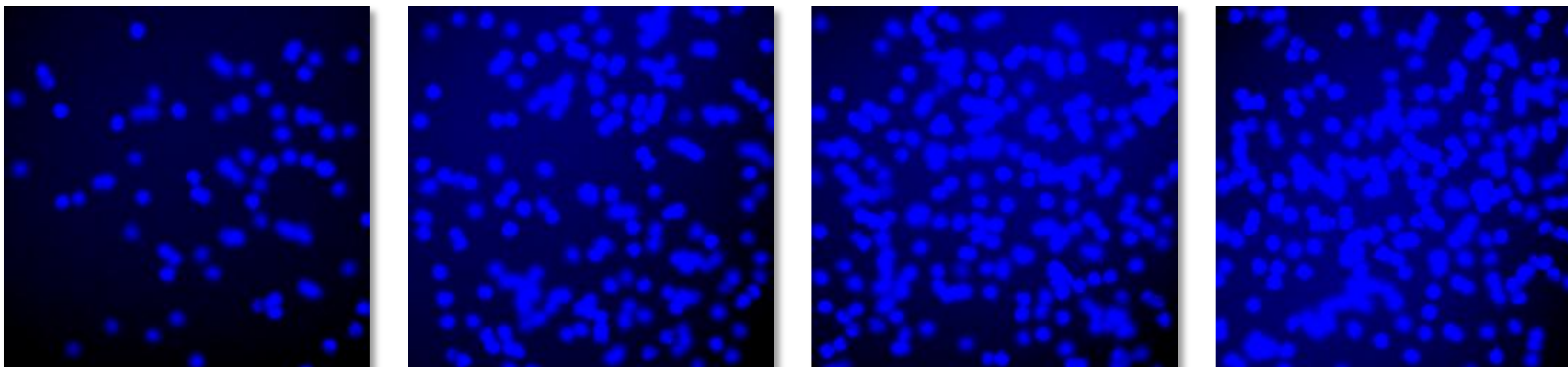
Learning To Count Objects in Images

NIPS 2010
Vancouver

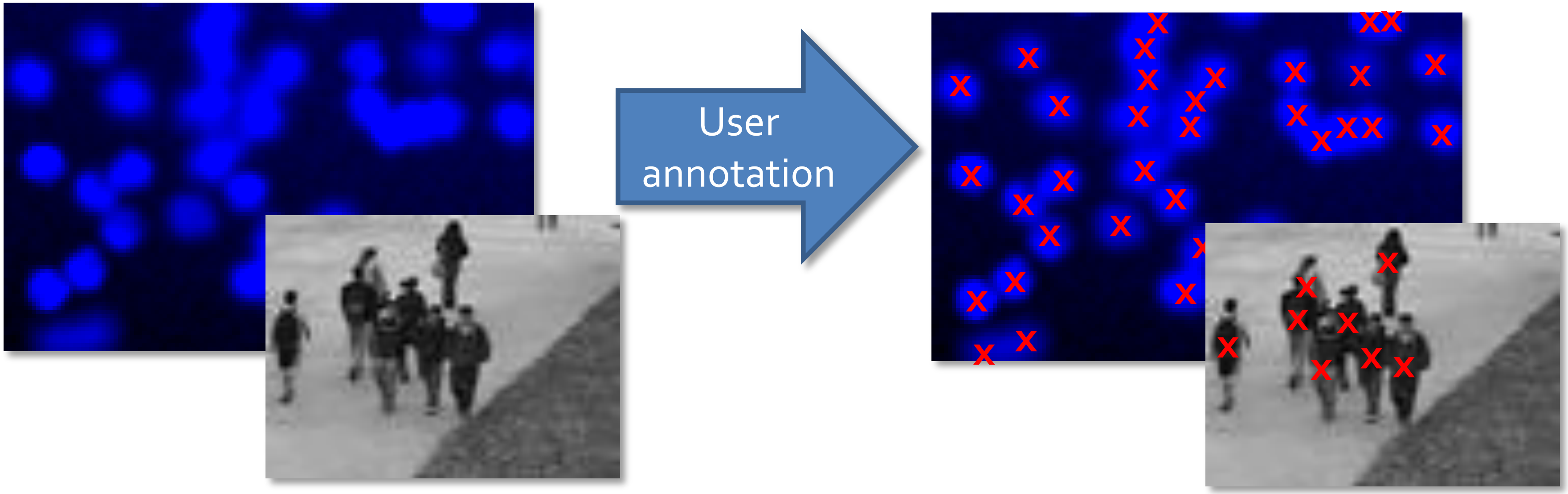
Victor Lempitsky **Andrew Zisserman**

Visual Geometry Group
University of Oxford



How many cells in each image?

How many people in each image?

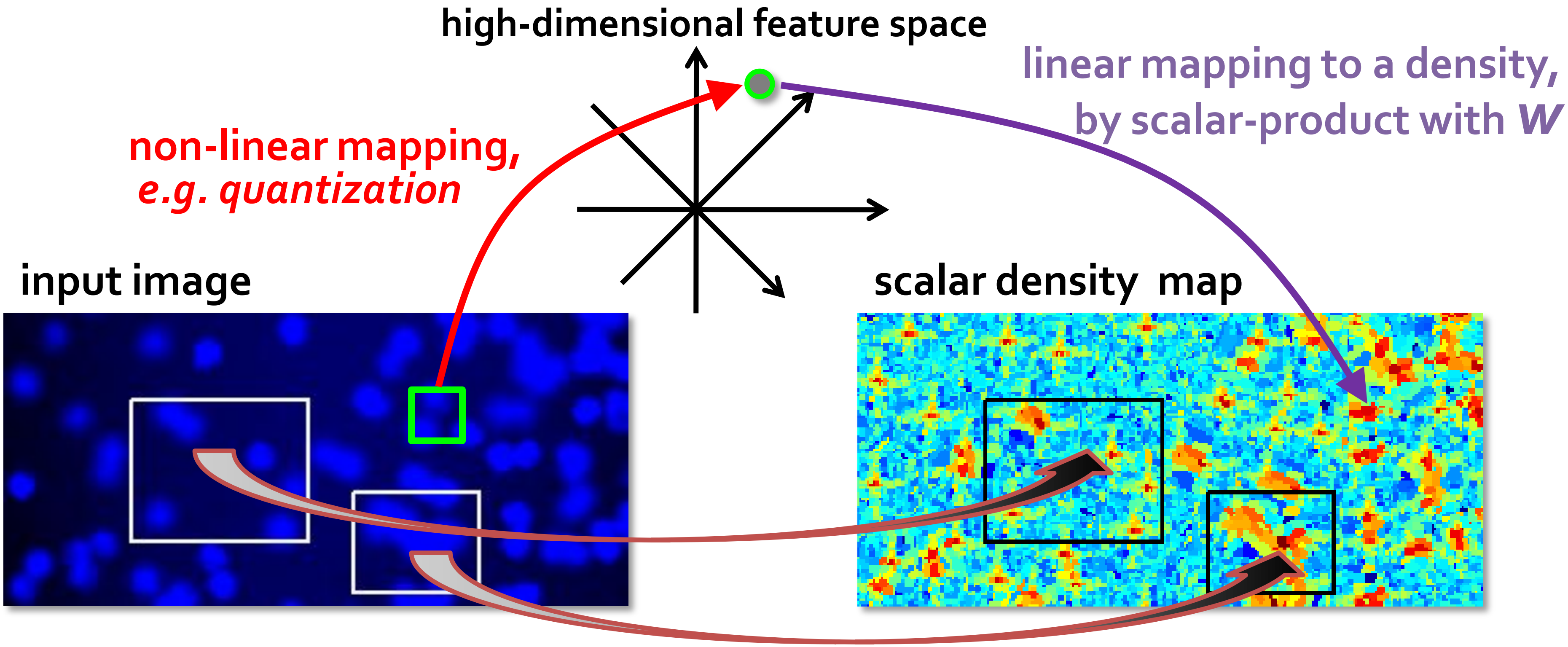


Our method:

- … is trained based on minimalistic *dotted* annotations
- … is trained discriminatively via convex optimization
- … is fast at test time
- … turns each training image into a combinatorial number of training samples

We avoid:

- … detection of individual object instances
- … building and fitting generative models



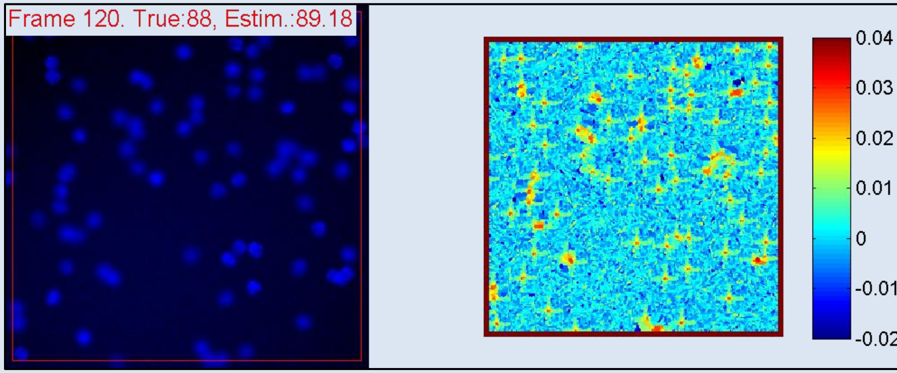
We learn ***w*** by minimizing

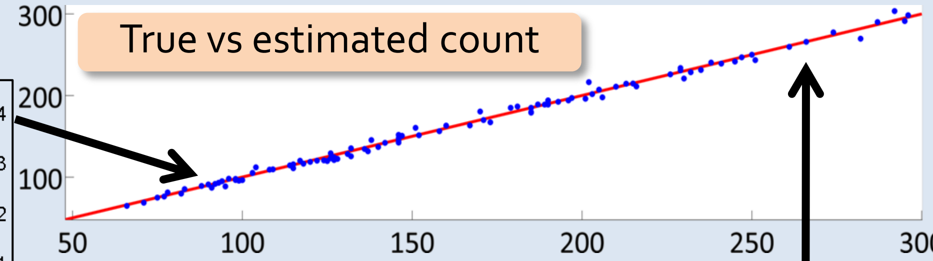
1. ***MESA-loss:*** the **sum** [over training images] of **max** [over <u>all</u> sub-images] of abs. difference between **the true counts** and **the density sums**
2. Regularization (e.g. squared L2-norm) on ***w***

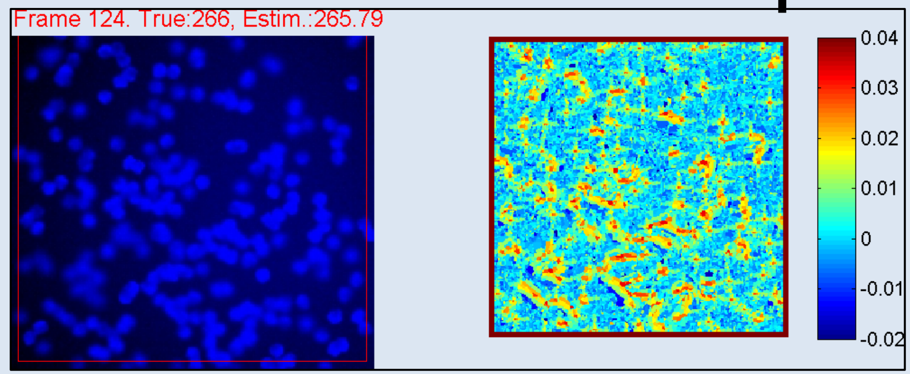
- Minimization can be posed as a **convex quadratic program** (with combinatorial number of constraints)

- Exact and efficient *constraint generation* via **2D-maximum subarray** algorithm

Cell counting

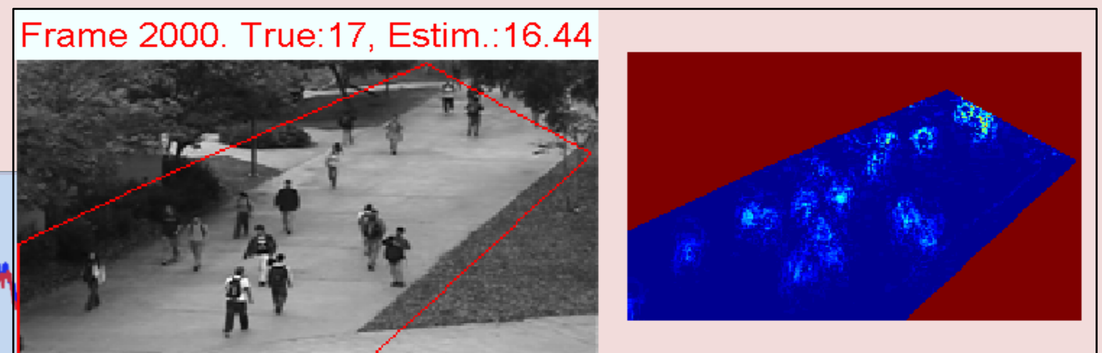






- Extensive comparisons with several baselines
- For all sizes of the training set (1 to 32) our method achieves the lowest counting error

Crowd counting



- Our method is on par with the state-of-the-art method that uses more detailed annotations for training