

# Оценивание параметров в задаче слежения за множеством объектов в видеопотоке

Григорьев Алексей Дмитриевич

Московский физико-технический институт  
Физтех-школа прикладной математики и информатики  
Кафедра интеллектуальных систем

Научный руководитель: к.ф.-м.н. А.Н. Гнеушев

Москва, 2020

## Задача

Разработать модель сопровождения объектов по видео последовательности кадров, основанную на ре-идентификации объектов.

## Проблема

Существующие решения являются либо ресурсоемкими, либо неустойчивыми к большой плотности объектов, что приводит к срыву слежения, перескоку траекторий, накоплению ошибок с течением временем.

## Решение

Использовать подход отбора объектов для ре-идентификации на основе оценки качества. Такое решение увеличит вычислительную эффективность метода, а также уменьшит число ошибок ре-идентификации.

## Задача слежения за множеством объектов

- *A. Bewley, G. Zongyuan, F. Ramos, B. Upcroft* Simple online and realtime tracking // ICIP, 2016, P. 3464–3468.
- *N. Wojke, A. Bewley, and D. Paulus* Simple online and realtime tracking with a deep association metric // ICIP, 2017, P. 3645–3649.

## Задача оценки качества

- *L. Best-Rowden, A. K. Jain.* Learning Face Image Quality From Human Assessments // IEEE Transactions on Information Forensics and Security, 2018.

## Существующая проблема

Некорректные переключения между объектами и разрывы траекторий при их продлении в сценариях с пересекающимися и временно исчезающими из кадра объектами, имеющими сложную модель движения.



Истинные траектории (слева), ложные переключения (справа)

## Возможное решение

Ре-идентификация – задача принятия решения о наличии объекта, присутствующего на предоставленном изображении, на множестве других имеющихся изображений.

## Обозначения

$Y_t = \{y_{j,1}, \dots, y_{j,t-1}, |j = 1, \dots, N\}$  – измерения всех  $N$  объектов на предыдущих кадрах,  $X_t = \{x_{j,t}, |j = 1, \dots, N\}$  – скрытые характеристики (состояния) всех  $N$  объектов на кадре  $t$ , подлежащие оцениванию,  $Z_t = \{z_{1,t}, \dots, z_{M_t,t}\}$  – измеренные характеристики не идентифицированных объектов на кадре  $t$  с помощью детектора объектов.

## Допущения

- $x_{j,t}$  зависит только от  $x_{j,t-1}$ ;  $y_{j,t}$  определяются только  $x_{j,t}$ ;
- процесс изменения характеристик  $x_{j,t}$  каждого объекта описывается моделью ЛДС, нормальность распределений:

$$p(x_{j,t}|x_{j,t-1}) = N(A_j x_{j,t-1}, \Gamma_j),$$

$$p(y_{j,t}|x_{j,t}) = N(B_j x_{j,t}, \Sigma_j),$$

$$p(x_{j,1}) = N(\mu_{j,0}, \Gamma_{j,0}).$$

- объекты не взаимодействуют и движутся независимо, одно обнаружение на кадре связано только с одним объектом;
- $x = (u, v, s, r, \dot{u}, \dot{v}, \dot{s})^T$ , где  $u$  и  $v$  – координаты,  $s$  и  $r$  – площадь и соотношение сторон прямоугольника локализации,  $\dot{u}, \dot{v}, \dot{s}$  – скорости.

Исходная задача – максимизация совместной апостериорной вероятности

$$\max_{X_t} p(X_t | Y_t, Z_t)$$

Факторизованная апостериорная вероятность

$$p(X_t | Y_t, Z_t) = \prod_{j=1}^N \left\{ \sum_{i=1}^{M_t} a_{i,j} p(x_{j,t} | y_{j,1}, \dots, y_{j,t-1}, z_{i,t}) \right\}$$

при условиях:  $a_{i,j} \in \{0, 1\}$ ,  $\sum_{i=1}^{M_t} a_{i,j} = 1$ ,  $\sum_{j=1}^N a_{i,j} = 1$ .

Подзадачи

$$\max_{a_{i,j}} \log p(\hat{X}_t | Y_t, Z_t) = \max_{a_{i,j}} \sum_{j=1}^N \sum_{i=1}^{M_t} a_{i,j} \log L_{\hat{x}_{j,t}}(y_{j,1}, \dots, y_{j,t-1}, z_{i,t})$$

$$\max_{X_t} \log p(X_t | Y_t, Z_t) = \sum_{j=1}^N \max_{x_{j,t}} \log p(x_{j,t} | y_{j,1}, \dots, y_{j,t-1}, y_{j,t}), y_{j,t} = \sum_{i=1}^{M_t} a_{i,j} z_{i,t}$$

## Шаг 1. Прогноз.

$$\hat{x}_{j,t} = \arg \max_{x_{j,t}} \log p(x_{j,t} | y_{j,1}, \dots, y_{j,t-1}) = \arg \max_{x_{j,t}} \log N(x_{j,t} | \hat{x}_{j,t}, \hat{V}_{j,t})$$

$$\hat{x}_{j,t} = \mathbf{A}_j \mu_{j,t-1}, \hat{V}_{j,t} = \Gamma_j + \mathbf{A}_j \hat{V}_{j,t-1} \mathbf{A}_j^T.$$

## Шаг 2. Задача о назначениях между $Z_t$ и $\{y_{j,t} | j = 1, \dots, N\}$

$$\min_{a_{i,j}} \sum_{j=1}^N \sum_{i=1}^{M_t} a_{i,j} \mathbf{C}_{i,j}, \quad y_{j,t} = \sum_{i=1}^{M_t} a_{i,j} z_{i,t}$$

где  $\mathbf{C}_{i,j} = -\log L_{\hat{x}_{j,t}}(y_{j,1}, \dots, y_{j,t-1}, z_{i,t})$ ,  $L_{\hat{x}_{j,t}}$  – функция правдоподобия;

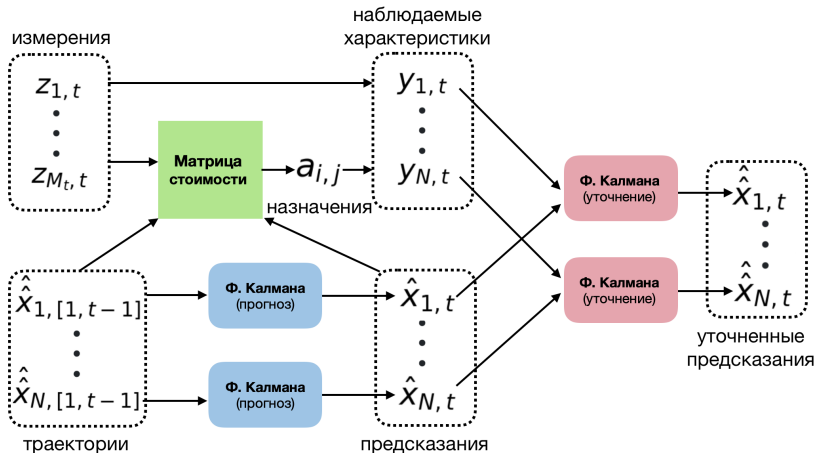
## Шаг 3. Коррекция.

$$\hat{\hat{x}}_{j,t} = \arg \max_{x_{j,t}} \log p(x_{j,t} | y_{j,1}, \dots, y_{j,t-1}, y_{j,t}) = \arg \max_{x_{j,t}} \log N(x_{j,t} | \hat{\hat{x}}_{j,t}, \hat{V}_{j,t})$$

$$\hat{\hat{x}}_{j,t} = \hat{x}_{j,t} + \mathbf{K}_{j,t}(y_{j,t} - \mathbf{B}_j \hat{x}_{j,t}), \hat{V}_{j,t} = (\mathbf{I} - \mathbf{K}_{j,t} \mathbf{B}_j) \hat{V}_{j,t},$$

$$\mathbf{K}_{j,t} = \hat{V}_{j,t} \mathbf{B}_j^T (\mathbf{B}_j \hat{V}_{j,t} \mathbf{B}_j^T + \Sigma_j)^{-1}.$$

# Схема решения задачи слежения на кадре $t$



Алгоритм слежения за множеством объектов для кадра  $t$



- Определяется отображение  $h(\cdot|\theta)$  пространства изображений, содержащих объект исследуемого класса, в пространство дескрипторов — векторное пространство фиксированной размерности, на котором определена операция скалярного произведения.
- Отображение  $h(\cdot|\theta)$  задается сверточной нейронной сетью архитектуры ResNet18 с  $L_2$ -нормализацией выхода (Re-ID сеть).
- Нейросеть обучается в задаче классификации объектов исследуемого класса с Cosine Softmax Cross-Entropy функцией потерь:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(\kappa \cdot \tilde{\mathbf{W}}_{c_i} \mathbf{d}_i)}{\sum_{j=1}^N \exp(\kappa \cdot \tilde{\mathbf{W}}_j \mathbf{d}_j)}, \quad \tilde{\mathbf{W}} = \frac{\mathbf{W}}{\|\mathbf{W}\|_2}, \quad \|\mathbf{d}_i\|_2 = 1,$$

где  $\mathbf{W}$  — матрица весов,  $\mathbf{d}_i = h(\mathbf{I}_i|\theta)$  — дескриптор,  $c_i$  — метка класса,  $\kappa$  — параметр масштабирования.

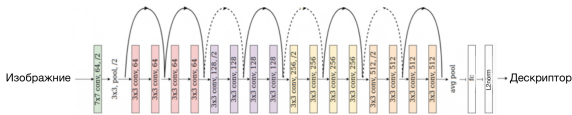


Схема слоев архитектуры Resnet18 с  $L_2$ -нормализацией выхода

## Связь с задачей сопровождения

После обнаружения и локализации объектов на кадре требуется их сопоставить с ранее найденными траекториями, чтобы для каждой из траектории произвести уточнение прогноза.

## Матрица стоимости

$$\begin{aligned} C_{i,j} &= -\log L_{\hat{x}_{j,t}}(\mathbf{y}_{j,1}, \dots, \mathbf{y}_{j,t-1}, \mathbf{z}_{i,t}) = -\log N(\mathbf{z}_{i,t} | \mathbf{B}_j \hat{\mathbf{x}}_{j,t}, \Sigma_j) = \\ &= \frac{1}{2} (\mathbf{z}_{i,t} - \mathbf{B}_j \hat{\mathbf{x}}_{j,t})^T \Sigma_j^{-1} (\mathbf{z}_{i,t} - \mathbf{B}_j \hat{\mathbf{x}}_{j,t}) + \frac{m}{2} \log 2\pi + \frac{1}{2} \log |\Sigma_j|, \end{aligned}$$

где  $m$  размерность вектора  $\mathbf{x}$ ;

## Поправка к матрице стоимости

$\mathbf{l}_i$  – изображение наблюдения  $i$  на текущем кадре,

$\mathbf{l}_{j,1}, \dots, \mathbf{l}_{j,\tau}$  – изображения в траектории объекта  $j$ .

Выделяются дескрипторы  $\mathbf{d}_i = h(\mathbf{l}_i | \boldsymbol{\theta})$ ;  $\mathbf{d}_{j,t} = h(\mathbf{l}_{j,t} | \boldsymbol{\theta})$ ,  $t \in \overline{1, \tau}$ .

$$\tilde{C}_{i,j} = C_{i,j} + r \left[ 1 - \max \left\{ \frac{\langle \mathbf{d}_i, \mathbf{d}_{j,1} \rangle}{\|\mathbf{d}_i\| \|\mathbf{d}_{j,1}\|}, \dots, \frac{\langle \mathbf{d}_i, \mathbf{d}_{j,\tau} \rangle}{\|\mathbf{d}_i\| \|\mathbf{d}_{j,\tau}\|} \right\} \right]^2,$$

где  $r$  – масштабный множитель.

## Качество в биометрии

Качество объекта – мера полезности объекта для задачи распознавания.

## Отбор кандидатов для ре-идентификации

- $I_1, \dots, I_\tau$  – изображения в траектории, функция оценки качества  $q : I_t \mapsto q_t$ ,  $q_t \in [0, 1]$  – показатель "качества".
- Производится сортировка элементов в порядке убывания показателя "качества", задающаяся перестановкой  $p(1)p(2) \dots p(\tau)$ :

$$q_{j,p(1)} \geq q_{j,p(2)} \geq \dots \geq q_{j,p(\tau)}.$$

- Отбор осуществляется выбором  $K$  кандидатов  $I_{p(1)}, \dots, I_{p(K)}$  с наибольшим значением качества.

## Методы оценки качества

- Оценка качества (*L. Best-Rowden*) [*Deep-QA-SORT*]
  - обучение с учителем, построение регрессии признакового представления объектов на ассессорскую оценку качества, модель SVM;
  - признаковое представление извлекается сверточной нейронной сетью, подобной сети ре-идентификации;
- Оценка качества на основе уверенности (confidence) детектора

## Показатели качества

- Precision – точность;
- Recall – полнота;
- IDS – суммарное число некорректных переключений при продлении траекторий;
- Hz – частота работы (в кадрах в секунду).

## Базы данных

- 1 MOT20-01 – данные камеры наблюдения, высокий ракурс;
- 2 MOT20-02 – данные камеры наблюдения, средний ракурс;

Выборка	FPS	Плотность на кадр	Длина	Траектории
MOT20-01	25	42.1	429	90
MOT20-02	25	72.7	2782	296

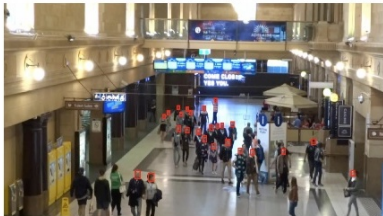
Описание выборок

## Локализация лиц

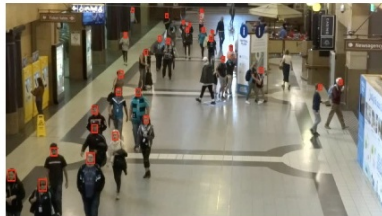
Для выделения областей, содержащих объекты, используются детекторы разной сложности: SSD и RetinaNet.

Архитектура	Число обучаемых параметров
SSD	4 М
RetinaNet	45 М

MOT20-01



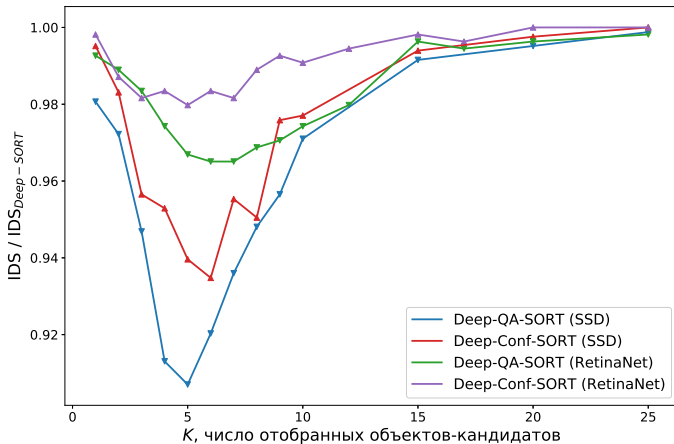
MOT20-02



Примеры локализаций лиц на кадрах из выборок MOT20-01, MOT20-02

Детектор	Метод слежения	Precision	Recall	IDS	Hz
выборка MOT20-02					
SSD	Deep-SORT	0.851	<b>0.910</b>	828	22.6
	Deep-Conf-SORT	<b>0.860</b>	0.904	778	<b>31.8</b>
	Deep-QA-SORT	0.855	0.907	<b>751</b>	31.8*
RetinaNet	Deep-SORT	0.887	<b>0.945</b>	543	22.4
	Deep-Conf-SORT	0.891	0.941	533	<b>31.4</b>
	Deep-QA-SORT	<b>0.896</b>	0.939	<b>526</b>	31.4*
выборка MOT20-01					
SSD	Deep-SORT	0.847	<b>0.895</b>	203	39.2
	Deep-Conf-SORT	<b>0.850</b>	0.890	202	<b>44.7</b>
	Deep-QA-SORT	0.848	0.892	<b>195</b>	44.7*
RetinaNet	Deep-SORT	0.871	<b>0.922</b>	168	40.9
	Deep-Conf-SORT	<b>0.875</b>	0.919	166	<b>46.0</b>
	Deep-QA-SORT	0.873	0.921	<b>162</b>	46.0*

\* Время работы алгоритма оценки качества не учтено



Зависимость отношения IDS метода к IDS базового метода от числа отобранных кандидатов  $K$

- Постановка задачи одновременного онлайн сопровождения объектов на видео изображении сведена к подзадачам о назначениях обнаруженных объектов и фильтрации параметров объекта.
- Предложен метод модификации матрицы стоимости в задаче назначений для использования метода ре-идентификации.
- Предложен подход отбора объектов на основе оценки качества изображения объекта при подсчете матрицы стоимости.
- Предложен метод оценки качества на основе уверенности детектора, не требующий дополнительных вычислений.
- Проведенные вычислительные эксперименты показали эффективность предложенных методов, уменьшение числа некорректных переключений между траекториями разных объектов при увеличении скорости работы системы слежения в 1.4 раза по сравнению со стандартным подходом.