ECCV 2012

UNIVERSIDADE DE COIMBRA
UNIVERSITY OF COIMBRA

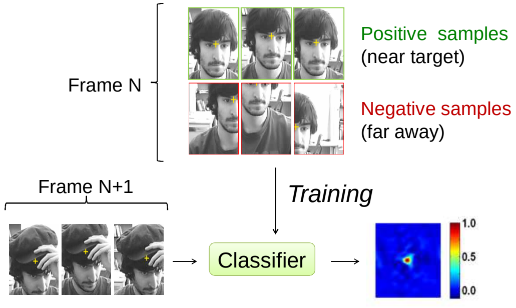
EXPLOITING THE CIRCULANT STRUCTURE OF TRACKING-BY-DETECTION WITH KERNELS

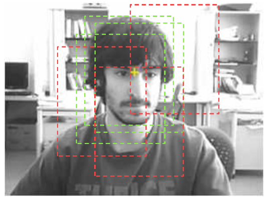
João F. Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista

1. Motivation

- In tracking-by-detection, a classifier is trained with **several patches** from a **single image**. ⇒

The data matrix has **extreme redundancies**.





This is a hint that **classic algorithms do unnecessary work.**

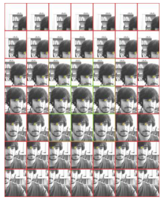
2. Idea

- Instead of random patches, use **all patches**.
- Look for a pattern that we can take advantage of.

⇒ It turns out that in this case, The data matrix is **Circulant.**

$$C(\mathbf{u}) = \begin{bmatrix} u_0 & u_1 & u_2 & \cdots & u_{n-1} \\ u_{n-1} & u_0 & u_1 & \cdots & u_{n-2} \\ u_{n-2} & u_{n-1} & u_0 & \cdots & u_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_1 & u_2 & u_3 & \cdots & u_0 \end{bmatrix}$$

- Image
- Image shifted by 1 element
- Image shifted by 2 elements
- ⋮
- Image shifted by n-1 elements



3. Circulant matrices

- Easy to manipulate algebraically.
- Allow **standard learning methods** to be transformed into efficient **Fourier domain** operations.

Circulant Matrix **Standard algebra operations**	⇔ Fourier Domain **Cheap element-wise operations**
$X = C(\mathbf{x}),\ Y = C(\mathbf{y}),\ Z = C(\mathbf{z})$	$\hat{\mathbf{x}} = \mathcal{F}(\mathbf{x}),\ \hat{\mathbf{y}} = \mathcal{F}(\mathbf{y}),\ \hat{\mathbf{z}} = \mathcal{F}(\mathbf{z})$
$Z = X + Y$	$\hat{\mathbf{z}} = \hat{\mathbf{x}} + \hat{\mathbf{y}}$
$Z = XY$	$\hat{\mathbf{z}} = \hat{\mathbf{x}}^* \odot \hat{\mathbf{y}}$
$\mathbf{z} = X\mathbf{y}$	$\hat{\mathbf{z}} = \hat{\mathbf{x}}^* \odot \hat{\mathbf{y}}$
$Z = X^{-1}$	$\hat{\mathbf{z}} = 1/\hat{\mathbf{x}}$

4. Linear classification

Ridge Regression + Circulant data ⇒ MOSSE filter (Bolme et al., CVPR 2010)

$$\mathbf{w} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathbf{x}) \odot \mathcal{F}^*(\mathbf{y})}{\mathcal{F}(\mathbf{x}) \odot \mathcal{F}^*(\mathbf{x}) + \lambda}\right)$$

- **Independent proof** from a risk minimization point-of-view.
- Circulant matrices **link** two fields:
  - **Generic learning algorithms**, and
  - **Classic signal processing** (frequency-domain filter synthesis)

5. Kernel classification

- **Kernel matrix** is Circulant for unitary kernels (Thm. 1).
- Compact and exact representation $\mathbf{k}$ — $\underbrace{K}_{n^2 \times n^2} = C\,(\underbrace{\mathbf{k}}_{n^2 \times 1})$ ⇒ The full kernel matrix is never built explicitly!

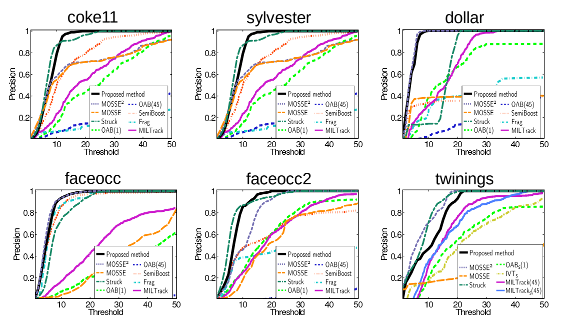
- **Solutions for**:
  - Kernel Ridge Regression — $\boldsymbol{\alpha} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathbf{y})}{\mathcal{F}(\mathbf{k}) + \lambda}\right)$
  - Classification — $\hat{\mathbf{y}} = \mathcal{F}^{-1}\left(\mathcal{F}(\bar{\mathbf{k}}) \odot \mathcal{F}(\boldsymbol{\alpha})\right)$
  - Computation of kernels — $\mathbf{k}^{\text{rbf}} = h\left(\|\mathbf{x}\|^2 + \|\mathbf{x}'\|^2 - 2\mathcal{F}^{-1}\left(\mathcal{F}(\mathbf{x}) \odot \mathcal{F}^*(\mathbf{x}')\right)\right)$
    - Linear
    - Polynomial
    - Gaussian ...

→ For **all patches** simultaneously

- **Kernel algorithm with same complexity as linear.**

Fast Fourier Transform ⇒ Tracking with kernels at 100-400 FPS

Results



Source code

Training image **x** (current frame) and test image **z** (next frame) must be pre-processed with a cosine window. **y** has a Gaussian shape centered on the target. **x**, **y** and **z** are M-by-N matrices. All FFT operations are standard in MATLAB.

```
function alphaf = training(x, y, sigma, lambda)    % Eq. 7
  k = dgk(x, x, sigma);
  alphaf = fft2(y) ./ (fft2(k) + lambda);
end

function yhat = detection(alphaf, x, z, sigma)     % Eq. 9
  k = dgk(x, z, sigma);
  yhat = real(ifft2(alphaf .* fft2(k)));
end

function k = dgk(x1, x2, sigma)                    % Eq. 16
  c = fftshift(ifft2(fft2(x1) .* conj(fft2(x2))));
  d = x1(:)'*x1(:) + x2(:)'*x2(:) - 2*c;
  k = exp(-1 / sigma^2 * abs(d) / numel(x1));
end
```

INSTITUTO DE SISTEMAS E ROBÓTICA – UNIVERSIDADE DE COIMBRA
FCTUC FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE DE COIMBRA
UNIÃO EUROPEIA Fundo Social Europeu

This work was supported by the Portuguese Science and Technology Foundation (FCT) through the project PTDC/EEA-CRO/122812/2010, and grants SFRH/BD75459/2010, SFRH/BD74152/2010, and SFRH/BD45178/2008.