Multi-Channel Correlation Filters

Hamed Kiani[1], Terence Sim[1] and Simon Lucey[2]
{hkiani,tsim}@comp.nus.edu.sg, simon.lucey@csiro.au
[1]School of Computing, NUS, Singapore
[2]CSIRO ICT Center, Australia

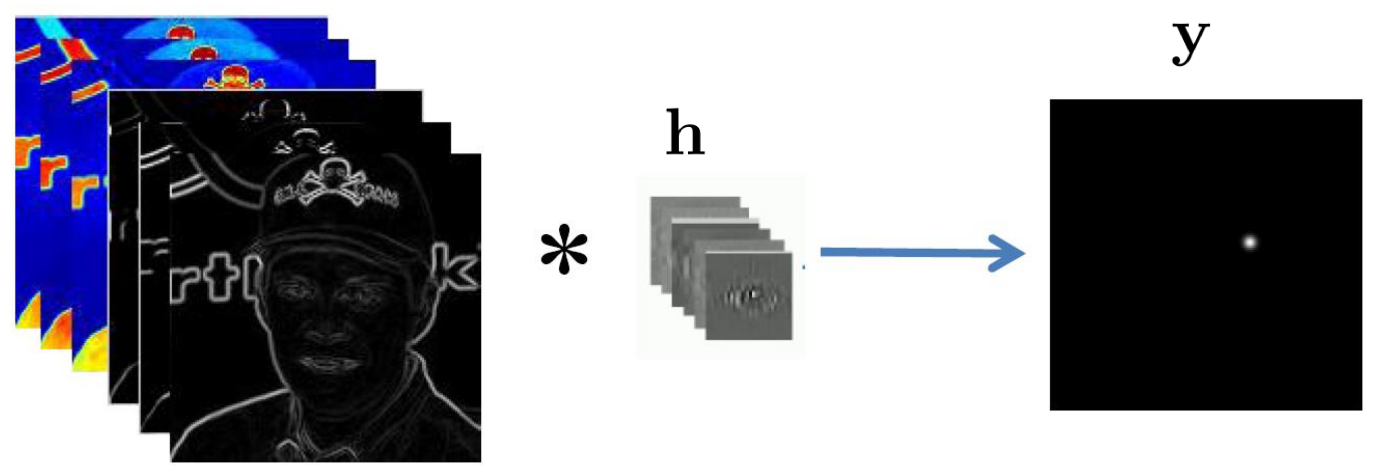
Abstract

From a signal processing perspective, pattern detection using modern descriptors like HOG can be efficiently posed as a correlation between a multi-channel image and a multi-channel detector/filter, which results in a single-channel response indicating where the pattern (e.g. object) has occurred. Here, we proposed a novel framework for learning multi-channel filters efficiently in the frequency domain, both in terms of complexity and memory usage.

Contributions

- Extending canonical correlation filter theory to efficiently handle multi-channel signals
- A multi-channel detector whose training memory is independent of the number of training samples
- Superior performance to current state of the art correlation filters, and superior computational and memory efficiency in comparison to spatial detectors (e.g. linear SVM) with comparable detection performance

Multi-Channel CFs



(i) Spatial domain:

$$\arg\min_{\mathbf{h}} \left\| \mathbf{y} - \sum_{k=1}^{K} \mathbf{h}^{(k)} * \mathbf{x}^{(k)} \right\|_2^2 + \lambda \sum_{k=1}^{K} \left\| \mathbf{h}^{(k)} \right\|_2^2$$

(ii) Fourier domain:

$$\arg\min_{\hat{\mathbf{h}}} \Big\| \underbrace{\hat{\mathbf{y}}}_{D \times 1} - \mathrm{conj}\Big(\underbrace{\begin{bmatrix} \mathrm{diag}(\hat{\mathbf{x}}^{(1)}) & \cdots & \mathrm{diag}(\hat{\mathbf{x}}^{(K)}) \end{bmatrix}}_{D \times KD} \Big) \underbrace{\begin{bmatrix} \hat{\mathbf{h}}^{(1)} \\ \vdots \\ \hat{\mathbf{h}}^{(K)} \end{bmatrix}}_{KD \times 1} \Big\|_2^2 + \lambda \left\| \begin{bmatrix} \hat{\mathbf{h}}^{(1)} \\ \vdots \\ \hat{\mathbf{h}}^{(K)} \end{bmatrix} \right\|_2^2$$

Complexity: $\mathcal{O}(D^3K^3)$

Memory: $\mathcal{O}(D^2K^2)$

(iii) Fourier domain with variable re-ordering:

$$\arg\min_{\nu(\hat{\mathbf{h}}(j))} \Big\| \underbrace{\hat{\mathbf{y}}(j)}_{\text{scalar}} - \underbrace{\mathrm{conj}(\nu(\hat{\mathbf{x}}(j)))}_{1 \times K} \underbrace{\nu(\hat{\mathbf{h}}(j))}_{K \times 1} \Big\|_2^2 + \lambda \left\| \nu(\hat{\mathbf{h}}(j)) \right\|_2^2 \quad \text{for } j = 1, \ldots, D.$$

Complexity: $\mathcal{O}(DK^3)$

Memory: $\mathcal{O}(DK^2)$

NOTATION: $*$: convolution operation, $|\mathbf{y}| = D$, K:# of channels and $\mathcal{V}(\mathbf{a}(j)) = [\mathbf{a}^{(1)}(j), \ldots, \mathbf{a}^{(K)}(j)]$

Comparison with Linear SVM

	250	500	1000	2000	4000	8000	16000	24000
MCCF	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
SVM	6.17	12.35	24.68	49.36	98.87	197.44	395.88	592.32

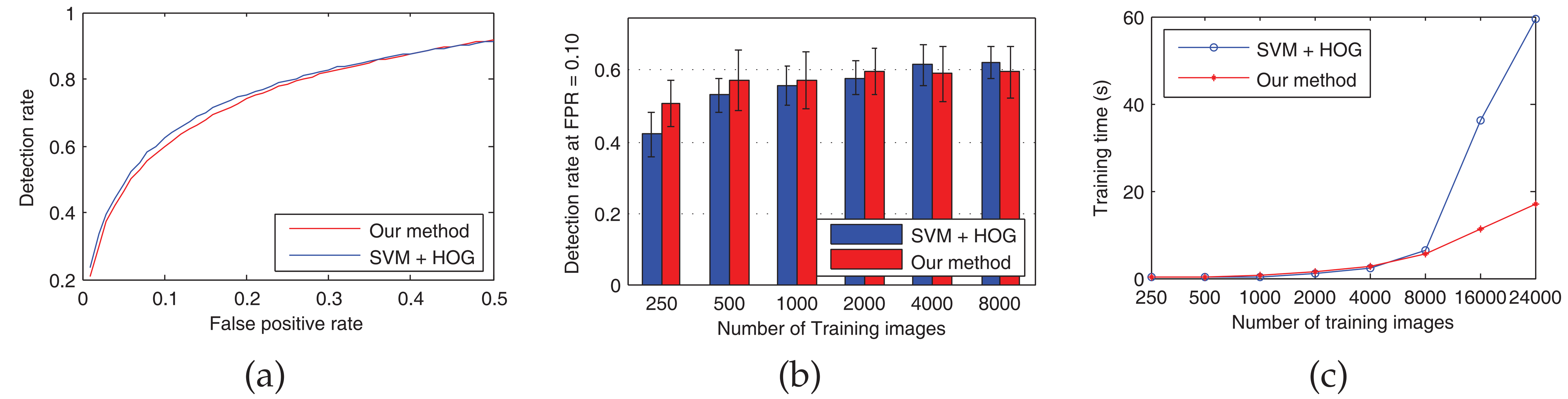


Figure 1. Comparing MCCF with SVM + HOG on the problem of pedestrian detection using Daimler dataset. Top: Memory usage (MB) of MCCF compared to SVM as a function of number of training images. Bottom: Detection rate as a function of (a) FPR, (b) number of training images at FPR = 0.10, and (c) training time versus training size.

Facial landmark detection

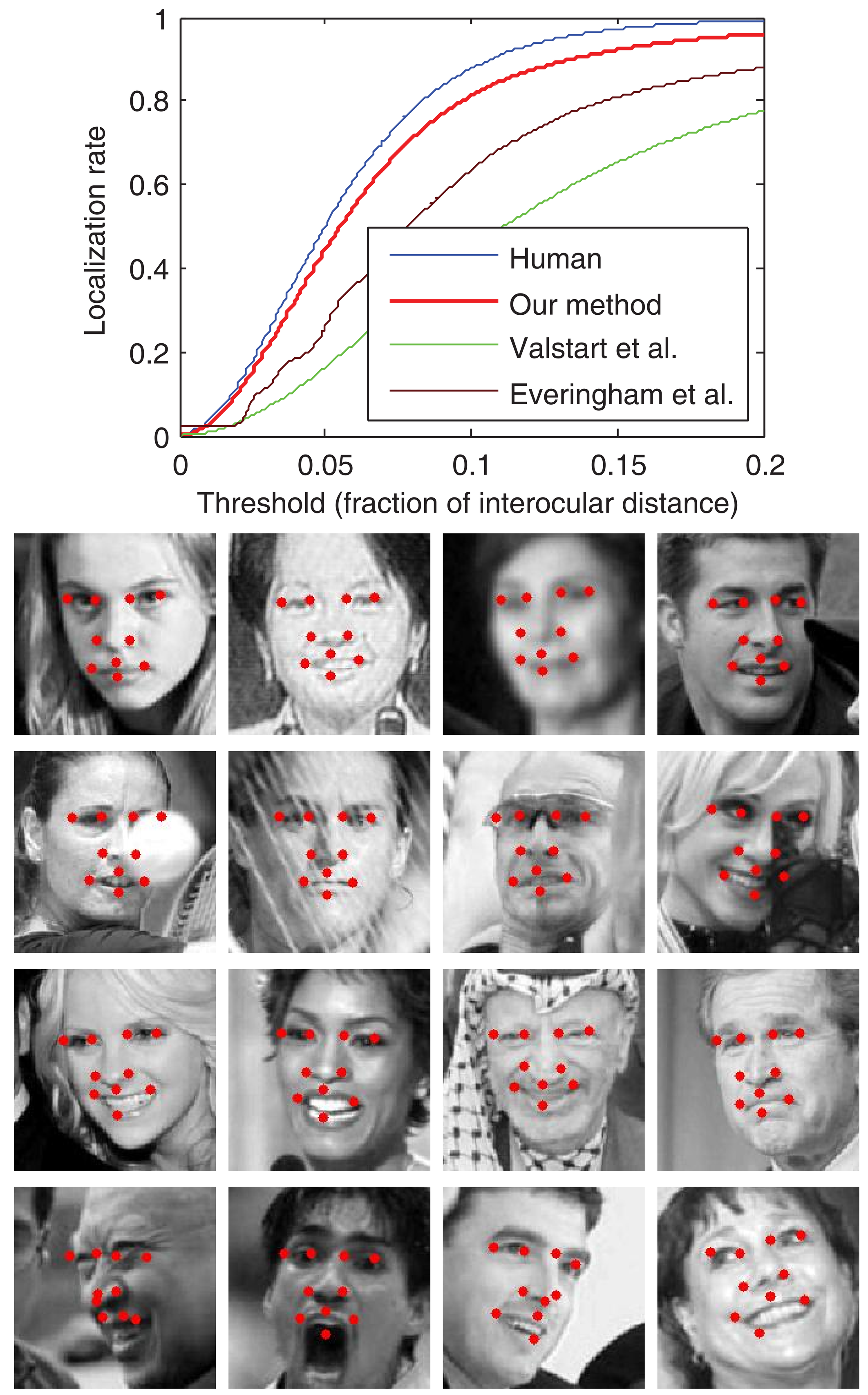


Figure 2. Facial landmark detection on the LFW dataset.

Car Detection

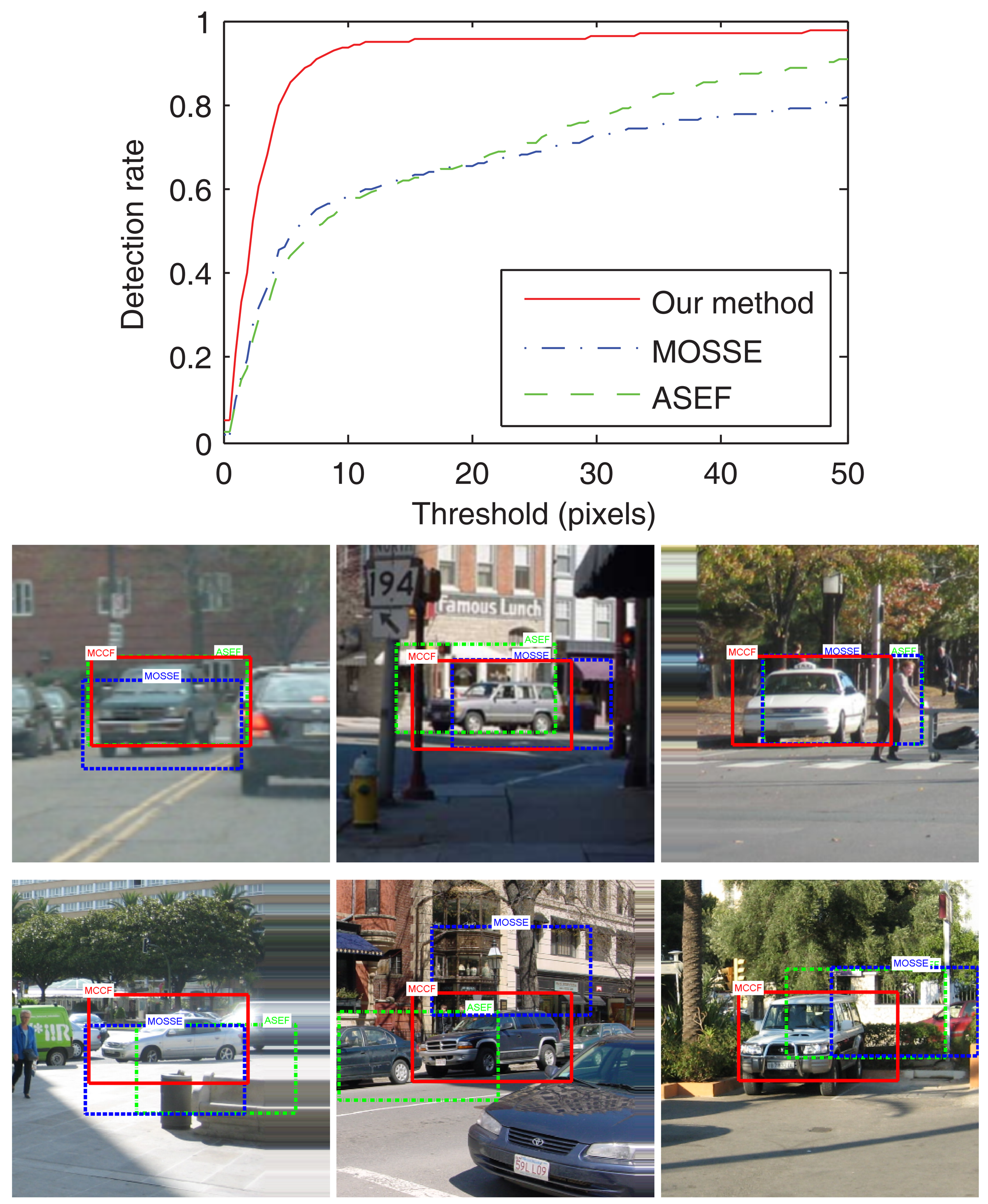


Figure 3. Car detection on the MIT Street Dataset.

References

[1] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui. Visual object tracking using adaptive correlation filters. In *CVPR '10*.