Rejection-Cascade of Gaussians: Real-time adaptive background subtraction framework

B Ravi Kiran[1], Arindam Das[2], Senthil Yogamani[3]
ravi.kiran@navya.tech, {arindam.das,senthil.yogamani}@valeo.com
Navya, Paris, France[1], Detection Vision Systems, Valeo India[2], Valeo Vision Systems, Galway, Ireland[3]

NCVPRIPG 2019
22-24 DECEMBER
KLE TECHNOLOGICAL UNIVERSITY

Problem Definition & Model

- **Background Subtraction** :
  - **Inputs** : Video stream containing static and dynamic backgrounds
  - **Output** : Binary classification problem per pixel b/w foreground/background classes.
  - **Model** : Gaussian Mixture Models (GMM) are parametric models used to estimate the background class at each pixel of the input image.
  - **Contribution** : Decomposition of GMM into Adaptive Rejection Cascade of binary classifiers using strong prior information. The classifiers are ordered by the negative class rejection rate following the **Viola-Jones rejection cascade**, as well as increasing computational complexity.
- Cascade of Gaussians (CoG) Framework
  - CoG constitutes of k+1 binary classifiers :
    - Consistent Hypothesis Propagation (CHP) classifier : Propagates previous time's class (FG/BG) if the value has not changed.
    - 1st dominant Gaussian $\omega_0.\eta(\mu_0, \sigma_0)$
    - 2nd dominant Gaussian $\omega_1.\eta(\mu_1, \sigma_1)$
    - kth dominant Gaussian $\omega_k.\eta(\mu_k, \sigma_k)$

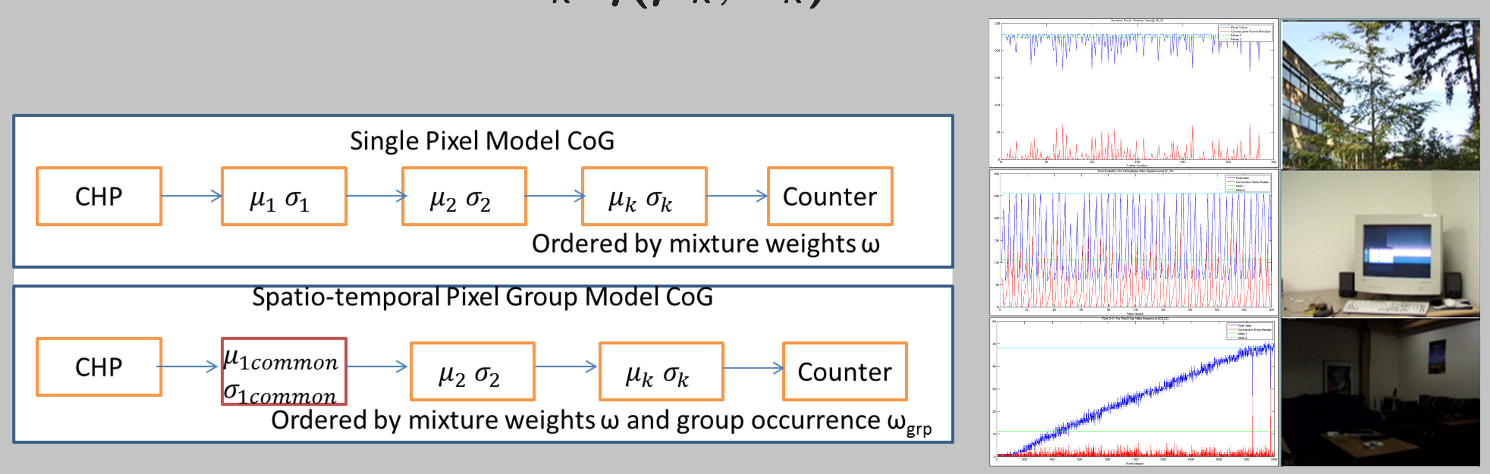

Figure 1:**Right** : Elements of CoG : CHP, first and second modes of gaussians and spatio-temporal window of a Cascade of Gaussians. **Left** : Different dynamics of a pixel : Dynamic Pixel Vs Oscillation Vs Pixel Drift.

CoG : Online parameter update

1. Get N frames & estimate pixel-wise $\mu(t), \sigma(t), \omega(t)$
2. Form matrix whose rows are adapted variance and ranked weight observations, while columns are variables V and R, $V(t_k, i) = I(t_k), k = 1 : N$
3. Obtain covariance matrices $R_{cov} = Cov(R), V_{cov} = Cov(V)$
4. Perform K-means clustering with K=3 (for temporal pixel residue due to dynamic, oscillating, or drifting BG).
5. Threshold for pixels within $0.7 - 0.5\sigma$
6. Calculate the KDE of given cluster & the joint occurrence distribution and associated weight ω_1, μ_1 and σ_1

Computational analysis of Rejection-Cascade of Gaussians

- **Average Speedup** : Over single Image I with N pixels

$$\frac{N}{\sum_i s_i n_i} \tag{1}$$

- n_i refers the ratio of background pixels labeled mean or mean with variance w.r.t the total number of background pixels in the image,
- s_i is the normalized ratio of the time it takes for cascade level i BG classifier model to evaluate and label a pixel as background.
- The values of n and s were profiled over various videos for different durations.

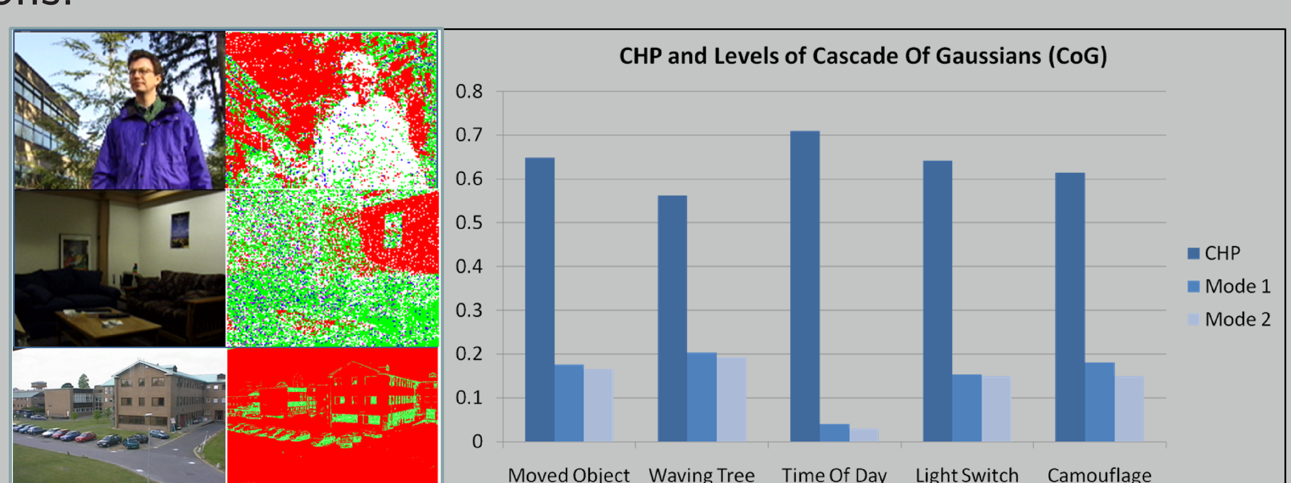

Figure 2:Left: Pixels in CHP(red), Mode 1(green), Mode 2(blue), Mode 3(violet) and Foreground(white). Right: Normalized pixel count over elements of Cascade of Gaussians CHP, first and Second modes of Gaussians.

Deep Rejection Cascade for VAEs

- **Variational Autoencoder (VAE)** : are generative models that approximate the data distribution $P(X)$ of a high dimensional input X, an image or video.

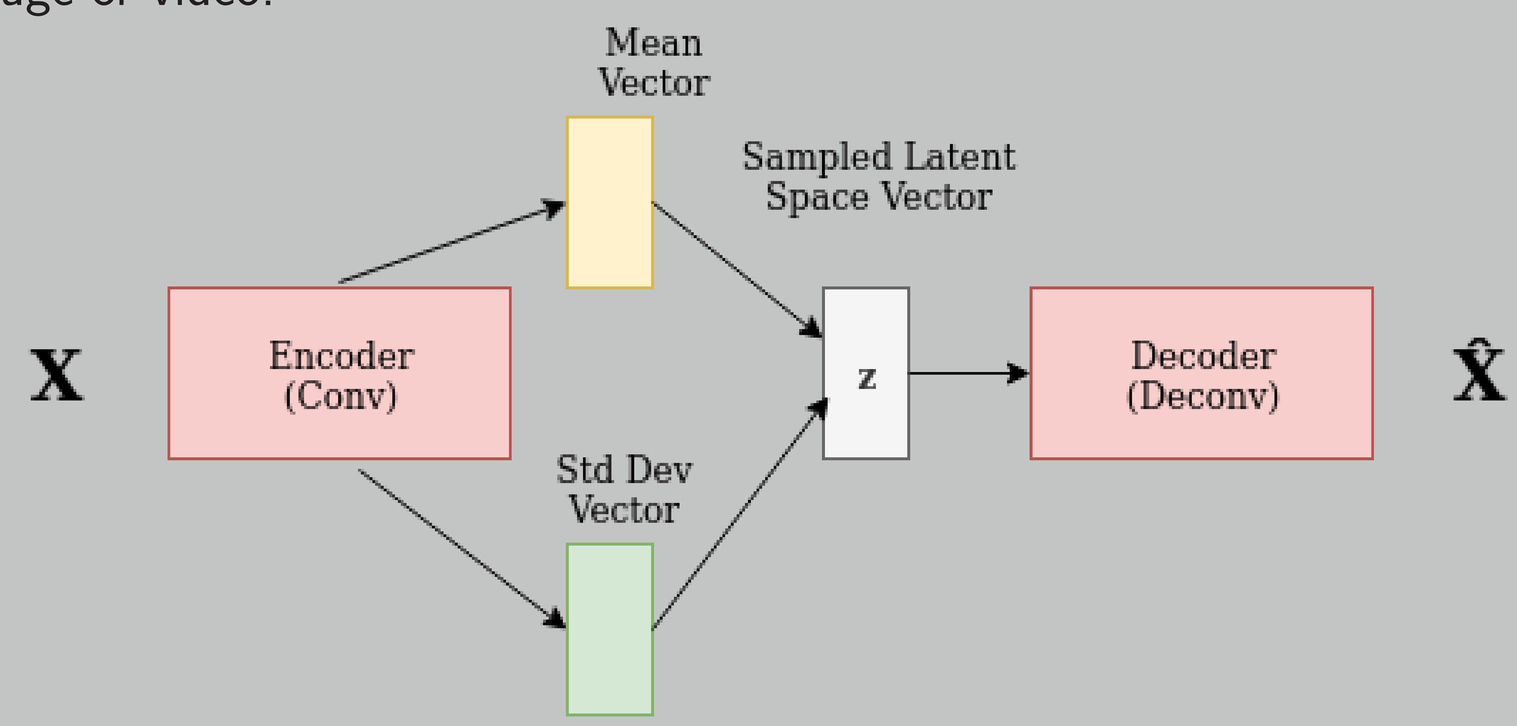

Figure 3:A VAE represents a variational approximation of the latent space with an autoencoder architecture, with a probabilistic encoder $q_\phi(x|z)$ that produces Gaussian distribution in the latent space z (represented by mean and standard deviation vectors), and a probabilistic decoder $p_\theta(z|x)$, which given a code produces distribution over the input space. Loss function : Reconstruction error + KL Divergence between training data latent space vector distribution & standard normal.

- **Deep Rejection Cascade over VAEs** : A Rejection cascade decomposition of the VAE can be achieved. The pixel-level tests in CoG are now performed by the VAE in the latent space.

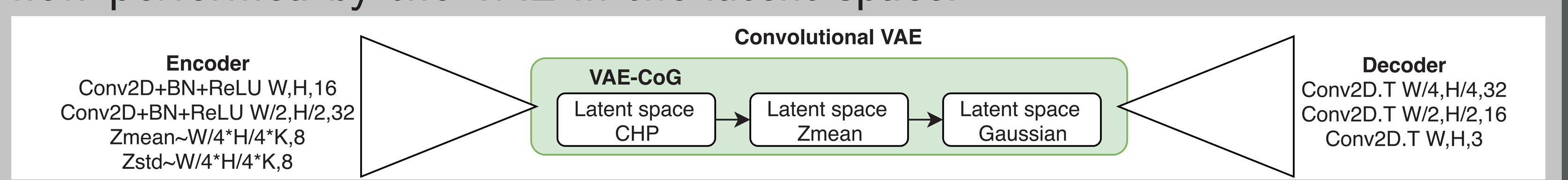

Figure 4:VAE-CoG on the latent space representation of a VAE. Filters are all 3x3. A convolutional VAE with latent space of 16 dimensions was trained on the CDW-2014 datasets.

- Invariance to positions, orientations, pixel level perturbations, and deformations due to convolutional architecture.

Figure 5:The input-output pairs and absolute value of residue between input-output pairs from a Convolutional VAE : top half without foreground bottom half with foreground. We remark that the dynamic background such as the snow has been removed. The right column demonstrates the 2d-Histogram over the latent space **z** of the CVAE (top) and the histogram over the temporal residue over **z** for the same test sequence.

- **Experiments and Analysis : VAE-COG**
  - The VAE is trained on frames with dynamic and static background to estimate the normal and standard deviation vectors.
  - The test samples are reconstructed and residue w.r.t input scaled by training error standard deviation is used as the output for background subtraction.
  - The Rejection Cascade elements : CHP and 1st level Gaussian frequency are measured over the latent space for the current video in CDW-2014 dataset. The plot demonstrates many images with dynamic BG are compressed and mapped to the same latent space vector for the CHP case.

Conclusions

- The CoG was evaluated on the wallflower dataset. We observed a speedup of 4-5x, over the baseline GMM, with an average improvement of 17% in the mis-classification rate.
- The VAE-CoG was evaluated on the CDW-2014 datasets, providing a first estimate in the speedup: CHP requires memory to store previous encoded latent space vector and output FG/BG image, while providing speedup by avoiding the VAE-COG's Decoding into output domain. A speedup can be achieved with the Gaussian test though this is not trivial.