ENTROPY-BASED APPROACH FOR DETECTING FEATURE RELIABILITY

Dragoljub Pokrajac, *Delaware State University dpokraja@desu.edu*
Longin Jan Latecki,*Temple University latecki@temple.edu*

Abstract – *Although a tremendous effort has been made to perform a reliable analysis of images and videos in the past fifty years, the reality is that one cannot rely in 100% on the analysis results. In this paper, rather than proposing yet another improvement in video analysis techniques, we discuss entropy-based monitoring of features reliability. Major assumption of our approach is that the noise, adversingly affecting the observed feature values, is Gaussian and that the distribution of noised feature approaches the normal distribution with higher magnitudes of noise. In this paper, we consider one-dimensional features and compare two different ways of differential entropy estimation—histogram based and the parametric based estimation. We demonstrate that the parametric approach is superior and applicable to identify time intervals of a test video where the observed features are not reliable for motion detection tasks.*

1. INTRODUCTION

The main effort in video analysis nowadays is still in making the feature computation more reliable. Our claim is that in spite of all such efforts we need to accept the fact that the computed features will never be absolutely reliable, and focus our attention also on computing reliability measures. This way system decisions will only be made when features are sufficiently reliable. For an intelligent system for video analysis, as supplement to feature computation, we propose instantaneous evaluation of computed features accompanied by assessing their reliability, and adapting the system behavior in accordance to features reliability. For example, if the goal of a system is to monitor motion activity, and to signal an alarm if the activity is high, the system can make reliable decisions only if there exist a subset of the computed motion activity features that is sufficiently reliable. The monitoring of features reliability and adjusting the system behavior accordingly therefore seems to be the only way to deploy autonomous video surveillance systems in the close future.

A good overview of the existing approaches to motion detection can be found in the collection of papers edited by Remagnino et al. [1] and in the special section on video surveillance in IEEE PAMI edited by Collins et al. [2]. A common feature of the existing approaches for moving objects detection is the fact that they are pixel based. Some of the approaches rely on comparison of color or intensities of pixels in the incoming video frame to a reference image. Jain et al. [3] use simple intensity comparison to reference images so that the values above a given threshold identify the pixels of moving objects. A large class of approaches is based on appropriate statistics of color or gray values over time at each pixel location. (e.g., the segmentation by background subtraction in W4 [4], eigenbackground subtraction [5], etc). Wren et al. [6] were the first who used a statistical model of the background instead of a reference image.

One of the most successful approaches for motion detection was introduced by Stauffer and Grimson [7]. It is based on adaptive Gaussian mixture model of the color values distribution over time at each pixel location. Each Gaussian function in the mixture is defined by its prior probability, mean and a covariance matrix. In [8], we applied the Stauffer-Grimson technique on spatial-temporal blocks and demonstrate the robustness of the modified technique in comparison to original pixel-based approach.

In [9] we have shown that the motion detection algorithms based on local variation are not only a much simpler but also can be a more adequate model for motion detection for infrared videos. This technique can significantly reduce the processing time in comparison to the Gaussian mixture model, due to smaller complexity of the local variation computation, thus making the real time processing of high-resolution videos as well as efficient analysis of large-scale video data viable. Moreover, the local-variation based algorithm remains stable with higher dimensions of input data, which is not necessarily the case for an EM type algorithm (used for Gaussian model estimation).

2. ENTROPY-BASED ASSESMENT OF FEATURE RELIABILITY

To determine whether a particular feature is reliable, we assume that the feature bears more information if its distribution differs more significantly from a normal (Gaussian) distribution. Obsereve that similar heuristics are used e.g., in Independent Components Analysis [10]. The follow-up assumption is that the feature becomes unreliable if an addition random noise is superimposed, which would lead the distribution of such noisy feature to become more Gaussian like. Hence, by measuring to what extent a feature distribution differs from a Gaussian distribution, one can not only get information to what extent the feature is useful but also when such usefulness drops (e.g., due to some external and often non-observed factor).

The Information Theory proposes *negentropy* as the measure of this discrepancy. Given a probability density *p(x)* of a feature, Differential Entropy is defined [11, 12] as:

$$H(x) = \int_{-\infty}^{\infty} -p(x)\log p(x)dx \tag{1}$$

For a given class of distributions *p(x)* that have the same variance σ^2, differential entropy is maximal for a Gaussian distribution where it is equal to:

$$H_{Gauss}(\sigma^2) = \frac{1}{2}\log 2\pi\, e\sigma^2 . \tag{2}$$

Hence, a negentropy, which is defined as

$$J(x) = H_{Gauss}(\sigma^2) - H(x) \tag{3}$$

or its normalized value

$$J(x)/H_{Gauss}(\sigma^2) = 1 - \frac{H(x)}{H_{Gauss}(\sigma^2)} \tag{4}$$

may be used to measure usefulness and reliability of a feature. Observe that the minimal value of negentropy is 0

(when $p(x)$ is Gaussian). Differential entropy and negentropy can be generalized for multidimensional variables. However, in this paper, we consider only one-dimensional feature x.

2.1 Computation of Differential Entropy Based on Piecewise Approximation

Let's approximate $p(x)$ with piecewise linear function $p'(x^*)$ such that:

$$p'(x^*) = K\,p(x_i), \text{for } x^* \in [x_i, x_i + \Delta x], i = 0,\ldots N_{BINS} - 1 \quad \text{where } x_{N_{BINS}} \equiv x_{N_{BINS}-1} + \Delta x. \tag{5}$$

Here, K is normalization constant chosen such that $p'(x^*)$ is distribution, i.e. $\int_{-\infty}^{\infty} p'(x^*)dx^* = 1$. Hence, we choose K such that $K = 1/\Delta x \sum_i p(x_i)$.

Let's denote $p_i \equiv Kp(x_i)$ such that $p'(x^*) = p_i, \text{for } x^* \in [x_i, x_i + \Delta x]$. The differential entropy can be approximated as:

$$\begin{aligned} H(x^*) &= \int_{-\infty}^{\infty} - p'(x^*)\log p'(x^*)dx \\ &= -\sum_i \int_{x_i}^{x_i+\Delta x} p_i \log p_i dx^* = -\sum_i p_i \log p_i \Delta x \end{aligned} \tag{6}$$

Let's now compute the variance $\hat{\sigma}^{*2}$ of the discretized variable x^* (corresponding to the distribution $p'(x^*)$)
The variance is:

$$\hat{\sigma}^{*2} = E\left(x^{*2}\right) - E\left(x^*\right)^2 \tag{7}$$

Where:

$$E(x^*) = \Delta x \sum_i p_i x_i + \frac{\Delta x}{2} \tag{8a}$$

$$E\left(x^{*2}\right) = \Delta x \sum_i p_i x_i^{\,2} + \Delta x^2 \sum_i p_i x_i + \frac{\Delta x^2}{3} \tag{8b}$$

The differential entropy of a Gaussian variable that has the same variance as x^* is (see eq. (2))

$$H_{gauss}\left(\hat{\sigma}^{*2}\right) = \frac{1}{2}\log 2\pi\, e\hat{\sigma}^{*2} \tag{9}$$

Using eq. (6) and eq. (9) we will approximate $J(x)$, eq. (3), with the negentopy $J(x^*)$ of discretized variable x^*:

$$J(x^*) = H_{gauss}\left(\hat{\sigma}^{*2}\right) - H(x^*) \tag{10}$$

Similar, we can approximate normalized negentropy, eq. (4), as:

$$J(x^*)/H_{gauss}\left(\hat{\sigma}^{*2}\right) = \frac{H_{gauss}\left(\hat{\sigma}^{*2}\right) - H(x^*)}{H_{gauss}\left(\hat{\sigma}^{*2}\right)}. \tag{11}$$

2.2 Computation of Entropy Based on Parametric Approximation

Entropy estimation based on histograms is a non-parametric method and hence suffers from drawbacks characteristic for this class of methods [13]. An alternative is to use parametric approach, such as that suggested in [11]. The main ideas of this approach are:

1) Instead of original feature x, use a standardized feature $x^*=(x\text{-}mean(x))/\text{std}(x)$ that have zero mean and unit standard deviation. This way, we could directly use negentropy to compare reliability with no need to normalize with the entropy of a Gaussian.

2) Use a first-order Taylor approximation of a logarithmic function in eq. (1) that leads to: $(1+\varepsilon)\log(1+\varepsilon) \approx \varepsilon + \varepsilon^2/2$;

3) Use conveniently chosen set of orthogonal functions of a feature x (denoted by $G_i(x)$) to expand probability density function $p(x)$ in vicinity of a Gaussian probability density.

In practice, the choice of orthogonal functions is based on practicability and sensitivity on outliers of the computation of estimates for expectations $E(G_i(x))$, integrability of the obtained probability density function approximation and last, but not the least, the properties of non-Gaussian distributions we want to capture.

Based on such consideration, [11] proposes the following two approximations of negentropy, that we use in this paper:

$$J_a(x^*) = k_1 E\left(x^* \cdot e^{-x^{*2}/2}\right)^2 + k_{2a}\left(E(|x^*|) - \sqrt{\frac{2}{\pi}}\right)^2 \tag{12a}$$

$$J_b(x^*) = k_1 E\left(x^* \cdot e^{-x^{*2}/2}\right)^2 + k_{2b}\left(E\left(e^{-x^{*2}/2}\right) - \sqrt{\frac{1}{2}}\right)^2 \tag{12b}$$

Here the coefficients are determined as:

$$k_1 = \frac{36}{8\sqrt{3}-9}, k_{2a} = \frac{24}{2-\frac{6}{\pi}}, k_{2b} = \frac{24}{16\sqrt{3}-27}. \tag{13}$$

3. EXPERIMENTAL RESULTS

We evaluated the proposed techniques for assessing feature reliability on a set of videos [14]. This set includes infrared videos, for which the same settings of parameters as for visual light videos were used. Here we focus on our results on a video sequence from the Performance Evaluation of Tracking and Surveillance (PETS) repository referred to as the *Campus 3* sequence [15].

We consider the reliability of a feature called *motion activity*, computed as follows [9]. First, we represent videos as three-dimensional (3D) arrays of gray level or monochromatic infrared pixel values. We divide each image in a video sequence into disjoint 8x8 squares and generate spatiotemporal (3D) blocks by three adjacent squares in consecutive frames at the same video plane location. The obtained 3D blocks are represented as 192-dimensional vectors (8*8*3) of gray level or monochromatic infrared pixel values. We then zero mean these vectors and project them to three dimensions using principal component analysis (PCA) [16]. For each block location (x,y), we consider set of vectors of the PCA projections corresponding to a symmetric window of 7 frames around the temporal instant t, and for each set, we compute the covariance matrix of vectors. Finally, we compute *motion activity* as the maximal eigenvalue of the covariance matrix.

We perform feature reliability assessment on a frame base. For each frame of the video, we estimate normalized negentropy and evaluate its applicability as a reliability measure. We compared histogram-based (Section 2.1) and parametric (Section 2.2) approach to estimate the normalized negentropy.

Two main issues in using histogram approach for negentropy estimation (Section 2.1) are: 1) to determine the number N_{BINS} of histogram bins 2) To specify the lower and upper bounds

(x_0 and x_{Nbins}) of the discretization interval. In our experiments, we set the discretization interval bounds as the minimal and maximal value of the feature observed. Alternatively, we could use e.g., 6-sigma interval with the estimated feature standard deviation.

The optimal number of histogram bins depends on the probability density function and number of samples. To determine optimal number of bins for histogram approximation of differential entropy, we performed the following experiment 200 times and average results.

For various number of samples $N_{samples}$ of normal standard univariate distribution we generated the values of the random variable x. We varied number of bins N_{BINS}, and computed histograms for each value of N_{BINS}. Using these histograms, we obtained estimates for p_i and then for estimated negentropy and normalized negentropy, eq. (10,11). Since x was generated as Gaussian, the *true* values of negentropy and normalized negentropy, $J(x)$ and $J(x)/H_{gauss}(\sigma^2)$, eq.(3-4), are equal to 0. However, due to the effect of a finite sample size used to estimate histograms, as well as due to the effects of the bin size (too small number of samples per bin would lead to incorrect estimation of p_i), estimated negentropy and its normalized value, $J(x^*)$ and $J(x^*)/H_{gauss}(\hat{\sigma}^{*2})$, would be different from zero.

Since the larger number of samples lead to better estimation, we may expect that estimated normalized entropy, eq. (11) decreases with $N_{samples}$. For a constant $N_{samplea}$, the increase of N_{BINS} initially leads to a better approximation of $p(x)$. However, when N_{BINS} becomes too large, we will not be able to properly estimate p_i and the $J(x^*)/H_{gauss}(\hat{\sigma}^{*2})$ should start to grow. Our results, see Fig. 1, confirm to this.

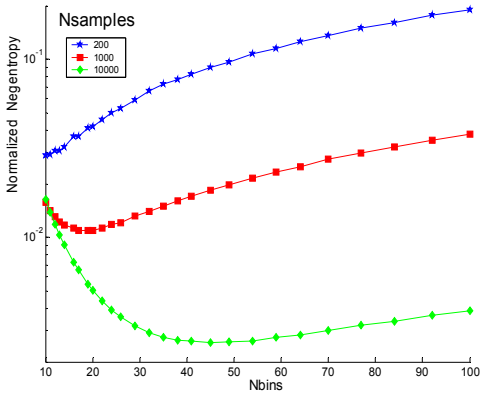


Fig.1. Estimated normalized negentropy of simulated Gaussian variable for different number of samples. The number of histogram bins is varied.

To determine the dependence of the optimal number of histogram bins on the number of distributions samples considered, we varied $N_{samples}$, gradually increased number of bins and determined the optimal number of bins for each sample size. The results of this experiment are shown on Fig. 2. As we can see, the optimal number of bins seem to depend linearly on the square root of the number of samples. Linear regression [17] provided the following empirical formula:

$$N_{BINSopt} \approx 6.2802+0.3985*sqrt(N_{samples}) \quad (14)$$

On Fig. 3a we demonstrated the application of histogram-based entropy method on Campus 3 video. The number of samples per frame was 1728=48*36 (corresponding to the number of 8*8 blocks in one video frame). The number of histogram bins, N_{BINS}=23, was obtained using the empirical formula, eq. (14).

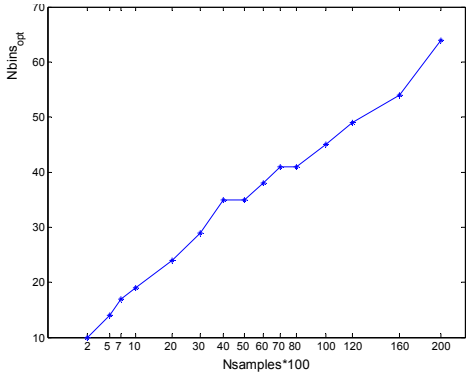
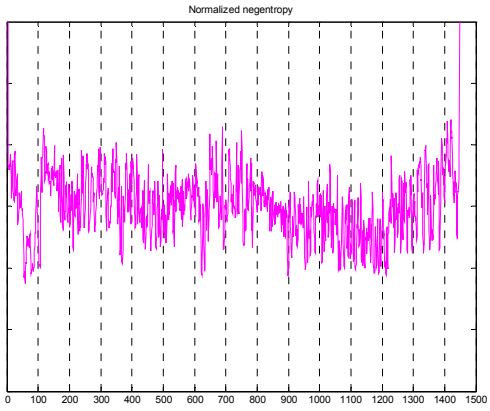
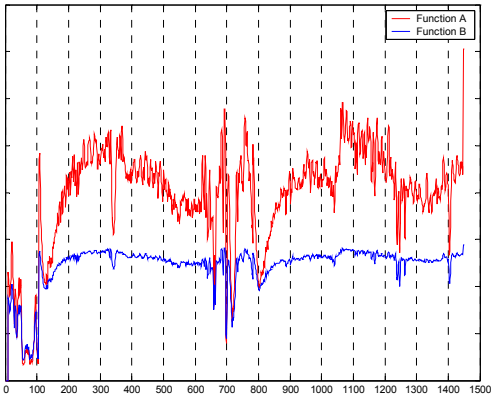


Fig.2. Optimal number of histogram bins (providing the minimal negentropy) for different number of observed samples of a Gaussian variable.



(a)



(b)

Fig.3. Estimated normalized negentropy using a) Piecewise approximation with 23 bins; b) Function A (eq. 12a) shown in red and Function B (eq. 12b) shown in blue for Campus 3 video. The negentropy is computed per video frame.

While we are able to observe fluctuations of estimated negentropy that could be associated with events on the video (e.g., between frames 600 and 800) non-parametric technique was not able to clearly identify the time intervals when the observed feature was unreliable. Generally, the estimated negentropy was noisy and hence not suitable for thresholding-based methods. One of possible explanations could be that we assumed a particular (Gaussian) shape of the distribution when estimating the optimal number of histogram bins, while the assumed shape does not properly correspond to the actual distribution of the feature. To compensate for this potential problem, we varied the number of histogram bins, but the results were always similar. Another reason why the histogram-based method failed could be that the estimation errors induced when estimating the variance of the feature (eq. 7) significantly affected the accuracy of denormalized negentropy estimation (eq 11). However, our experiments suggest that even if this normalization is not performed, the negentropy as estimated using the histogram-based method still does not seem useful for estimating feature reliability. The analog results with parametric techniques for estimating normalized negentropy are shown in Fig. 3b. Both considered parametric methods (eq. (12a) and (12b)) identified drop around frames 330, 660, a strong drop around 700, a drop around 720 and the relatively long-term drop between 800 and 900. Also, there were some small oscillations between 1200 and 1300 and one drop around 1400. All these events correspond to frames in the video sequence with significant amount of motion artifacts. Fig. 4 illustrates motion artifacts in *Campus 3* video introduced by some reflections in windows that are probably caused by cars passing by, in to the normal motion. Due to these artifacts, the feature of motion activity is unreliable. Function B (eq. 12b) was able to exactly identify the time intervals in which the motion activity feature was computed unreliable, as shown in Fig. 3b. Function B (eq. 12b) performed better, by having less oscillations and fluctuations, and hence being more suitable for thresholding.

4. CONCLUSIONS

In this paper, we discussed an entropy-based technique for feature reliability assessment. We have shown that a parametric negentropy estimation can be efficiently used to evaluate the usability of a motion measure employed for detection of moving objects. The discussed technique is useful for one-dimensional features and works under assumption that the noise, which corrupts the observed feature, is additive Gaussian. Work in progress includes generalization of the proposed technique into multidimensional case and for the feature reliability detection with non-gaussian and non-additive noise.

5. ACKNOWLEDGEMENTS

This work has been partially supported by NSF-funded "Seeds of Success: A Comprehensive Program for the Retention, Quality Training, and Advancement of STEM Student" Grant (award # 0310163). Additional support was provided by NSF-funded Infrastructure Grant (award # 0320991), NIH-funded Delaware BRIN Grant (P20 RR16472) and DoD HBCU/MI Infrastructure Support Program (45395-MA-ISP Department of Army).

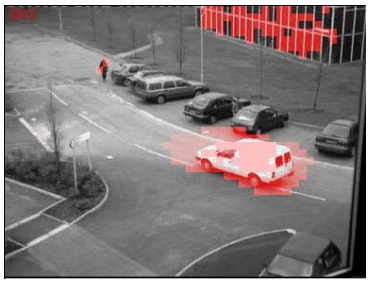


Figure 4.A frame from Campus 3 video with "false motion" artifacts introduced by reflections in the windows.

REFERENCES

[1]Remagnino, P., G. A. Jones, N. Paragios, and C. S. Regazzoni, eds., *Video-Based Surveillance Systems*, Kluwer Academic Publishers, 2002.

[2]R.T. Collins, A.J. Lipton, and T. Kanade, "Introduction to the Special Section on Video Surveillance", *IEEE PAMI* 22(8) (2000), pp. 745–746.

[3]Jain, R., Militzer, D., and Nagel, H. Separating nonstationary from stationary scene components in a sequence of real world TV images. In Proc. IJCAI, 612–618, Cambridge, MA, 1977

[4]I. Haritaoglu, D. Harwood, and L. Davis, "W4: Real-Time Surveillance of People and Their Activities", *IEEE PAMI* 22(8) (2000), pp. 809–830.

[5]N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions", *IEEE PAMI* 22(8) (2000), pp. 831–843.

[6]C. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland, "Pfinder: Real-time Tracking of the Human Body", *IEEE PAMI* 19(7) (1997), pp. 780–785.

[7]C. Stauffer, W. E. L. Grimson, "Learning patterns of activity using real-time tracking", *IEEE PAMI* 22(8) (2000), pp. 747–757.

[8]D. Pokrajac and L. J. Latecki: Spatiotemporal Blocks-Based Moving Objects Identification and Tracking, *IEEE Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, October 2003.

[9]L. J. Latecki, R. Miezianko, and D. Pokrajac. Motion Detection Based on Local Variation of Spatiotemporal Texture. *IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS)*, Washington, DC, July 2004.

[10]A. Hyvärinen, J. Karhunen, and E. Oja. Independent Component Analysis. Wiley, 2001.

[11]A. Hyvärinen. New approximations of differential entropy for independent component analysis and projection pursuit. In *Advances in Neural Information Processing Systems*, volume 10, pages 273-279. MIT Press, 1998.

[12]T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[13]C. Bishop. *Neural Networks for Patter Recognition.* Oxford University Press, 1995.

[14]R. Miezianko, L. J. Latecki, D. Pokrajac. Link to test results. *http://knight.cis.temple.edu/~video/VA*

[15]VS PETS 2002 bechmark videos, ftp://pets.rdg.ac.uk/PETS2002/

[16]Jolliffe, I. T, *Principal Component Analysis*, 2nd edn., Springer Verlag, 2002.

[17]Devore, J. L., *Probability and Statistics for Engineering and the Sciences*, 5th edn., Int. Thomson Publishing Company, Belmont, 2000.