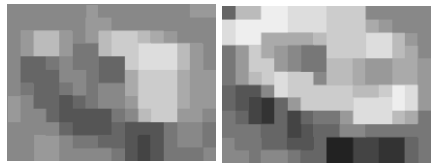
Unsupervised Object Counting without Object Recognition

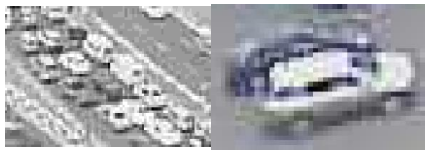
Takayuki Katsuki, Tetsuro Morimura (IBM Research - Tokyo), and
Tsuyoshi Idé (IBM T. J. Watson Research Center)

How many vehicles are there in this road?

Identification of independent objects is often impossible when we handle



Very low resolution



Overlapped

Count objects without recognizing any objects or using any labeled training data

Observation x

~~Recognize objects~~ ~~Count objects~~

Estimate

# of objects

$$\boldsymbol{h} \in \{0,1\}^{\infty}, \quad \sum_{d=0}^{\infty} h_d = 1$$

We use the training data $\boldsymbol{X} \equiv \{x_1, x_2, \ldots, x_N\}$ **without any label information as to the count.**

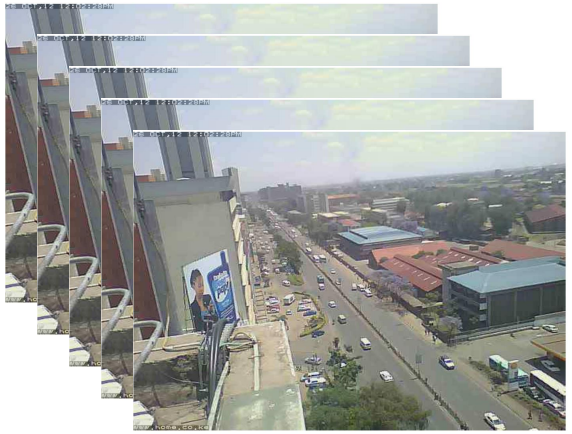



.....

Formalize the problem as a density estimation using a particular type of mixture model

Learning phase

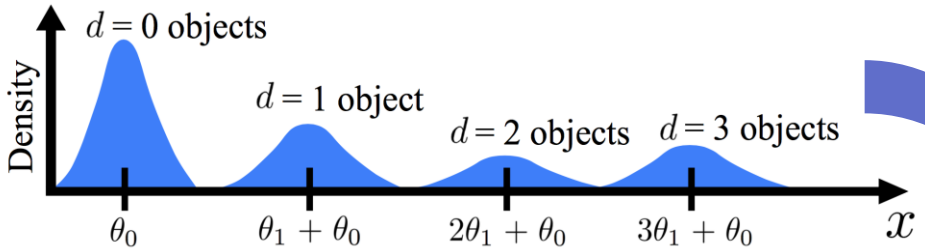
Training data without labels



Feature extraction

x_1
x_2
x_3
x_4
x_5
$\vdots$

Learn a Gaussian mixture model (GMM) whose mixture index is equated with the count of the objects



Our idea is to use the **stick-breaking process** as a constraint to make it possible to interpret the mixture indexes as the count.

Runtime phase

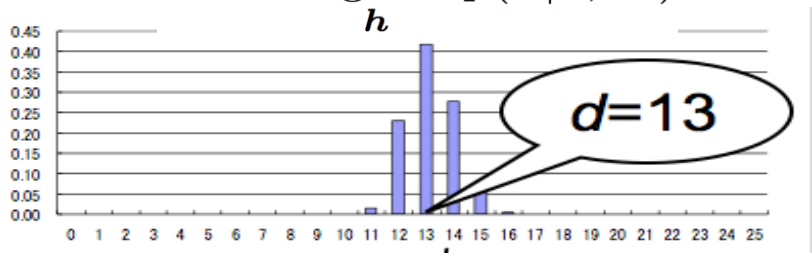
New observation

Feature extraction

x

To find the count for a new observation, we pick the cluster of the highest posterior for the count given x :

$$h^* \equiv \underset{h}{\operatorname{argmax}}\, p(h|x, X)$$

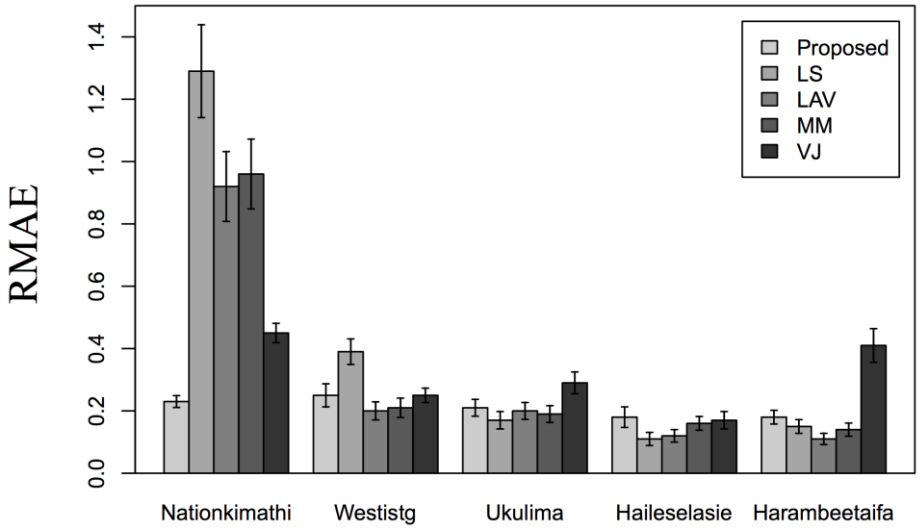


Experimental Results: Vehicle-Counting from Web Camera Images

We confirmed that the proposed framework can apply to the task of counting vehicles in the web camera images.

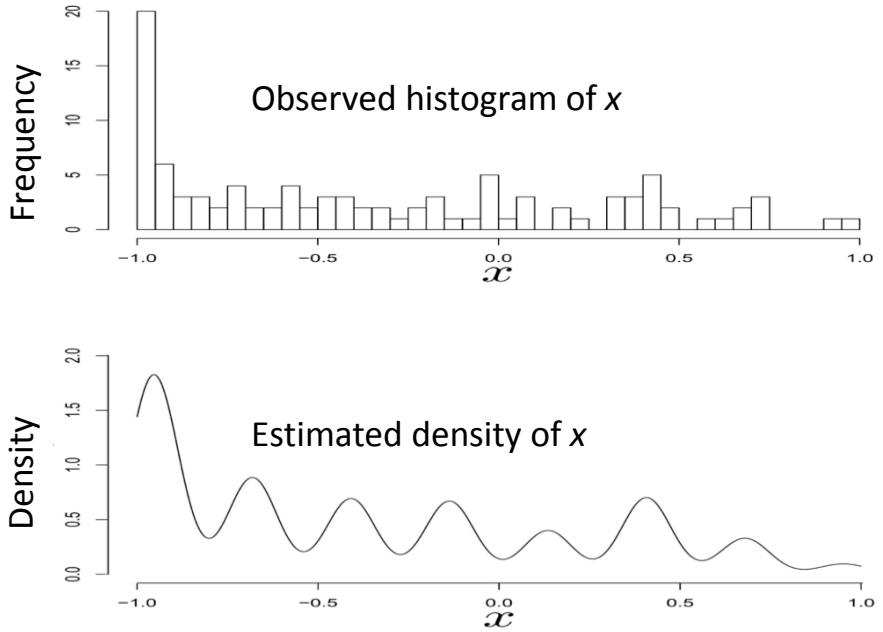
(a) Nationkimathi (b) Westistg (c) Ukulima (d) Haileselasie (e) Harambeetaifa

We compared our unsupervised approach with several *supervised* alternatives in RMAEs for all of the camera locations (smaller is better). Error bars represent the standard error.



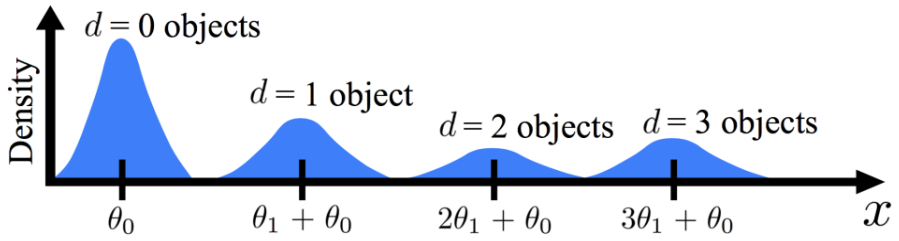
We can see that the overall performance of our method is comparable to or even better than those of the supervised alternatives. This is rather surprising, because our method does *not* use any labeled training data.

Comparison of the estimated $p(x)$ distribution with the true one created from the data.



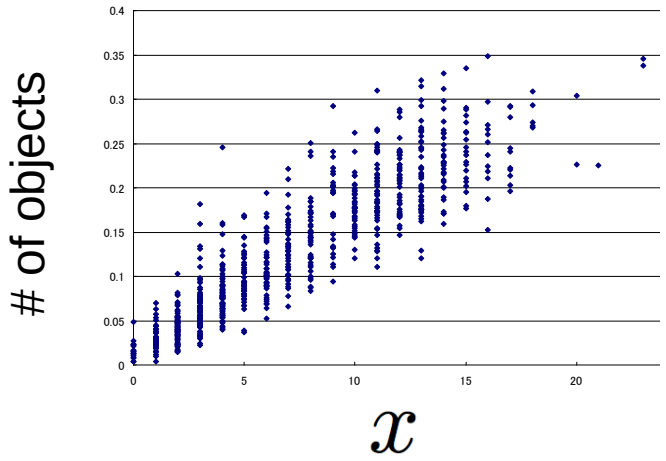
Gaussian Mixture Model whose *d*-th component is responsible for *x* having *d* number of objects

The proposed GMM has **a restriction on its mean parameter**:



$$p(x|\boldsymbol{h}, \boldsymbol{\theta}, \beta) \equiv \prod_{d=0}^{\infty} \mathcal{N}\left(x \middle| \theta_1 d + \theta_0, \beta^{-1}\right)^{h_d}$$
$$= \frac{\exp\left(-\frac{\beta}{2}\sum_{d=0}^{\infty} h_d (x - \theta_1 d - \theta_0)^2\right)}{\left(2\pi\beta^{-1}\right)^{\frac{1}{2}}}$$

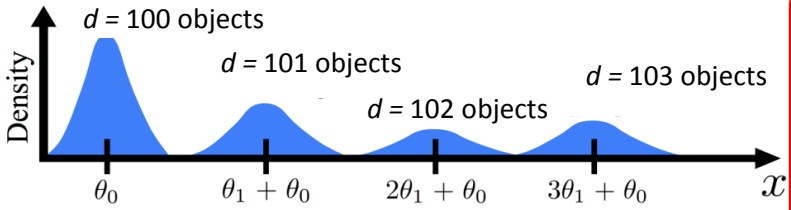
We loosely assume that feature *x* is a good enough feature in the sense that it is proportional to the count.

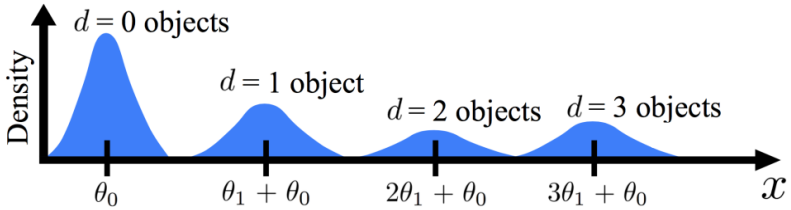


However, the counting results of the proposed GMM without any additional constraint will become ***linearly proportional to the true count.***

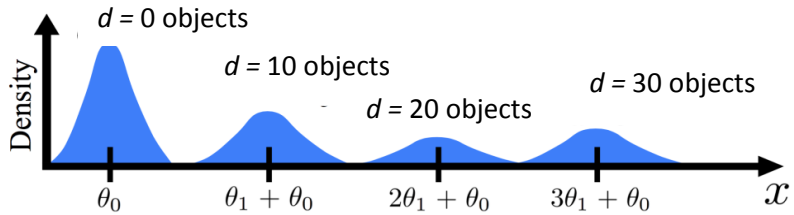
Regularize density estimation results of the GMM by stick-breaking process prior

The likelihood of the count ***h*** in the proposed GMM is invariant with respect to the simultaneous translation of x and ϑ_0, as well as the simultaneous scaling between count d and ϑ_1, such as,





Assumption: the smallest one in the possible count sets has to be the true one



We introduce the stick-breaking process as the prior for the count, which can represent the desired property of the count set: the assigned count values for the observations are consecutive natural numbers from zero.

Stick-breaking Process $$p(\boldsymbol{h}|\boldsymbol{v}) \equiv \prod_{d=0}^{\infty} \left(v_d \prod_{k=0}^{d-1} (1 - v_k) \right)^{h_d}$$

Interestingly, in traditional Bayesian nonparametric literature, this nature, which is each component is always associated with the same index, is known as a drawback; that is, it can cause the solution to get stuck at a local minimum in practical use.

Estimate posterior through efficient variational Inference algorithm

Using the proposed model and the conjugate priors, we can rewrite the posterior as

$$p(\boldsymbol{h}|x,\boldsymbol{X}) \propto \int p(x,\boldsymbol{h},\boldsymbol{X},\boldsymbol{H},\boldsymbol{\theta},\beta,\boldsymbol{v})\mathrm{d}\boldsymbol{H}\mathrm{d}\boldsymbol{\theta}\mathrm{d}\beta\mathrm{d}\boldsymbol{v},$$

$$\text{where}\quad p(x,\boldsymbol{h},\boldsymbol{X},\boldsymbol{H},\boldsymbol{\theta},\beta,\boldsymbol{v}) \equiv p(x|\boldsymbol{h},\boldsymbol{\theta},\beta)$$

$$\times p(\boldsymbol{h}|\boldsymbol{v})\left[\prod_{n=1}^{N} p(x_n|\boldsymbol{h}_n,\boldsymbol{\theta},\beta)p(\boldsymbol{h}_n|\boldsymbol{v})\right]p(\boldsymbol{\theta})p(\beta)p(\boldsymbol{v})$$

We approximates the posterior in a factorized form:

$$q(\boldsymbol{h},\boldsymbol{H},\boldsymbol{\theta},\beta,\boldsymbol{v}) \equiv q(\boldsymbol{h},\boldsymbol{H})q(\boldsymbol{\theta})q(\beta,\boldsymbol{v})$$

We then identify the optimal *q* that minimizes

$$D_{\mathrm{KL}}(q\|p) \equiv \int q(\ln q - \ln p)\mathrm{d}\boldsymbol{H}\mathrm{d}\boldsymbol{\theta}\mathrm{d}\beta\mathrm{d}\boldsymbol{v}$$

Finally, thanks to the conjugate modeling, we can get the iterative updating equations as

$$q(\boldsymbol{h})q(\boldsymbol{H}) = \mathrm{Categorical}\left(\boldsymbol{h}\middle|\boldsymbol{\mu}_{\boldsymbol{h}}\right)$$

$$\times \prod_{n=1}^{N} \mathrm{Categorical}\left(\boldsymbol{h}_n\middle|\boldsymbol{\mu}_{\boldsymbol{h}_n}\right),$$

$$q(\boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{\theta}\middle|\boldsymbol{\mu}_{\boldsymbol{\theta}},\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right),\quad \text{and}$$

$$q(\beta,\boldsymbol{v}) = \mathrm{Gamma}\left(\beta\middle|a_\beta,b_\beta\right)\mathrm{Beta}\left(v_d\middle|a_{v_d},b_{v_d}\right),$$

We stop the VB iterations when this condition is satisfied

$$\frac{\left(D_{\mathrm{KL}}(q\|p) - D_{\mathrm{KL}}(q'\|p)\right)^2}{D_{\mathrm{KL}}(q'\|p)^2} < 10^{-10}$$

We can estimate the number of objects in the new observation as

$$\boldsymbol{h}^* \simeq \underset{\boldsymbol{h}}{\mathrm{argmax}}\, q(\boldsymbol{h})$$

Future work

- Including many other features and introducing a non-linear relationship in the proposed GMM would be an important research area

- Applying the proposed approach to other applications, such as crowd counting and cell counting, would be another promising area of study