# Normalized Autobinomial Markov Channels For Pedestrian Detection

Cosmin Ţoca
cosmin.toca@gmail.com
Mihai Ciuc
mihai.ciuc@upb.ro
Carmen Pătraşcu
cpatrascu@ceospacetech.pub.ro

University Politehnica of Bucharest, Faculty of Electronics, Telecommunications and Information Technology, Applied Electronics and Information Engineering Department, RO

Pedestrian detection represents one of the most important components of engineering devices that use automated vision to help decision systems take quick and accurate actions. Such systems are defined and customized to be useful for different needs, such as monitoring and aided surveillance, or increasing safety features in automotive industry. Given the large spectrum of applications that use pedestrian detection, demand has increased in recent years for the development of feasible solutions which can be integrated in devices such as smartphones or action cameras.

This paper focuses on finding probabilistic features that highlight the human body characteristics regardless of contextual information in images. Adjacent pixels are often spatially correlated, which means that they are likely to have similar values. We view the image as a collection of random variables indexed by certain locations, called sites. The state of a site $\xi$ is conditionally independent of all variables in the random field, except the neighbouring system $\mathcal{N}_\xi = \{\eta \in \Omega \mid \eta \neq \xi, \mathrm{d}^2(\xi,\eta) \leq \Delta\}$, where $\Delta$ is a positive integer and $\mathrm{d}^2(\xi,\eta)$ is the squared Euclidean distance between $\xi$ and $\eta$. The neighbouring system strictly depends on a collection of cliques $\mathcal{C} = \sum_{k=1}^{\omega(\delta)} \mathcal{C}_k$, where $\omega(\Delta)$ is the number of cliques for each local specification.

**Energy function:** An unpublished manuscript [2] describes how to interpret the local property of a Markov random field in terms of energy and potential, claiming that the probability at a site $\xi$ is given by:

$$\pi^\xi(\gamma) = \frac{\mathrm{e}^{-\sum_{C \ni \xi} V_C(\gamma)}}{\sum_{\varphi \in \mathcal{F}} \mathrm{e}^{-\sum_{C \ni \xi} V_C(\varphi, \gamma_{\Omega \setminus \xi})}}, \tag{1}$$

where $V_C$ is the potential function, and $\varphi \in \mathcal{F}$. To get the probability that at a site $\xi$ the state is $\gamma$, we need to define a potential function $V_C(\gamma)$ in the neighbouring system, here denoted by a collection of cliques $\mathcal{C}$. To be able to do this, we refer to the auto-binomial model that was introduced by Besag [1] to describe types of spatial processes, examining some stochastic models that occur in the texture of various physical materials.

**Potential function:** The potential at a certain state is given by:

$$V_C(\gamma) = \begin{cases} -\ln \binom{\Gamma}{\gamma_\xi} + \gamma_\xi & \text{if } \mathcal{C} = \{\xi\} \\ \dfrac{\gamma_\xi \cdot \gamma_\eta}{\nu} & \text{if } \mathcal{C} = \{\xi, \eta\} \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $\nu$ is a normalization constant. If we replace Eq. 2 in Eq. 1, we get the probability assigned to a local system, and namely Eq. 3.

**Feature calculation:** The normalized autobinomial Markov channels are given by the probability that at a site $\xi$ the state is $\gamma$:

$$\pi^\xi(\gamma) = Z^{-1} \binom{\Gamma}{\gamma_\xi} \sigma^{\gamma_\xi} (1-\sigma)^{\Gamma - \gamma_\xi}, \tag{3}$$

where $Z = \binom{\Gamma}{\Gamma/2}$ is a normalization constant for the binomial distribution, and $\sigma = \sigma(\mathcal{N}_\xi) = (\mathrm{e}^{\langle \alpha, \beta \rangle})/(1+\mathrm{e}^{\langle \alpha, \beta \rangle})$. The scalar product $\langle \alpha, \beta \rangle$ of the two vectors of size $\omega(\delta)$ sums up and weighs (with $\alpha$) the absolute difference of gray levels $\beta_k = \beta_k(\gamma_{\mathcal{N}_\xi}) = |\gamma_\eta - \gamma_{\eta'}|$. The only sites participating are pairs of cliques found in the same neighbourhood of the analyzed pixel. In other words $\eta \neq \eta'$, and $\eta, \eta' \in \mathcal{N}_\xi$, where $\{\eta, \xi\}$ and $\{\eta', \xi\}$ represent two pairs in $\mathcal{C}_k$ containing $\xi$.

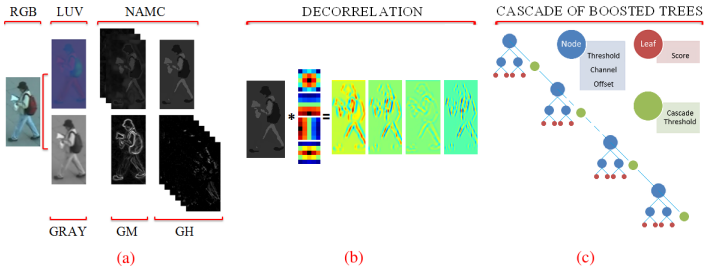


Figure 1: **a) Feature representation.** Several channels (NAMC-normalized autobinomial Markov channels, GM-gradient magnitude and GHs-gradient histograms) have been computed using different transformations of the input image. **b) Local decorrelation.** Multiple decorrelation filters have been learned separately for each channel. **c) Cascade of boosted trees.** The classification is based on a cascade of boosted trees, where each node examines a feature value in a certain channel and each leaf returns a score.

The main advantage of using the normalized autobinomial Markov channels as feature descriptor comes from the property of randomly selecting pixels to be part of the neighbourhood system, having a significant contribution for pedestrian detection in noisy scenarios. Another benefit is given by the fact that it shows several possibilities of optimizations by turning many of the computations into memory accesses.

This paper introduces the normalized autobinomial Markov channels and proves that the general idea of feature scaling is applicable. Moreover, it uses a method of feature decorrelation [3] to substitute the need for oblique splits, and shows that such a cascade of boosted trees [4] outperforms the majority of the existing features and methods for pedestrian detection. The results shown in Fig. 2.a and Fig. 2.b demonstrate the efficiency of our approach against the tested state-of-the art solutions.

[1] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, 1974.

[2] J. M. Hammersley and P. Clifford. Markov field on finite graphs and lattices. *Unpublished*, 1971.

[3] W. Nam, P. Dollar, and J.H. Han. Local decorrelation for improved pedestrian detection. *NIPS*, 2014.

[4] Z. Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. *ICCV*, 2005.

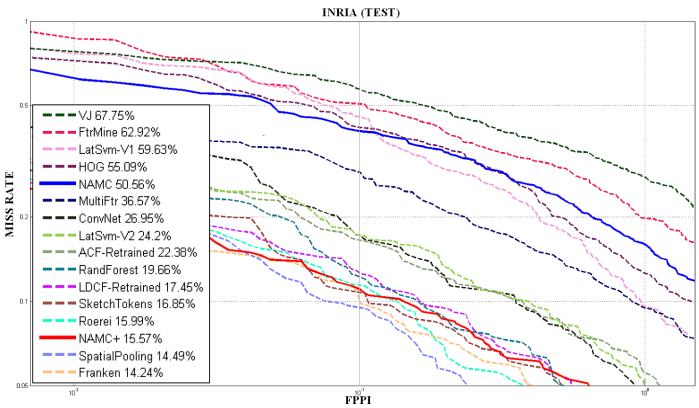


Figure 2.a: **Results on INRIA Pedestrian Dataset.**

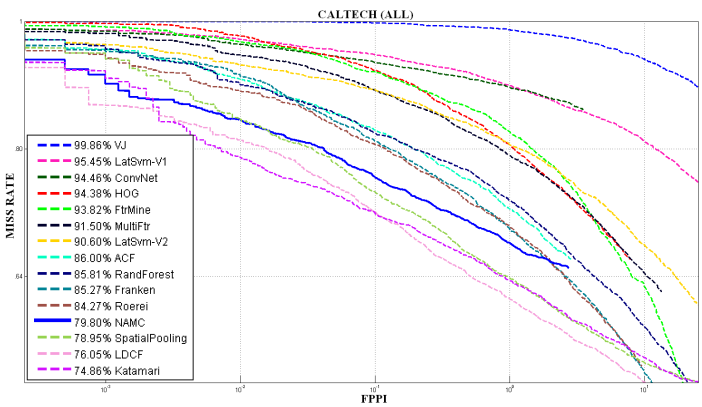


Figure 2.b: **Results on CALTECH Pedestrian Dataset.**