DSST & FDSST

1.MOSSE

- 两个信号的相关性表示：

$$\text{连续}(f \otimes g)(\tau) = \int_{\infty}^{-\infty} f^*(t)g(t+\tau)dt$$
$$\text{离散}(f \otimes g)(n) = \sum_{-\infty}^{\infty} f^*[m]g(m+n)$$

f表示输入的图像特征，h表示滤波器模板，g表示响应输出。

- FFT

$$F(g) = F(f \otimes g) = F(f) \cdot F(h)^*$$
$$G = F \cdot H^*$$
$$H^* = \frac{G}{F}$$

- 构建目标函数

$$\min_{H^*} = \sum_{i=1}^{m} |H^* F_i - G_i|^2$$

上式的操作都是元素级别的，因此想要得到H，只要使其中的每个元素的MOSSE最小即可。上式转化为：

$$\min_{H^*_{wv}} \sum_{i=1}^{m} |H^*_{wv} F_{wvi} - G_{wvi}|^2$$

- 求偏导，使其为0，得到结果：

$$H_{wv} = \frac{\sum_i F_{iwv} G^*_{iwv}}{\sum_i F_{iwv} F^*_{iwv}}$$

- 最终得到H

$$H = \frac{\sum_i F_i \cdot G^*_i}{\sum_i F_i \cdot F^*_i}$$

- F_i与G_i的获取

对跟踪框进行随机仿射变换，获得一系列的训练样本f_i，而g_i则是由高斯函数产生，并且其峰值位置是在f_i的中心位置。

- 更新策略

$$H_t = \frac{A_t}{B_t}$$
$$A_t = \eta F_t \cdot G^*_t + (1-\eta) A_{t-1}$$
$$B_t = \eta F_t \cdot F^*_t + (1-\eta) B_{t-1}$$

2. DCF for Multidimensional Features

- 考虑到图像的特征是多维度的，设f为特征，有d维，f^l为其中的第l维，损失函数变为：

$$\varepsilon = \left\| \sum_{l=1}^{d} h^l \star f^l - g \right\|^2 + \lambda \sum_{l=1}^{d} \left\| h^l \right\|^2$$

- H^l

$$H^l = \frac{\bar{G} F^l}{\sum_{k=1}^{d} \overline{F^k} F^k + \lambda}$$

- 更新策略

$$A_t^l = (1-\eta) A_{t-1}^l + \eta \bar{G}_t F_t^l$$
$$B_t = (1-\eta) B_{t-1} + \eta \sum_{k=1}^{d} \overline{F_t^k} F_t^k$$

- 计算新一帧图像的响应

$$y = \mathscr{F}^{-1} \left\{ \frac{\sum_{l=1}^{d} \bar{A}^l Z^l}{B + \lambda} \right\}$$

- y的最大值被认为是目标新位置的估计

3.Fast Scale Space Tracking

- 将位置滤波器与尺度滤波器分开计算，利用二维的位置滤波器获得目标的中心位置后，再用一维的尺度滤波器估计目标在图片中的尺度。
- 训练样本f从目标中心扣取，假设当前帧的目标大小为$P \times R$，尺度为S，我们扣取目标中心的大小为$a^n P \times a^n R$的窗口标记为J^n。a表示尺度因子，取$a = 1.02$，$S = 33$，n的取值范围如下：

$$n \in \left\{ \left\lfloor -\frac{S-1}{2} \right\rfloor, \cdots, \left\lfloor \frac{S-1}{2} \right\rfloor \right\}$$

尺度等级为n的训练样本 f 是J^n的d维的特征描述子。

在连续的两帧中，位置的变化往往大于尺度的变化，因此，文中先采用位置滤波器确定位置信息，在位置的基础上再使用尺度滤波器确定尺度信息。

- 算法流程

Algorithm 1 Proposed tracking approach: iteration at time step t.

Input:

Image I_t.
Previous target position $\boldsymbol{p}_{t-1}$ and scale s_{t-1}.
Translation model A_{t-1}^{trans}, B_{t-1}^{trans} and scale model A_{t-1}^{scale}, B_{t-1}^{scale}.

Output:

Estimated target position $\boldsymbol{p}_t$ and scale s_t.
Updated translation model A_t^{trans}, B_t^{trans} and scale model A_t^{scale}, B_t^{scale}.

Translation estimation:

1: Extract a translation sample z_{trans} from I_t at $\boldsymbol{p}_{t-1}$ and s_{t-1}.
2: Compute the translation correlation y_{trans} using z_{trans}, A_{t-1}^{trans} and B_{t-1}^{trans} in (6).
3: Set $\boldsymbol{p}_t$ to the target position that maximizes y_{trans}.

Scale estimation:

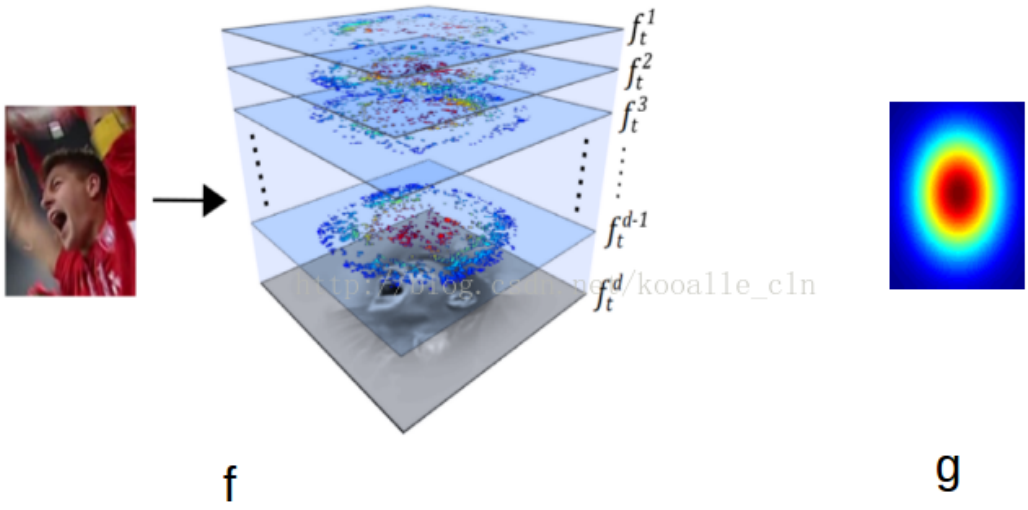
4: Extract a scale sample z_{scale} from I_t at $\boldsymbol{p}_t$ and s_{t-1}.
5: Compute the scale correlation y_{scale} using z_{scale}, A_{t-1}^{scale} and B_{t-1}^{scale} in (6).
6: Set s_t to the target scale that maximizes y_{scale}.

Model update:

7: Extract samples f_{trans} and f_{scale} from I_t at $\boldsymbol{p}_t$ and s_t.
8: Update the translation model A_t^{trans}, B_t^{trans} using (5).
9: Update the scale model A_t^{scale}, B_t^{scale} using (5).

- 位置估计的训练

目标所在的图像块P的大小为$M \times N$，提取P的特征（例如Fhog特征），得到大小为 $M \times N \times d$ 的特征 f ，其中特征的维度为 d维。如下图所示（作者选取灰度图为第1维的特征，后续的特征是Fhog的前27维特征），响应g是根据高斯函数构造的，大小为$M \times N$，中间响应值最大，向四周依次递减：



根据下述公式，需要对 f 的每一个维度的特征做二维的DFT，得到 F^l，对 g 做二维的DFT得到 G：
（下式中的乘法都是点乘，矩阵的对应位置相乘）

$$H^l = \frac{\bar{G}F^l}{\sum_{k=1}^{d} \overline{F^k}F^k + \lambda}$$

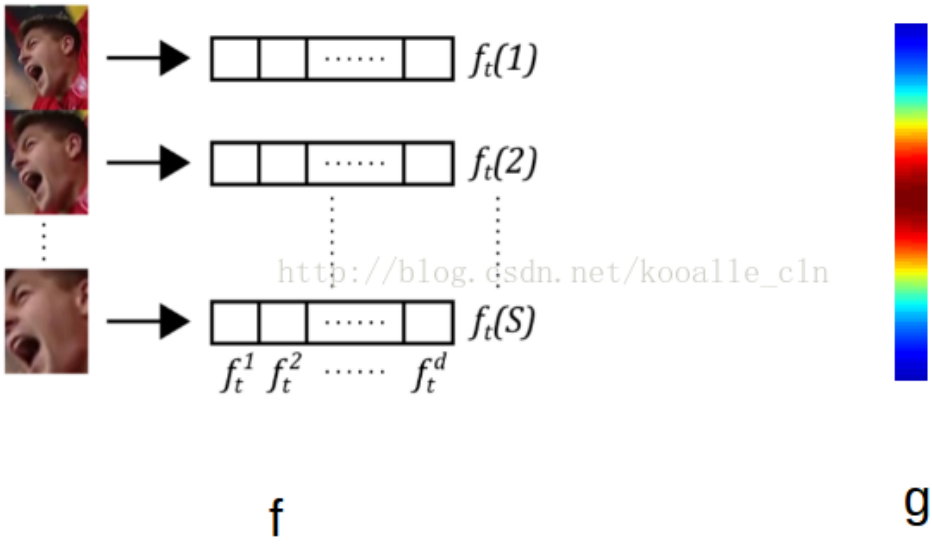
- 位置估计的检测

新的一帧的图片的特征为z，同样求取每一维度特征的DFT得到Z^l,按下式求出响应y。

$$y = \mathscr{F}^{-1}\left\{\frac{\sum_{l=1}^{d}\bar{A}^l Z^l}{B+\lambda}\right\}$$

- 尺度估计的训练

目标所在的图像块P的大小为$M \times N$，以图像块的正中间为中心，截取不同尺度的图片。这样就能够得到一系列的不同尺度的图像Patch（搜索范围为S个尺度，就会有S张图像Patch），针对每个图像Patch求其特征描述子（维度为 d 维，这里的 d和位置估计中的维度d没有任何关系，只与截取图像resize的尺寸有关），每一个维度的特征f^l是一个 $1 \times S$的向量，g是高斯函数构造的输出响应大小为 $1 \times S$，中间值最大，向两端依次减小 。
如上一样得到滤波器模板H



- 尺度估计的检测

新的一帧的图片以位置估计得出的位置为中心，截取S个不同尺度的图像Patch，分别求其特征描述子，组成新的特征为z，同样求取每一维度的一维的DFT得到Z^l， 进而得到y的值（求取方法和位置估计类似），y为 $1 \times S$ 维的向量，y 中最大值的所对应的尺度为最终尺度估计的结果。

- 总结

1. Input： 前一帧的目标位置P_{t-1}和尺度S_{t-1}，位置滤波器参数A_{t-1}^{trans} 、B_{t-1}^{trans}和尺度滤波器参数 A_{t-1}^{scale} 、B_{t-1}^{scale}
2. Output：估计的目标位置和尺度 P_t、S_t
3. 更新：利用新一帧的位置和尺度，更新滤波器参数。

4. Fast Discriminative Scale Space Tracking

4.1 Sub-Grid Interpolation of Correlation Scores

4.2 Dimensionality Reduction

为了减少傅里叶变换的计算量，更新一个目标模板：

$$u_t = (1-\eta)u_{t-1} + \eta f_t$$

则滤波器的分子A_t^l可以通过$\bar{G}\mathscr{F}\left\{u_t^l\right\}$得到。利用模板 u_t构造了一个投影矩阵P_t （$\tilde{d} \times d$）。这个矩阵定义了特征被投影到上面的低维子空间，$\tilde{d}$为特征压缩后的维数。

- 通过最小化目标模板u_t的重构误差来获得P_t

$$\varepsilon = \sum_n \left\| u_t(\boldsymbol{n}) - P_t^{\mathrm{T}} P_t u_t(n) \right\|^2$$

$$s.t.\, P_t P_t^T = I$$

- 通过对自相关矩阵进行特征值分解，得到一个解：

$$C_t = \sum_{\boldsymbol{n}} u_t(\boldsymbol{n}) u_t(\boldsymbol{n})^{\mathrm{T}}$$

- C_t的前$\tilde{d}$个最大的特征值对应的特征向量构成了P_t的行向量
- 压缩维度后的训练样本：

$$\tilde{F}_t = \mathscr{F}\left\{P_t f_t\right\}$$

- 压缩目标模板：

$$\tilde{U}_t = \mathscr{F}\left\{P_t u_t\right\}$$

- FDSST更新策略：

$$\tilde{A}_t^l = \bar{G}\tilde{U}_t^l, \quad l = 1, \ldots, \tilde{d}$$

$$\tilde{B}_t = (1-\eta)\tilde{B}_{t-1} + \eta \sum_{k=1}^{\tilde{d}} \tilde{\tilde{F}}_t^k \tilde{F}_t^k$$

- 检测：

$$\tilde{Z}_t = \mathscr{F}\left\{P_{t-1} z_t\right\}$$

$$Y_t = \frac{\sum_{l=1}^{\tilde{d}} \overline{\tilde{A}_{t-1}^l} \tilde{Z}_t^l}{\tilde{B}_{t-1} + \lambda}$$