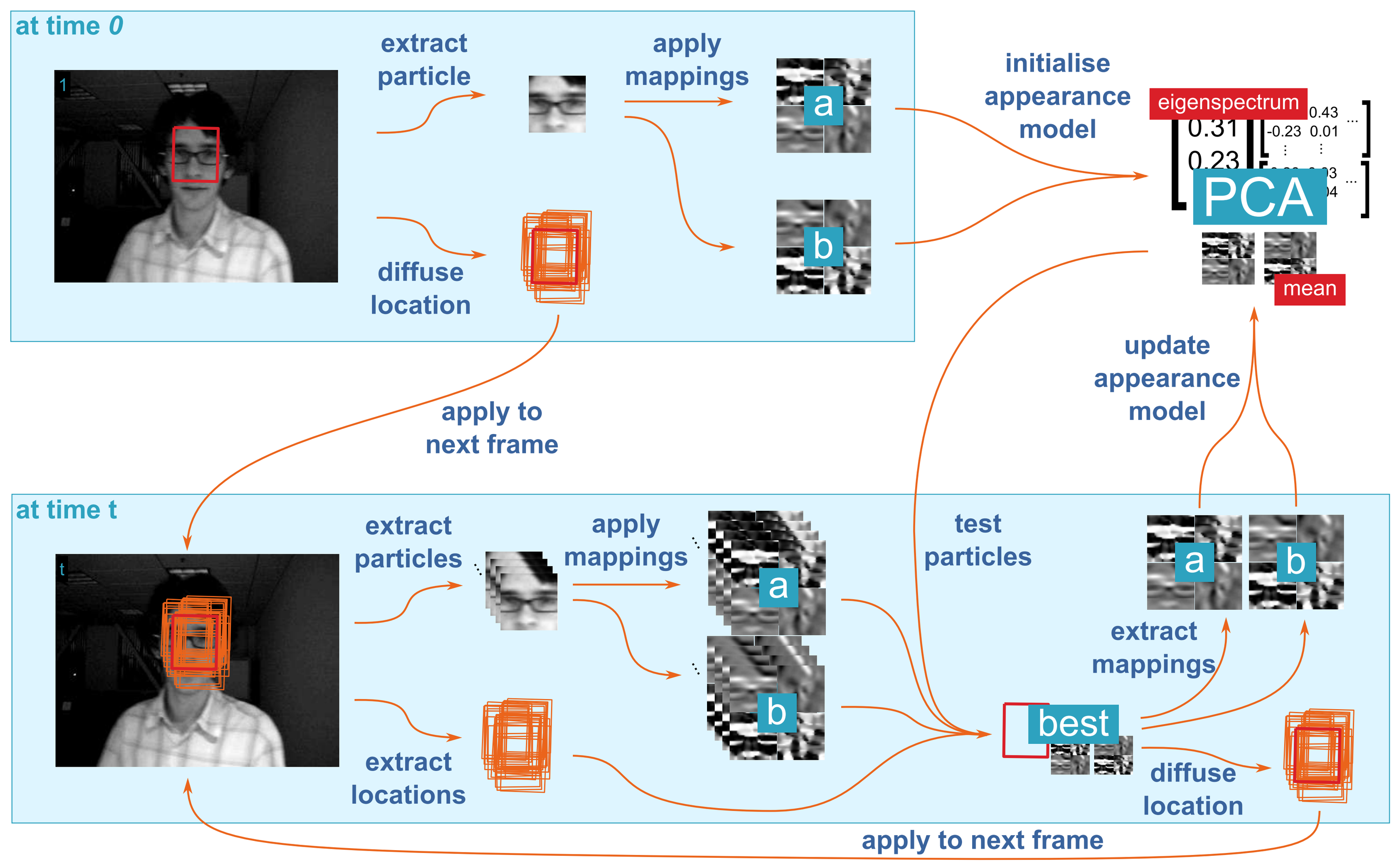

Figure 2: Our Direct Incremental Kernel-PCA Tracker (DIKT).

We propose a non-object-specific tracking algorithm which learns the appearance of the target online. All that is needed is the object's initial position in the first video frame.

> **At each new frame in time-step t, we**
>
> - extract particles around the best previous location at $t-1$,
> - find the sample that best matches our appearance model,
> - update the principal component analysis (PCA) with the sample.

Our algorithm combines efficiency with robustness, as we utilise our kernel which can be computed directly. This removes the need for hard to find pre-images, used in other kernel-based trackers.

4. Qualitative Evaluation

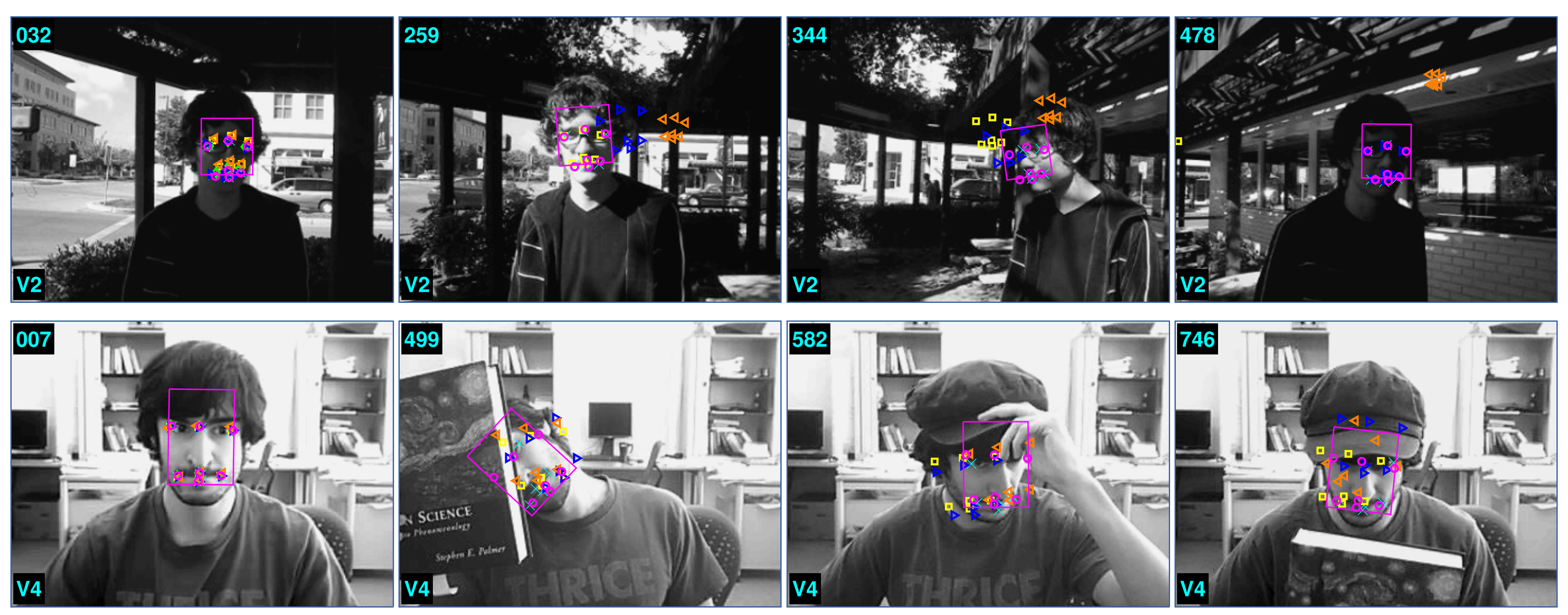

Figure 3: DIKT (as ○ with bounding box) versus 4 other recent trackers.

We compare our framework to 4 other recent trackers (IVT [7], MIL [8], L1 [9] and IKPCA [10]). A total of 9 popular and challenging videos are used. Fig. 3 shows a subset of the qualitative results.

5. Quantitative Evaluation

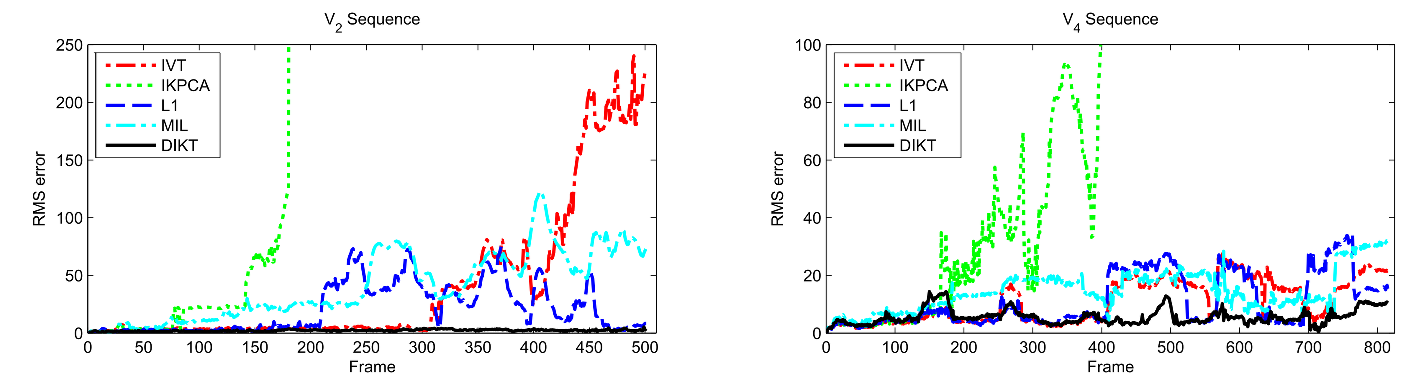

Figure 4: Comparison of the root mean square error (RMS) for 2 videos.

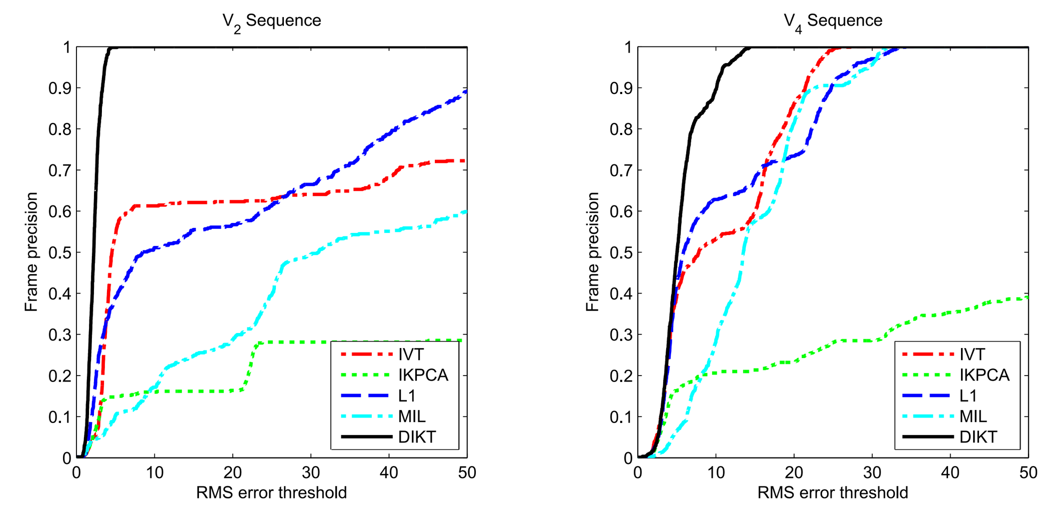

Figure 5: Comparison of the precision for 2 videos.

As before, we compare to the 4 trackers in 9 videos. Fig. 4 and Fig. 5 plots the accuracy and precision respectively. Table 1 shows the average root mean square error of all frames for each video.

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
IVT	6.82	(lost)	4.07	10.79	(lost)	**3.31**	**1.78**	2.62	(lost)
IKPCA	(lost)	(lost)	(lost)	(lost)	(lost)	(lost)	(lost)	(lost)	(lost)
L1	6.17	(lost)	2.87	11.10	12.68	9.53	1.62	13.58	(lost)
MIL	16.95	(lost)	13.61	14.62	37.56	12.73	4.14	23.87	17.62
DIKT	**4.48**	**2.27**	**2.49**	**5.62**	**11.28**	3.40	1.80	**1.96**	**5.90**

Table 1: Average root mean square error for all videos.

6. Conclusions

Our Direct Incremental Kernel-PCA Tracker

- outperforms in precision and accuracy for most videos,
- follows the target with occlusions, varying pose and lighting,
- adapts to appearance changes while suppressing noise,
- performs in real-time.

> **Related Journal Paper:**
> S. Liwicki, S. Zafeiriou, G. Tzimiropoulos, and M. Pantic, "Efficient Online Subspace Learning with an Indefinite Kernel for Visual Tracking and Recognition," IEEE Transaction on Neural Networks and Learning Systems, *in print*.

References

[1] J. Hsieh, S. Yu, Y. Chen, and W. Hu, "Automatic Traffic Surveillance System for Vehicle Tracking and Classification," *IEEE Trans. Intell. Transportation Sys.*, pp. 175 – 187, 2006.

[2] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 780 – 785, 1997.

[3] Y. Luo, T. Wu, and J. Hwang, "Object-based analysis and interpretation of human motion in sports video sequences by dynamic bayesian networks," *Computer Vision and Image Understanding*, pp. 196 – 216, 2003.

[4] R. Collins, J. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for Cooperative Multisensor Surveillance," in *The IEEE*, 2001, p. 89.

[5] H. Gunes and M. Pantic, "Automatic, Dimensional and Continuous Emotion Recognition," *Int. Journal of Synthetic Emotion*, pp. 68 – 99, 2010.

[6] G. Tzimiropoulos, V. Argyriou, S. Zafeiriou, and T. Stathaki, "Robust FFT-Based Scale-Invariant Image Registration with Image Gradients," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1899 – 1906, 2010.

[7] D. Ross, J. Lim, R. Lin, and M. Yang, "Incremental Learning for Robust Visual Tracking," *Int. Journal of Comp. Vision*, pp. 125 – 141, 2008.

[8] B. Babenko, M. Yang, and S. Belongie, "Robust Object Tracking with Online Multiple Instance Learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1619 – 1632, 2011.

[9] X. Mei and H. Ling, "Robust Visual Tracking using L1 Minimization," in *ICCV'09*, 2009, pp. 1436 – 1443.

[10] T. Chin and D. Suter, "Incremental Kernel Principal Component Analysis," *IEEE Trans. Image Process.*, pp. 1662 – 1674, 2007.