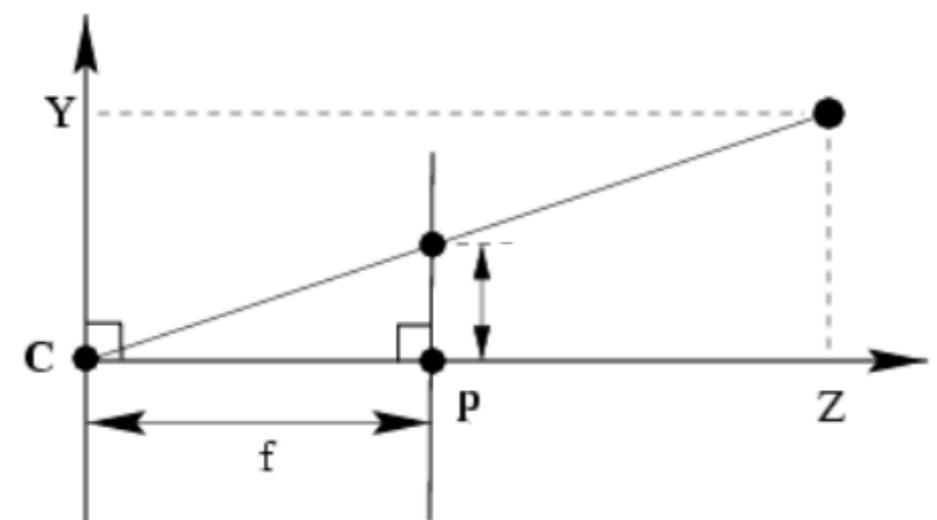
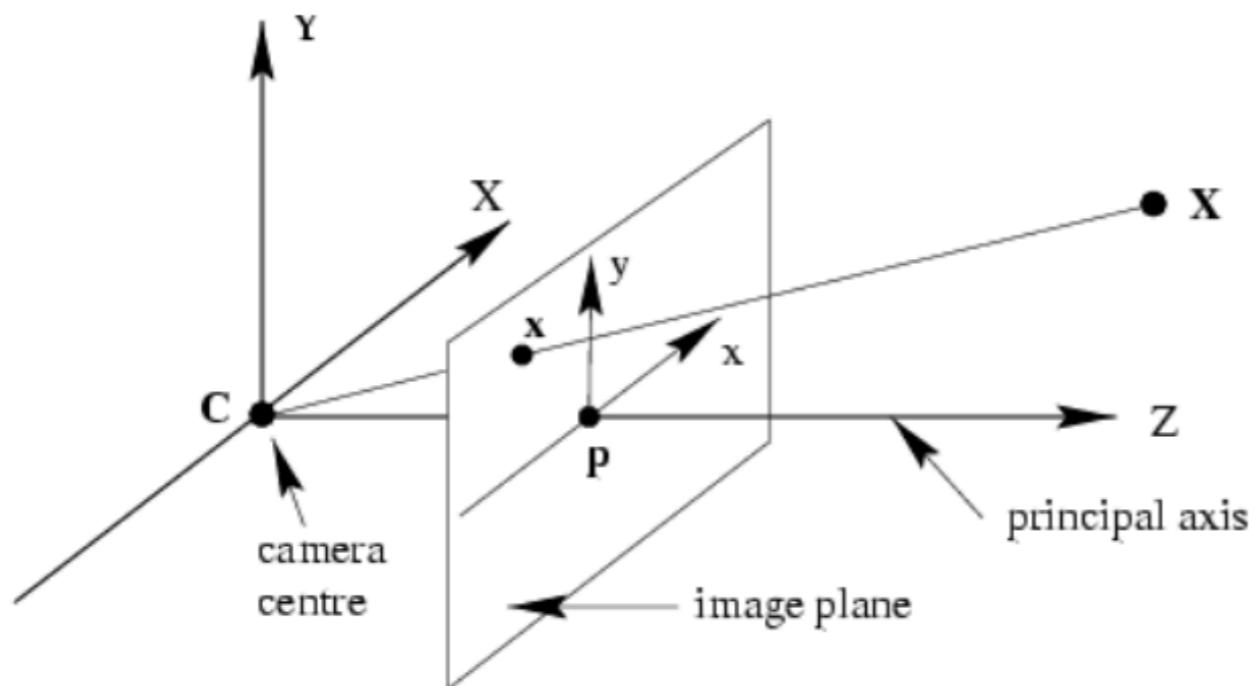


Camera Geometry and Sensor Fusion

Camera Geometry

- Камера отображает реальность, исходя из законов геометрии
- Зная эти законы, мы можем делать выводы о состоянии 3D мира

Pinhole Camera Projection



По подобию треугольников: $x/f = X/Z$, $y/f = Y/Z$
где f - это фокальное расстояние

Pinhole Camera Projection

- Но это не удобно
- Проекция - нелинейная, а хочется линейную
- Зачем? Например объединять несколько преобразований в одну матрицу

Гомогенные координаты

- Давайте перейдем в новые координаты
- Для точки (x, y) на плоскости гомогенные координаты это $(ax, ay, a) = (X, Y, W)$, для любой $a > 0$
- Можем положить $a = 1$
- Чтобы вернуться в исходные координаты, нужно лишь поделить на W : $(X/W, Y/W, 1)$

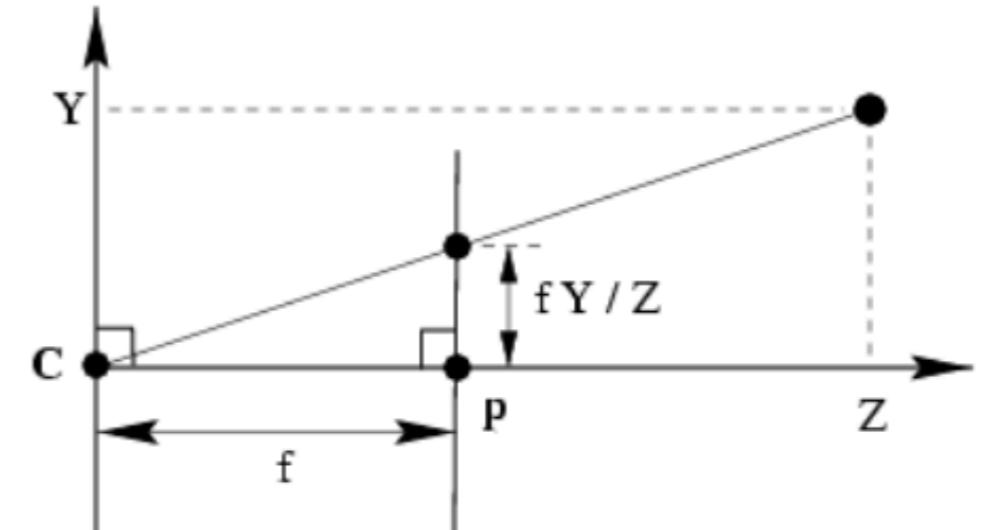
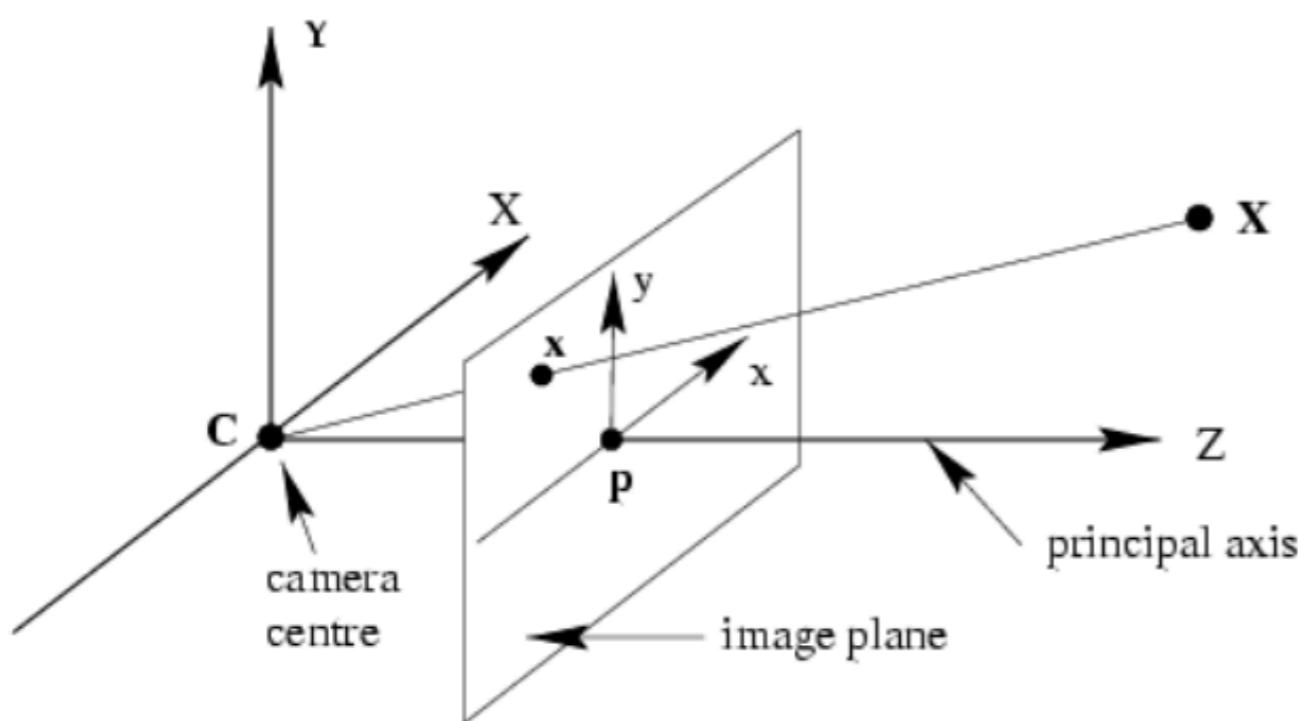
Гомогенные координаты Линии.

- Уравнение прямой: $ax + by + c = 0$
- Что меняется?
- $aX + bY + cW = 0$
- Ничего! Мы просто умножили вектор на константу, линии остались линиями.

Pinhole Camera Projection

$$\begin{pmatrix} fx \\ fy \\ z \end{pmatrix} = \begin{bmatrix} f & & & \\ & f & & \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

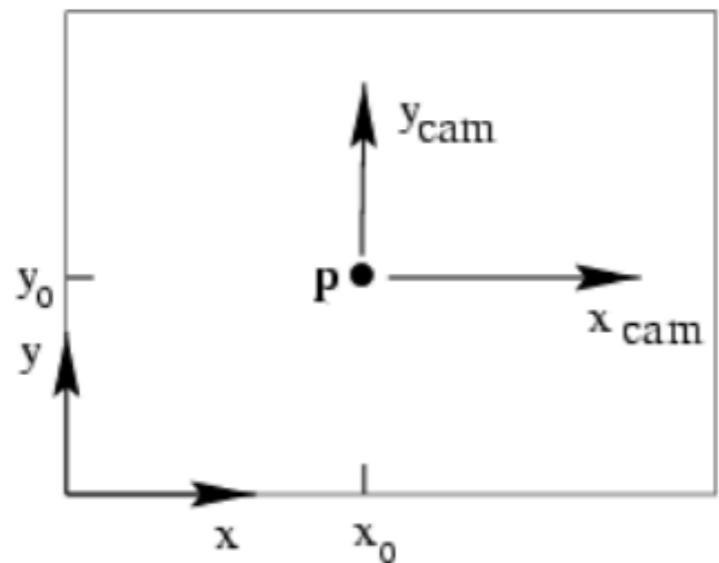
3x4
Projection
Matrix



Матрица калибровки камеры

- Пересечение оси Z и плоскости картинки, лежит не в начале координат

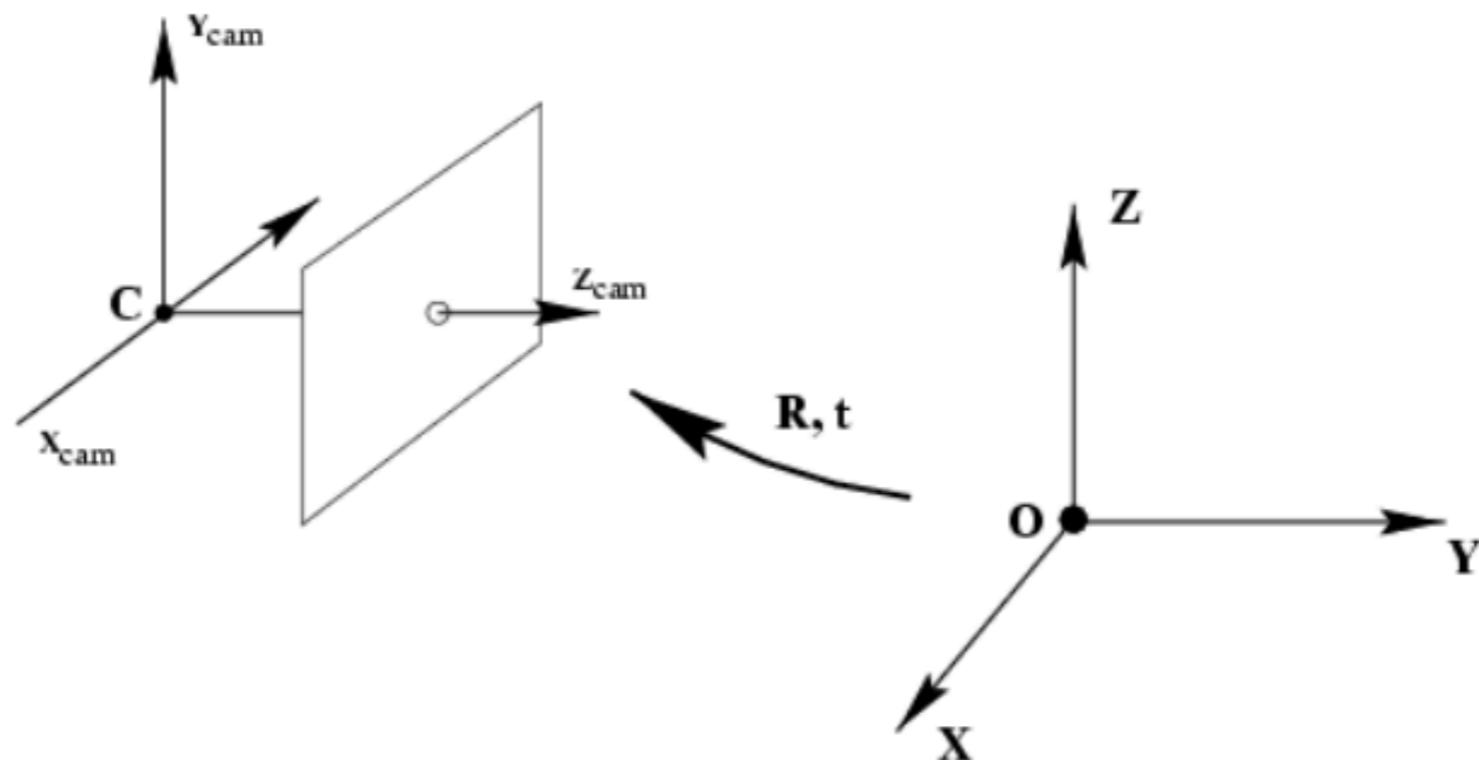
$$\begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \begin{bmatrix} f & p_x & 0 \\ f & p_y & 0 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$



$$K = \begin{bmatrix} f & p_x \\ f & p_y \\ 1 \end{bmatrix} \quad \text{(Intrinsic) Calibration matrix}$$

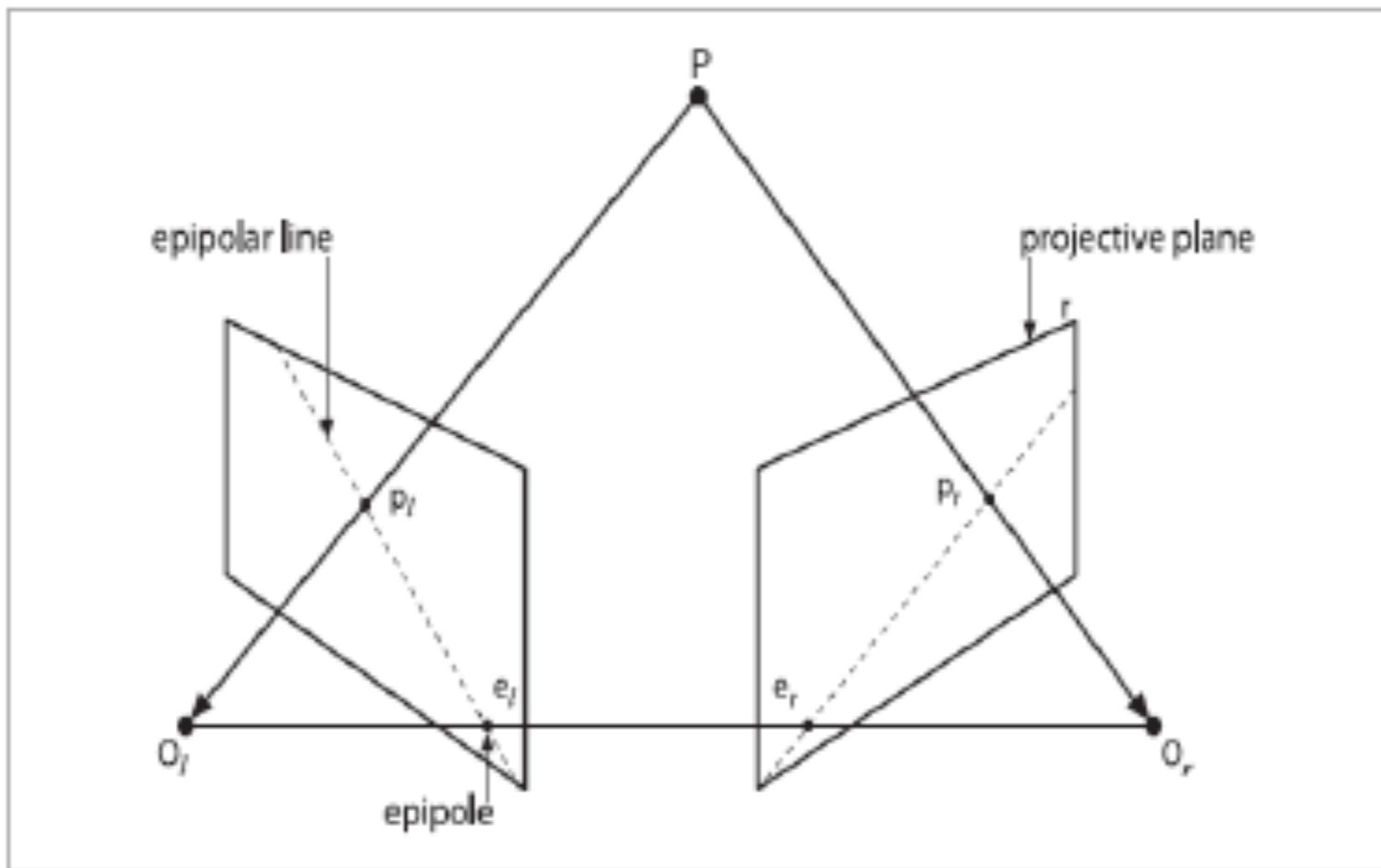
Экстринсики камеры

- Камера можно двигаться в 3D мире
- Нам часто хочется знать ее положение в 3D мире
- Хотим вычислить Rotation и Transform относительно референсной точки 3D мира



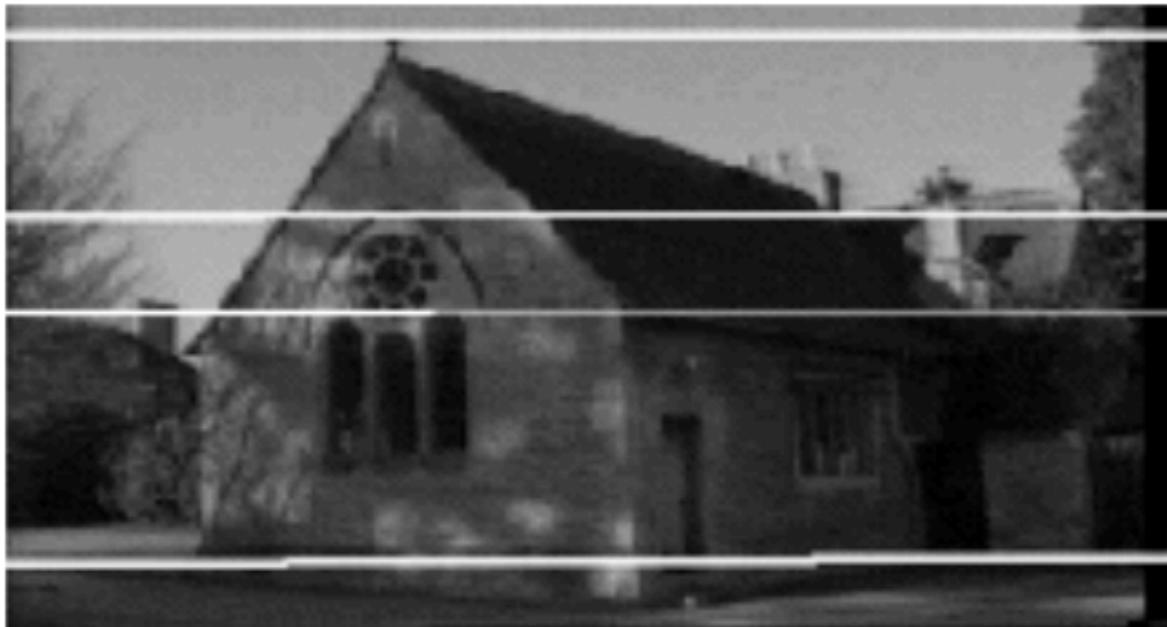
Стерео пара

- Скалированная система из двух камер позволяет видеть в 3D



Epipoles

- Зная все параметры камеры (интринсики и экстринсики), мы можем найти на картинках линии соответствующие одним и тем же точкам в 3D



MultiView Pose Estimation

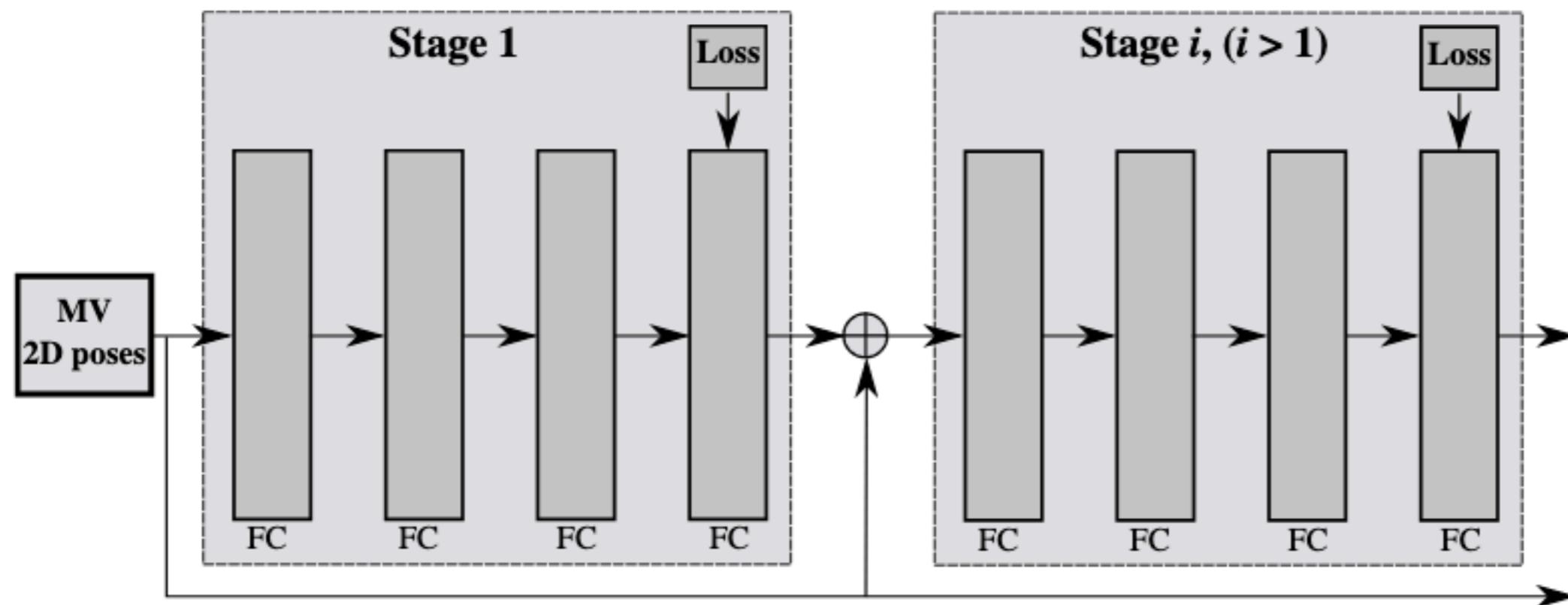
- Имея несколько камер, необходимо определить позу человека
- Датасеты:
 - Human3.6M
 - CMU Panoptic



A generalizable approach for multi-view 3D human pose regression



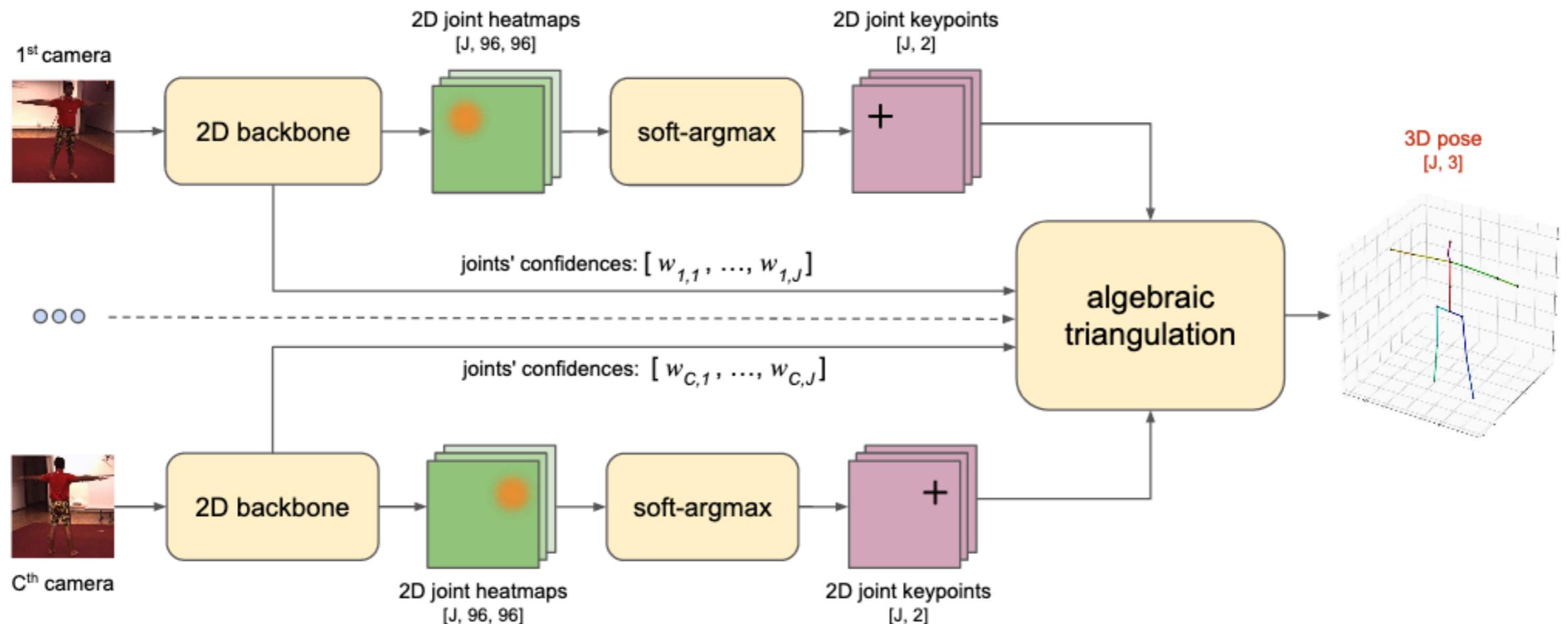
A generalizable approach for multi-view 3D human pose regression



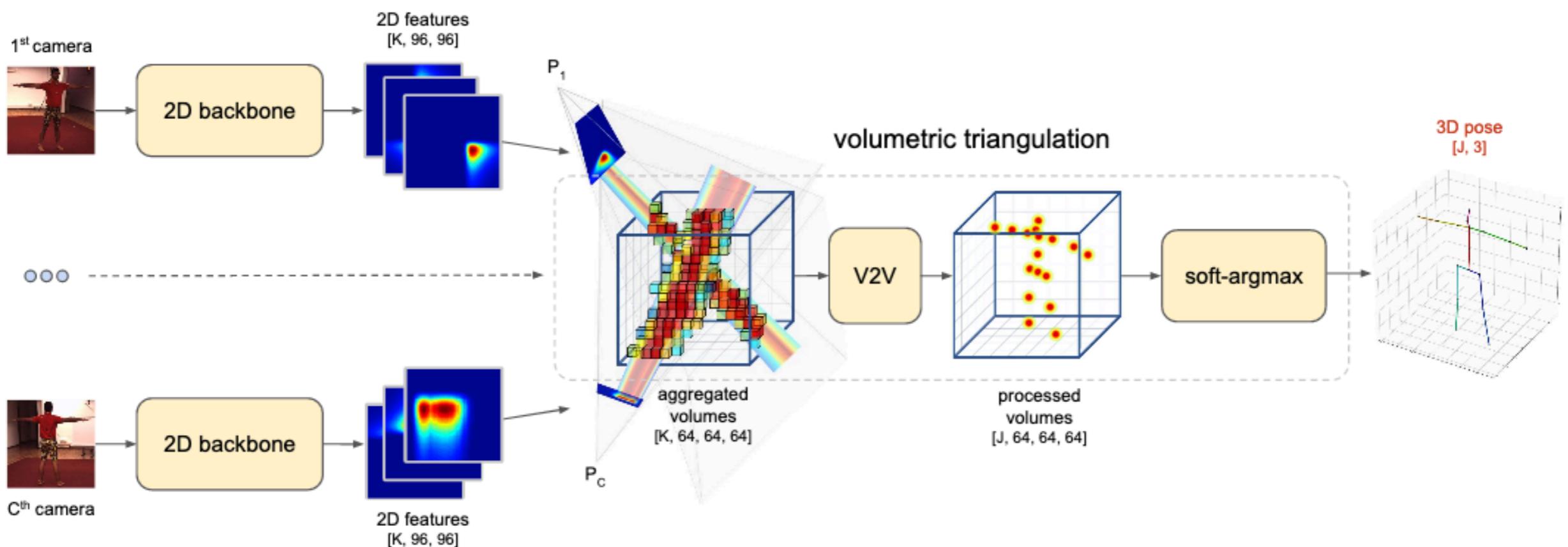
Learnable Triangulation of Human Pose

<https://www.youtube.com/watch?v=z3f3aPSuhqg>

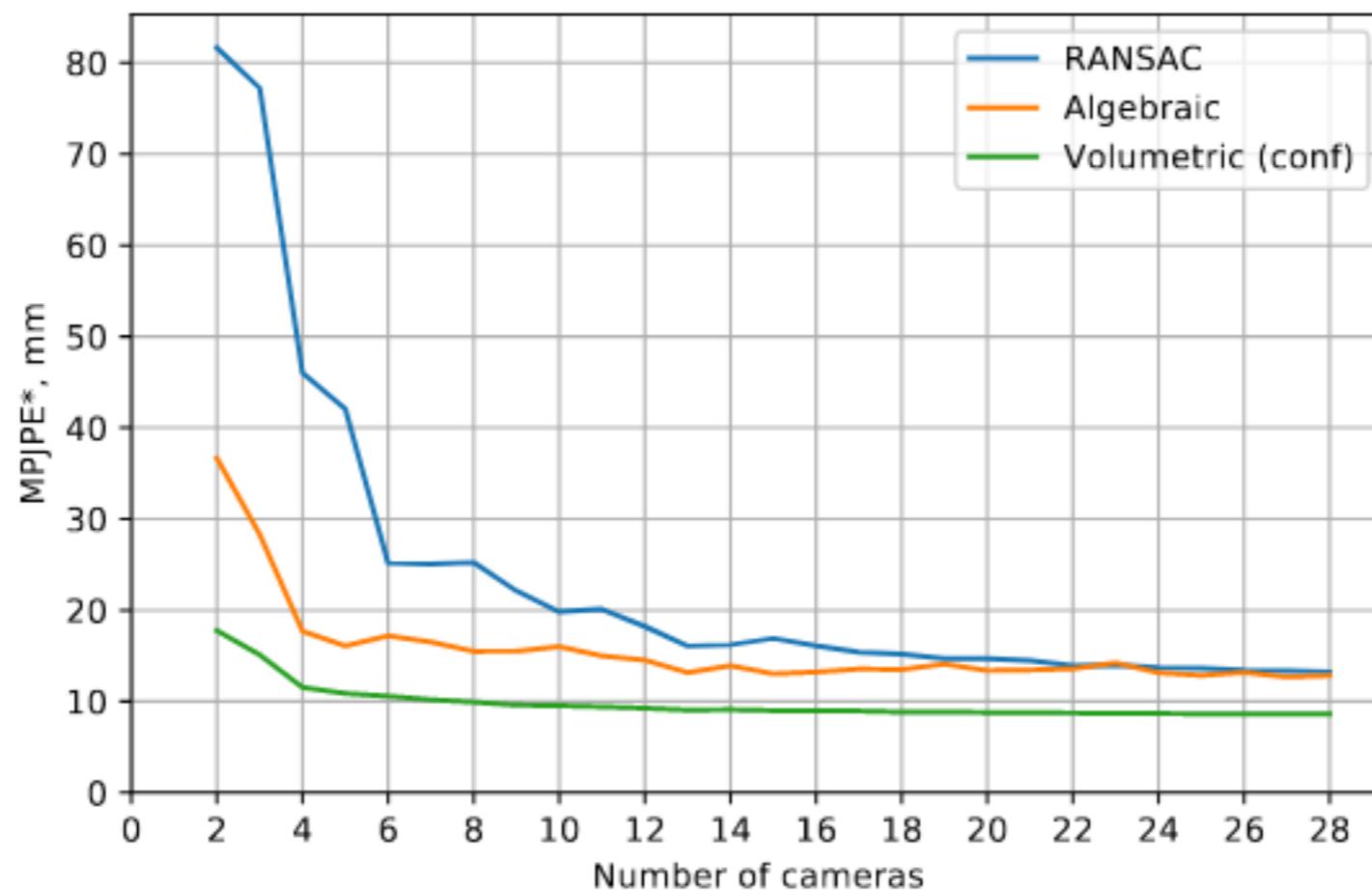
Learnable Triangulation of Human Pose: algebraic triangulation



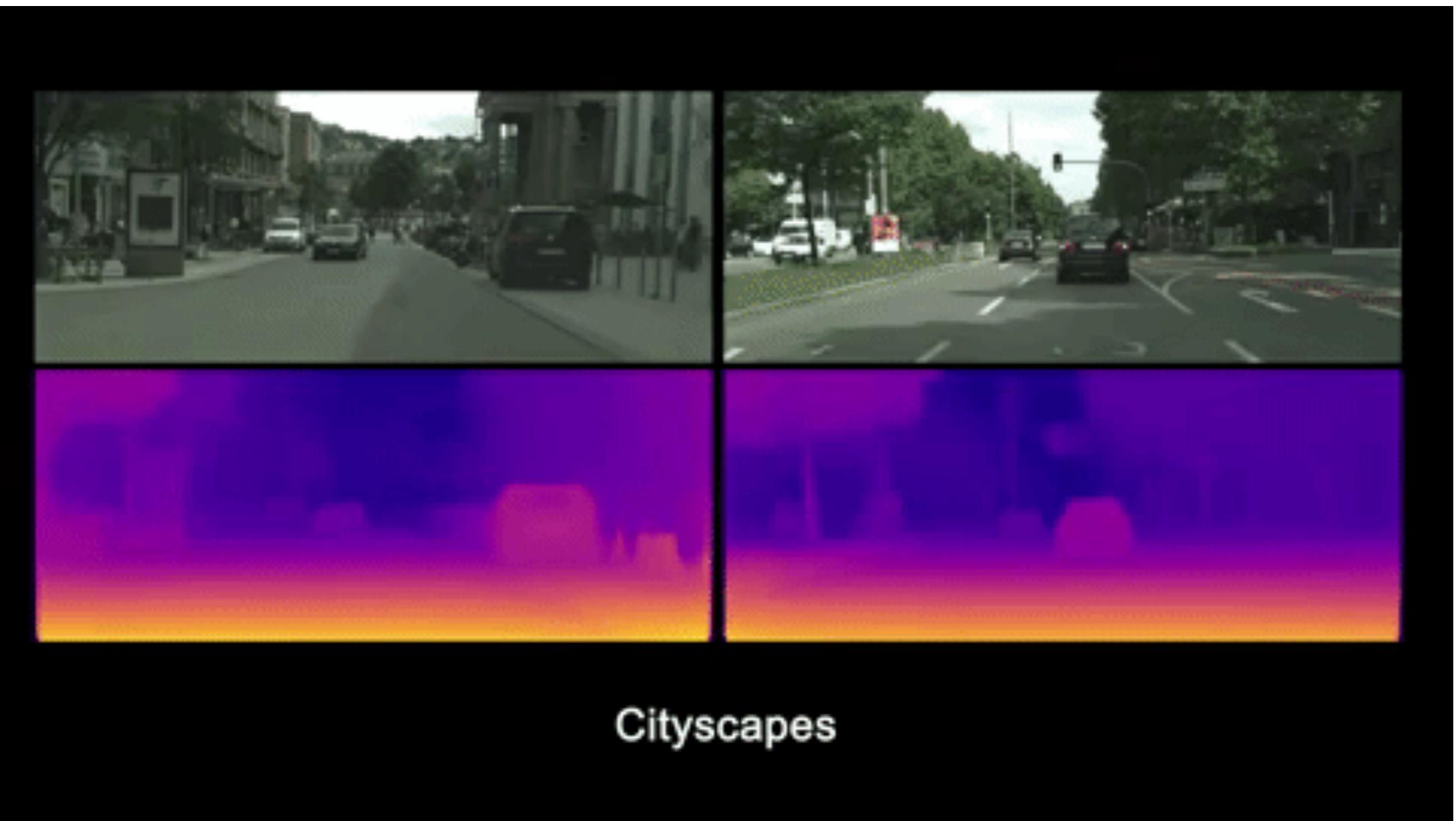
Learnable Triangulation of Human Pose: volumetric triangulation



Learnable Triangulation of Human Pose: multiple cameras

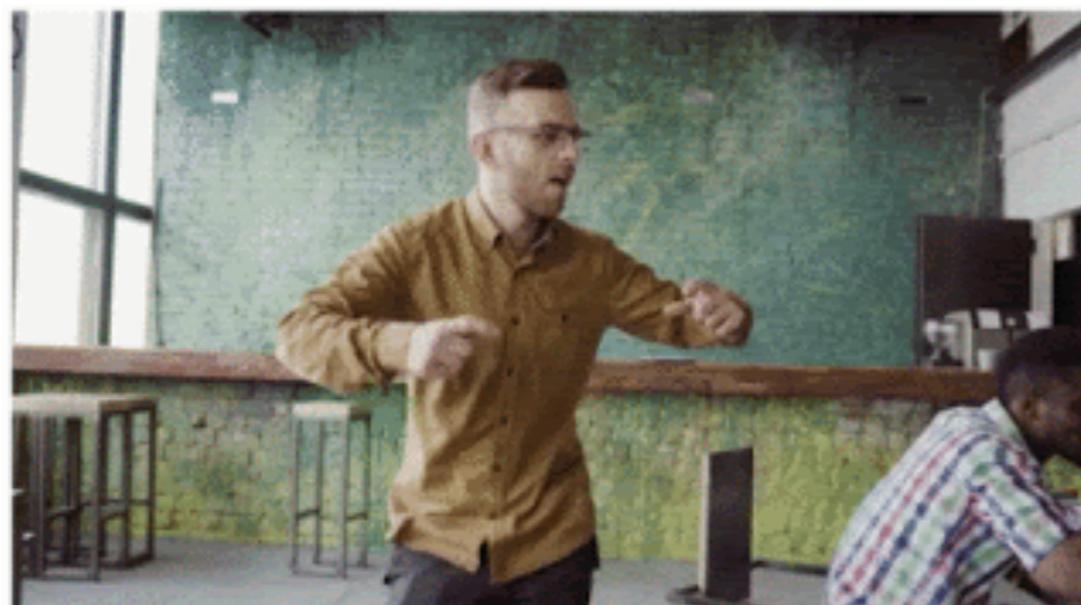


Monocular Depth



Learning the Depths of Moving People by Watching Frozen People

- Одна камера
- Двигаются только люди
- Камера тоже двигается



Input video



Our depth predictions

Идея

- Давайте восстановим движение камеры
- Для статических объектов оценим глубину с помощью стерео геометрии
- Людей отфильтруем через Mask-RCNN
- Будем учить сеть закрашивать глубину на людях

Идея



(a) Reference image I^r

(b) Human mask M

(c) Input depth D_{pp}

- Но где взять ground truth?

Идея

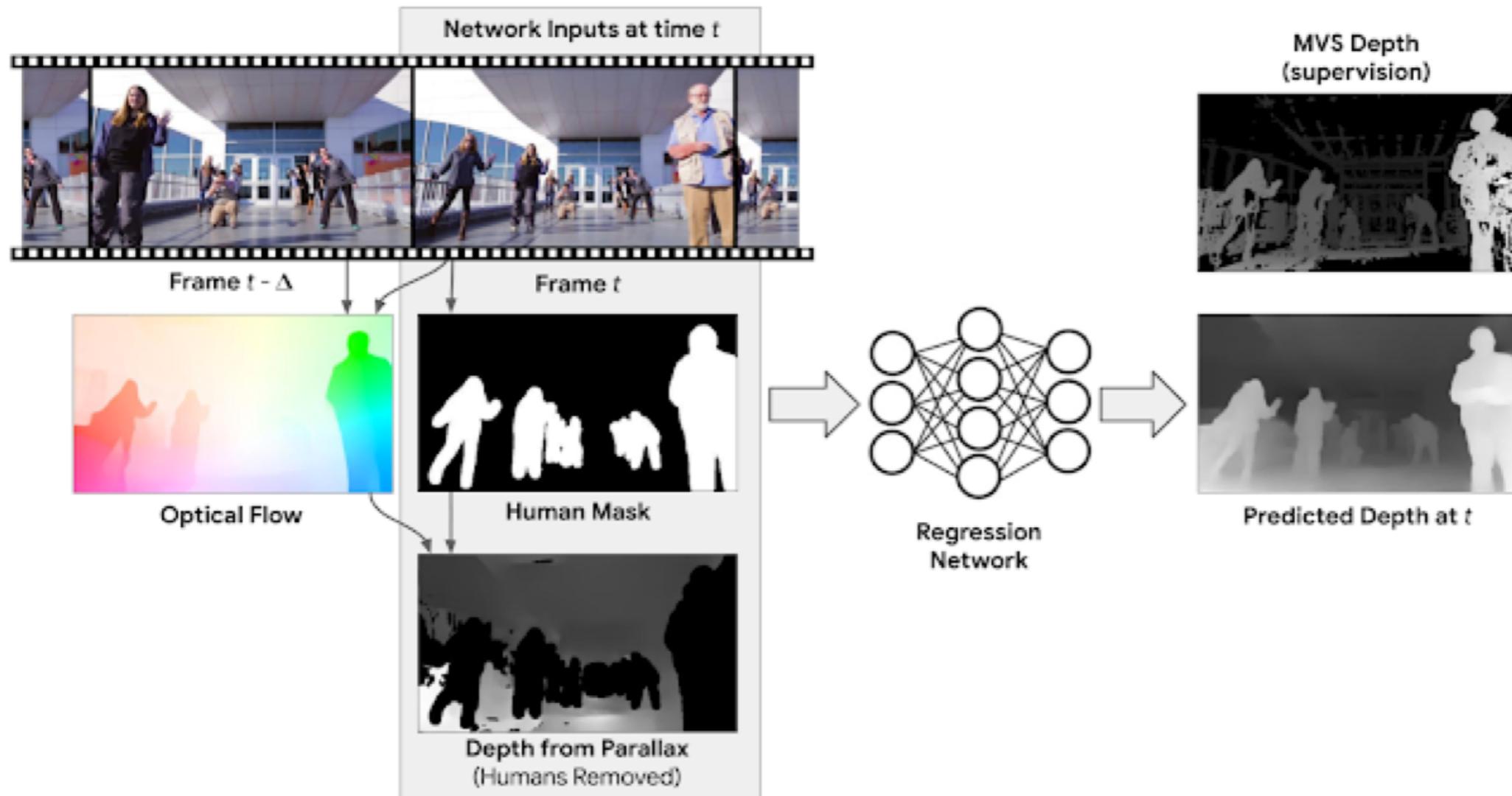
- Давайте использовать видео, где люди тоже не двигаются
- Тогда мы сможем оценить глубину и для людей используя стерео геометрию

https://www.youtube.com/watch?v=z_asR0mTeQQ

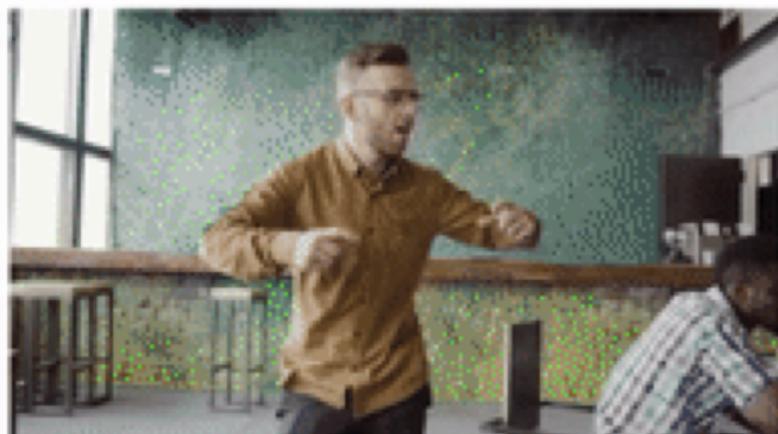
Как восстановить движение камеры?

- SLAM!
 - Матчим точки
 - Обновляем информацию о параметрах камеры и ее движении

Модель



Эксперименты



Input



DORN (monocular)



Chen et al. (monocular)



DeMoN (stereo)



Ours

ЧТО МОЖНО С ЭТИМ делать?

Synthetic Defocus



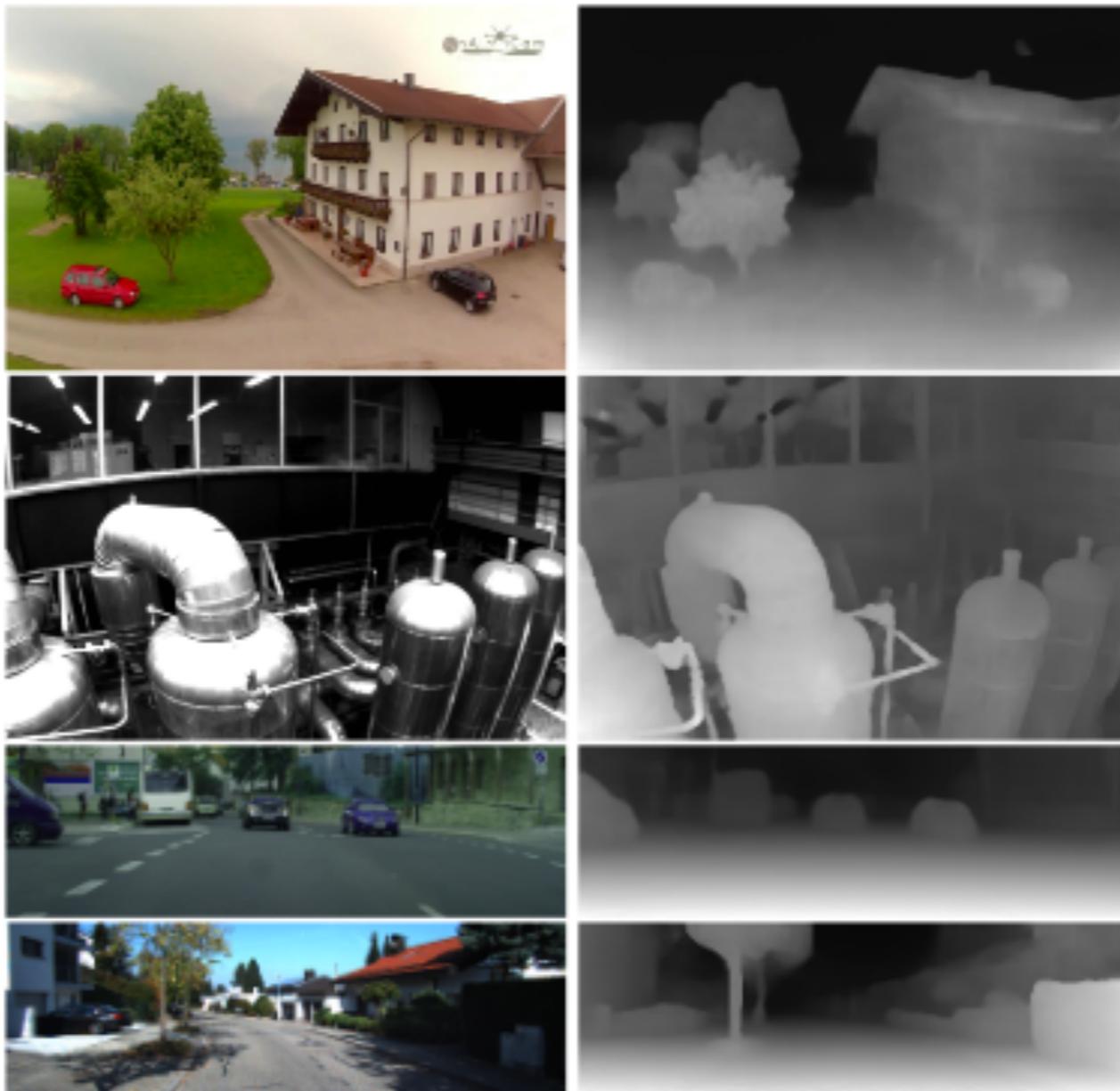
Input video



Estimated depth

https://www.youtube.com/watch?v=fj_fK74y5_0

Depth from Videos in the Wild



Идея

- Будем предсказывать параметры камеры, движение камеры и глубину
- Имея предсказание глубины для двух последовательных кадров, и зная наше движение, мы можем проверить качество нашего предсказания через соответствие двух кадров

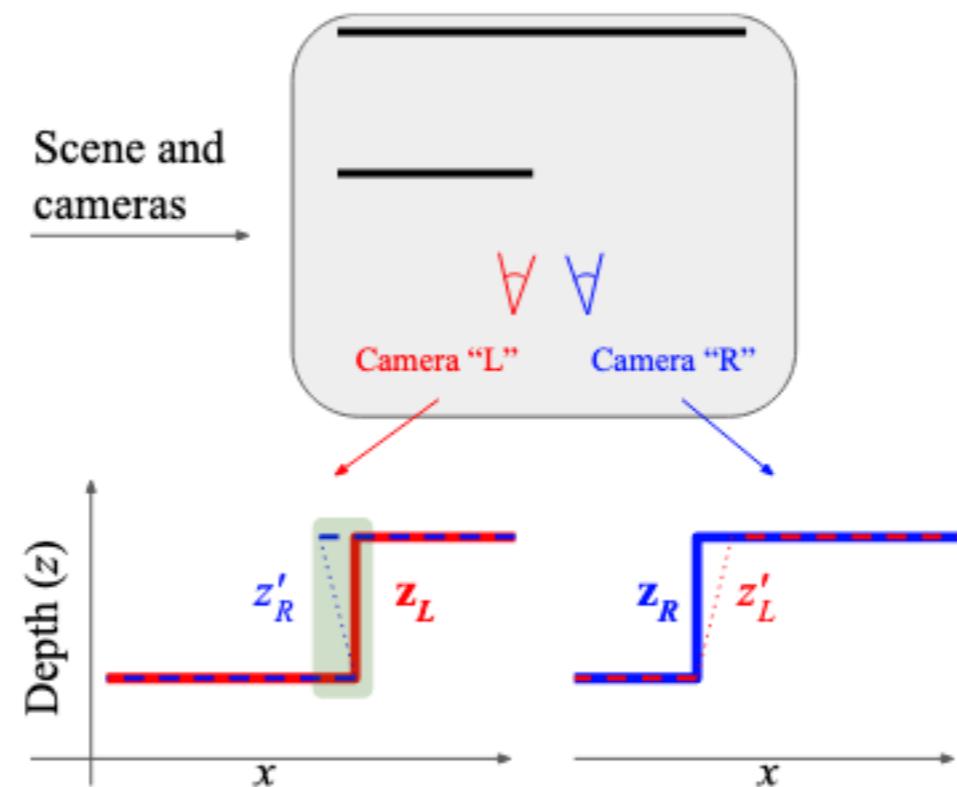
Идея

- Динамические объекты мешают
- Давайте замаскируем их детектором

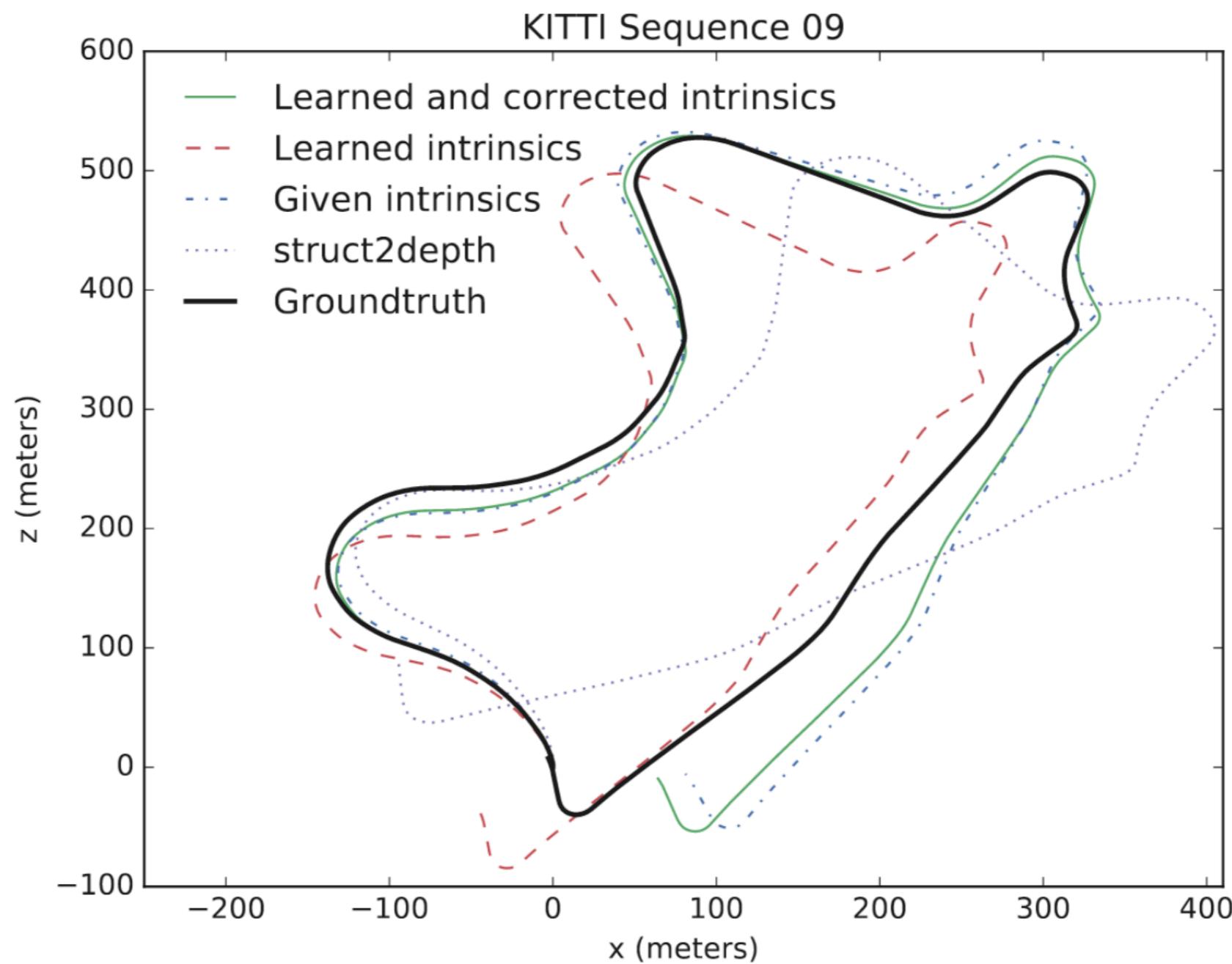


Идея

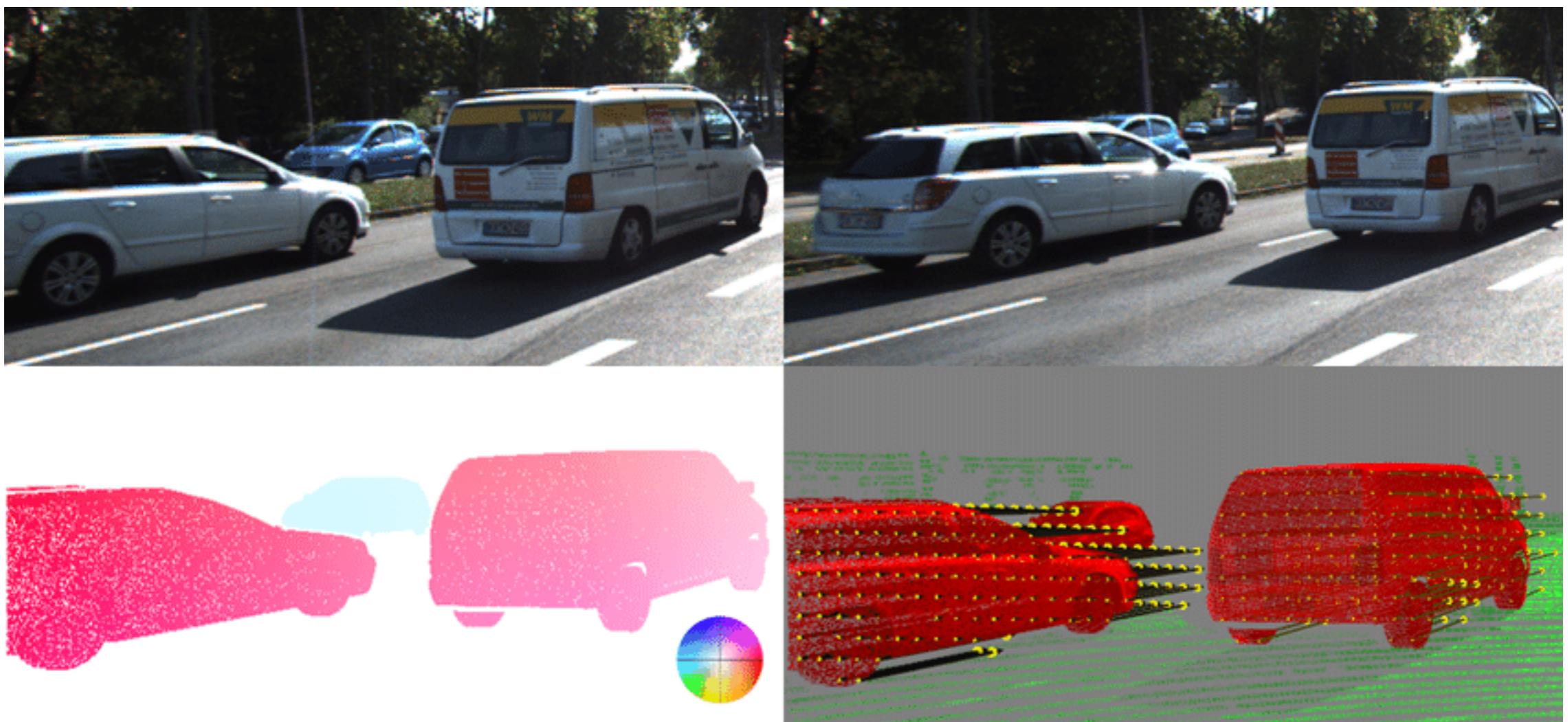
- Оклюзии мешают
- Мы можем пропускать лосс только на тех пикселях, которые видим на обеих кадрах



Движение камеры



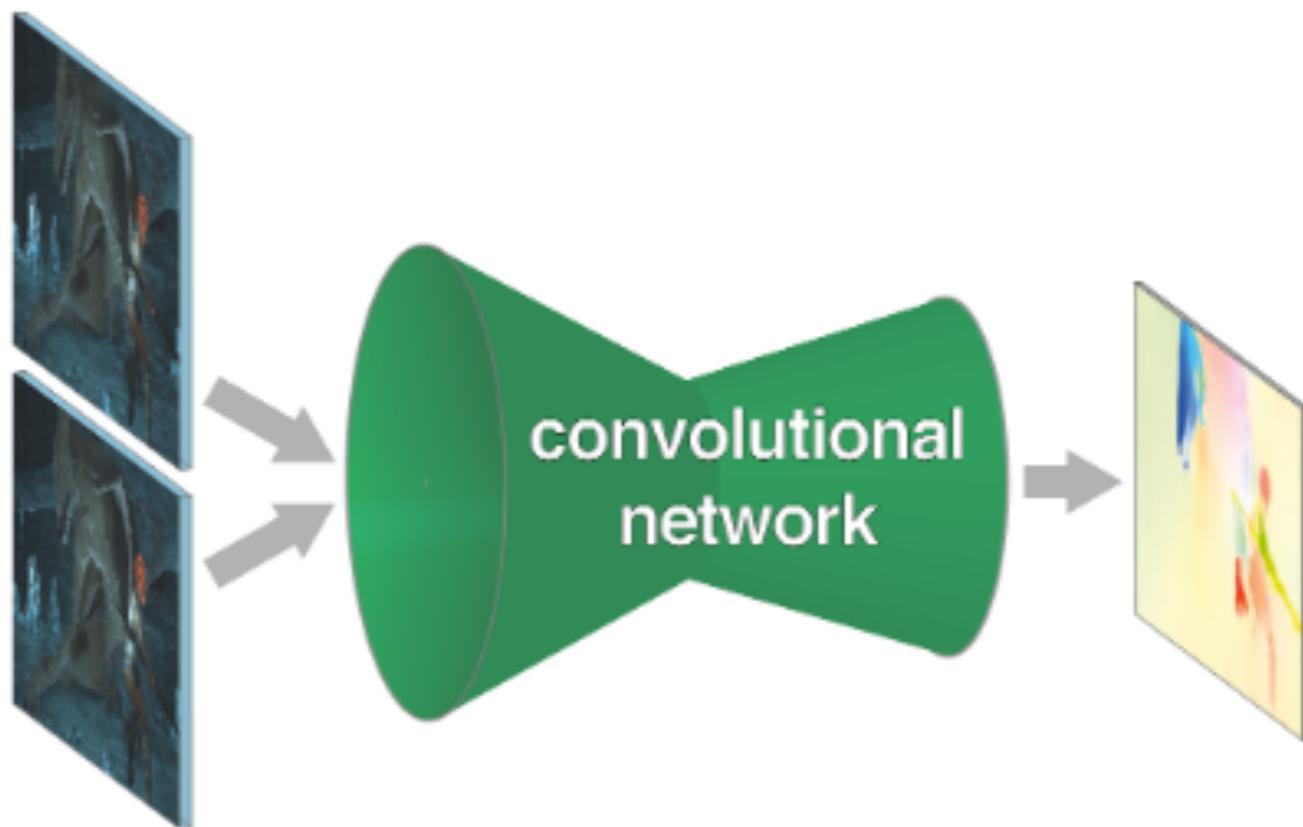
Optical Flow



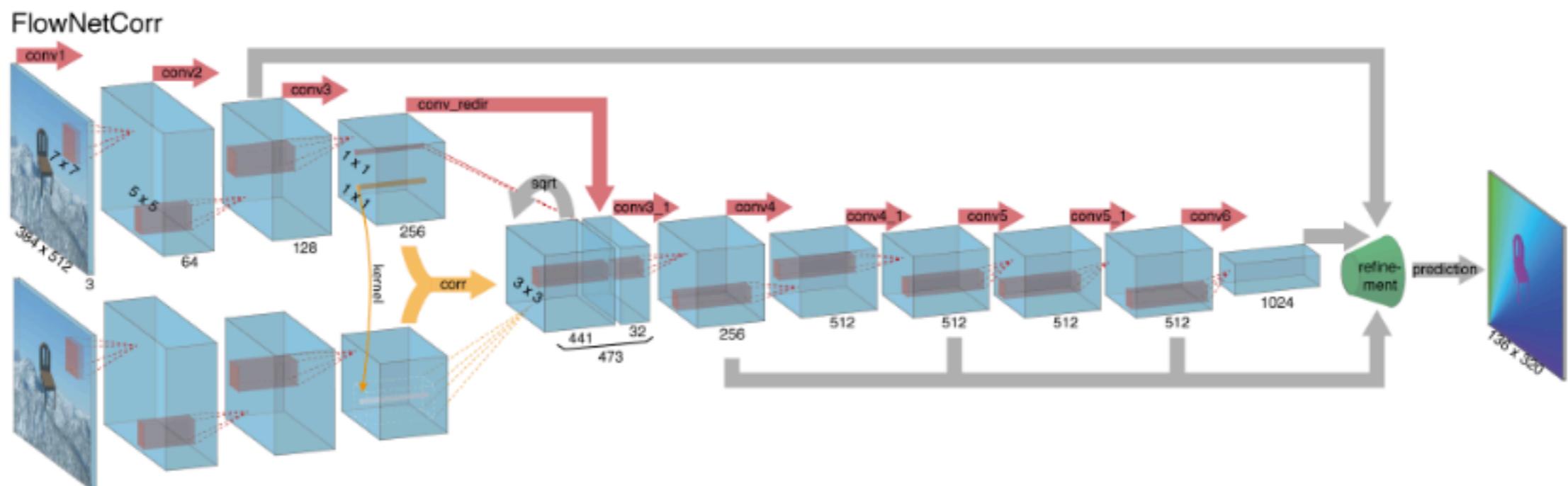
Датасеты

- Тяжело (невозможно?) получить ground truth на реальных данных
- На помощь приходят симуляторы: <https://www.youtube.com/watch?v=1iAQp6KmhE4>

FlowNet



FlowNet



<https://www.youtube.com/watch?v=JSzUdVBmQP4>