Feature Mining for Localised Crowd Counting

Queen Mary University of London

School of Electronic Engineering and Computer Science

London E1 4NS, UK

Ke Chen, Chen Change Loy, Shaogang Gong and Tao Xiang

Queen Mary Vision Laboratory

{cory.sgg,txiang}@eecs.qmul.ac.uk
ccloy@visionsemantics.com

1 .: Introduction :.

(a) UCSD Dataset [1] (b) Mall Dataset

Figure 1. Example public scenes of groups of people.

- Crowd counting for public space safety and management;
- Applications include crowd control, public space design, pedestrian behaviour profiling

.: Motivation :.

Limitations of existing techniques:
- Counting-by-detection: Slow due to exhaustive scanning at multi-scales;
- Counting-by-clustering: Requires large quantity of data, e.g. high frame rate;
- Counting-by-regression:
  1. Global models [1]: Spatial information is lost due to using global features;
  2. Multiple local models [2]: Not scalable and missing shared information across spatially correlated regions

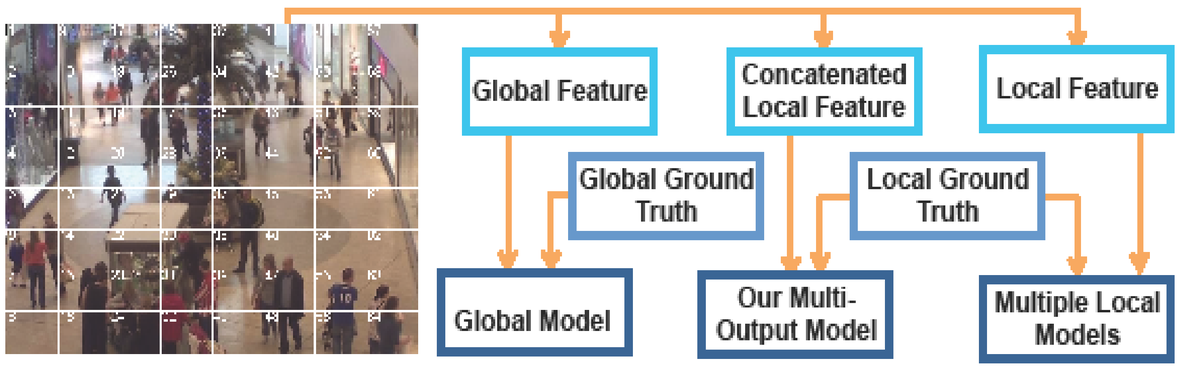

Figure 2. Comparing the processing flow charts of conventional global and local counting-by-regression models and that of our proposed multi-output model.

.: Contribution :.

- Consider local and correlated feature mining for crowd counting;
- Exploit a multi-output ridge regression model for localised crowd counting;
- Provide local estimates with scalability also achieved;
- Release the largest Mall crowd dataset of over 60,000 pedestrian instances (http://www.eecs.qmul.ac.uk/~ccloy/downloads_mall_dataset.html).

3 .: Settings :.

Datasets:

Data	N_f	R	FPS	D	Tp
UCSD	2000	238 × 158	10	11–46	49885
Mall	2000	320 × 240	<2	13–53	62325

Table 1. Dataset properties: $\mathbf{N_f}$ = number of frames, **R** = Resolution, **FPS** = frame per second, **D** = Density (minimum and maximum number of people in the ROI), and **Tp** = total number of pedestrian instances.

Evaluation Metrics:

Three evaluation metrics, namely mean absolute error (mae), ε_{abs} ; mean squared error (mse), ε_{sqr}; and mean deviation error (mde), ε_{dev} were employed.

$$\varepsilon_{\text{abs}} = \frac{1}{N}\sum_{i=1}^{N}|v_i - \hat{v}_i|, \quad \varepsilon_{\text{sqr}} = \frac{1}{N}\sum_{i=1}^{N}(v_i - \hat{v}_i)^2, \quad \text{and} \quad \varepsilon_{\text{dev}} = \frac{1}{N}\sum_{i=1}^{N}\frac{|v_i - \hat{v}_i|}{v_i},$$

where N is the total number of test frames, v_i is the actual count in each cell region or the whole image, and $\hat{v}_i$ is the estimated count of ith frame.

Comparative Evaluation:

- Global model with global feature : (1) ridge regression (RR), (2) Gaussian processes regression [1] (GPR) with linear + RBF kernel.
- Multiple localised regressors (MLR) [2]: Multiple ridge regression models learning the mapping between local feature and corresponding people count.

2 .: Methodology :.

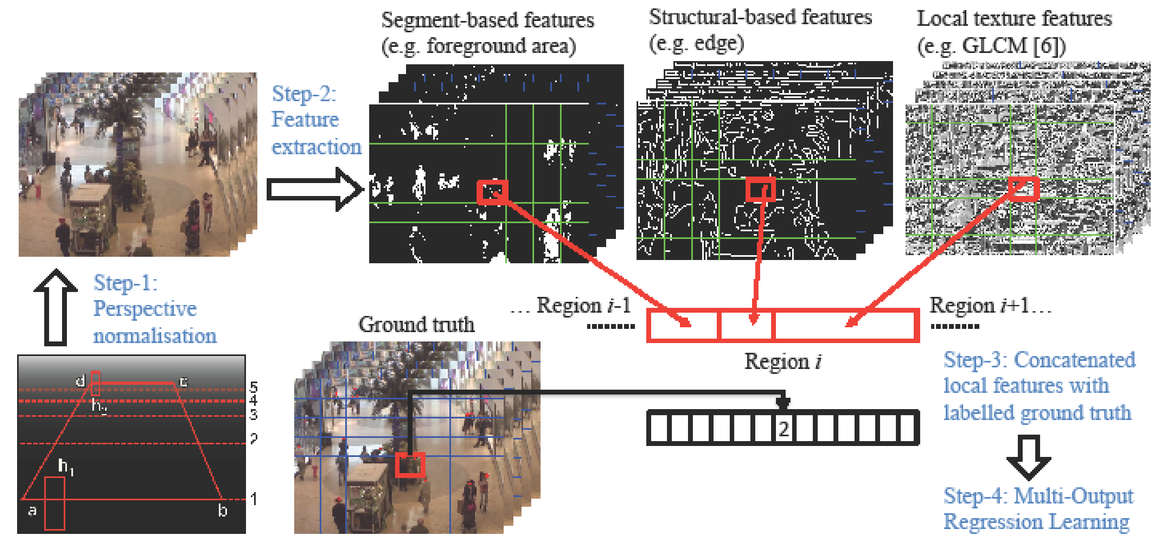

Figure 3. A multi-output regression framework for localised crowd counting by feature mining.

Multivariate Ridge Regression for multi-output regression learning (MORR):

• Given the concatenated intermediate feature vector xi and the concatenated localised labelled ground truth $\mathbf{y}_i$, Multivariate Ridge Regression is presented as

$$\min \quad \frac{1}{2}\|\mathbf{W}\|_F^2 + C\sum_{i=1}^{N}\|\mathbf{y}_i^T - \mathbf{x}_i^T\mathbf{W} - \mathbf{b}\|_F^2$$

where $\mathbf{W} \in \mathbb{R}^{d\times K}$ and $\mathbf{b} \in \mathbb{R}^{1\times K}$ denote a weight matrix and a bias vector.

The importance of weight matrix W
- capture the local feature importance
- facilitate the sharing of features
- jointly weigh features from specific local cell and other cells

4 .: Experiments :.

Method	Features Level		Learning Level		UCSD			Mall		
	Global	Local	Global	Local	mae	mse	mde	mae	mse	mde
RR	✔	–	✔	–	2.25	**7.82**	0.1101	3.59	19.0	0.1109
GPR	✔	–	✔	–	**2.24**	7.97	0.1126	3.72	20.1	0.1159
MLR	–	✔	–	✔	2.60	10.1	0.1249	3.90	23.9	0.1196
MORR	–	✔	✔	–	2.29	8.08	**0.1088**	**3.15**	**15.7**	**0.0986**

Table 2. Performance comparison between different methods and our multi-output ridge regression (MORR) model on global crowd counting.

Figure 4. Using the Mall dataset as a study case: the figures depict the weight contributions of neighbouring cells to cells 11, 51, and 55,which are highlighted using black boxes. Red colour in the heat maps represents higher weight contribution i.e. more information sharing.

	Region 1(R1)		
	mae	mse	mde
MLR	0.82	1.45	0.3611
MORR	**0.76**	**1.22**	**0.3317**

	Region 2 (R2)		
	mae	mse	mde
MLR	0.71	1.24	0.3317
MORR	**0.67**	**1.12**	**0.3061**

	Scalability (seconds)	
	Time-tr	Time-te
MLR	17.274	0.1028
MORR	14.848	0.0196

Figure 5. Localised counting performance on two busy localised regions in the Mall dataset. Region 1 consists of Cells 11, 12, 19, and 20, while Region 2 includes Cells 43, 44, 51, and 52. Time-tr and Time-te denote the training time and testing time respectively

Future work will focus on exploring dynamic and temporal segmentation of crowd structure.

[1] A.B. Chan, Z.-S. J. Liang, and N. Vasconcelos. Privacy preserving crowd monitoring: counting people without people models or tracking. *International Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2008

[2] X.Wu, G. Liang, K.K. Lee, and Y. Xu. Crowd density estimation using texture analysis and learning. *International Conference on Robotics and Biomimetics*, pages 214–219, 2006.