Closed-loop active object recognition with constrained illumination power

Noom, Jacques; Soloviev, Oleg; Smith, Carlas; Verhaegen, Michel

DOI
10.1117/12.2618750

Publication date
2022

Document Version
Final published version

Published in
Proceedings Real-Time Image Processing and Deep Learning 2022

Closed-loop active object recognition with constrained illumination power

Jacques Noom, Oleg Soloviev, Carlas Smith, Michel Verhaegen

Jacques Noom, Oleg Soloviev, Carlas Smith, Michel Verhaegen, "Closed-loop active object recognition with constrained illumination power," Proc. SPIE 12102, Real-Time Image Processing and Deep Learning 2022, 1210203 (27 May 2022); doi: 10.1117/12.2618750

SPIE.

Event: SPIE Defense + Commercial Sensing, 2022, Orlando, Florida, United States

Closed-loop active object recognition with constrained illumination power

Jacques Noom[a], Oleg Soloviev[a,b], Carlas Smith[a], and Michel Verhaegen[a]

[a]Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628CD Delft, The Netherlands
[b]Flexible Optical BV, Polakweg 10–11, 2288 GG Rijswijk, The Netherlands

ABSTRACT

Some applications require high level of image-based classification certainty while keeping the total illumination energy as low as possible. Examples are minimally invasive visual inspection in Industry 4.0, and medical imaging systems such as computed tomography, in which the radiation dose should be kept "as low as is reasonably achievable". We introduce a sequential object recognition scheme aimed at minimizing phototoxicity or bleaching while achieving a predefined level of decision accuracy. The novel online procedure relies on approximate weighted Bhattacharyya coefficients for determination of future inputs. Simulation results on the MNIST handwritten digit database show how the total illumination energy is decreased with respect to a detection scheme using constant illumination.

Keywords: Active fault diagnosis, Auxiliary signal design, Machine Vision, Computational Tomography, Medical imaging, Industry 4.0

1. INTRODUCTION

Visual observation of objects often induces the dilemma of maximizing information gain, while minimizing damage to the object. For example, product inspection in Industry 4.0 is often required to be minimally invasive, such that the products maintain their usability after inspection. Also in medical imaging systems such as computed tomography, the radiation dose should be kept "As Low As is Reasonably Achievable", following the internationally known ALARA-principle. Most of these imaging systems are working with single-shot acquisition, using only one single image for recognizing the object. Counter-intuitively, multiple-shot acquisition can actually lead to a lower total illumination energy while increasing the recognition performance.

Bajcsy and Aloimonos *et al.* recently revisited[1] their contributions on active perception from the year 1988,[2,3] advocating the use of "intelligent control strategies applied to the data acquisition process which will depend on the current state of data interpretation." This would lead to improved robotic perception of the world. Recently, this paradigm was successfully implemented in two applications under the name Closed-Loop Active Model Diagnosis (CLAMD).[4,5] The inputs to the imaging system were determined sequentially after each measurement, in order to ensure quick and reliable diagnosis. Theory for CLAMD is developed under the names active fault diagnosis[6] and auxiliary signal design.[7,8] Although these fields report absence of "widespread adoption in practical applications," the theory is actually very applicable, for instance in imaging systems.

This paper uses CLAMD for high-performance object recognition with minimal phototoxicity. The proposed closed-loop imaging scheme consists of a neural network cluster for diagnosing the object with a certain confidence, and of a controller for determination of the next minimally invasive yet discriminating illumination input. The vast majority of the computations can be done offline, such that fast online execution is ensured. Simulation experiments test the procedure on the MNIST handwritten digit dataset,[9] after which the results can be compared to an open-loop approach. The methods in Section 2 consist of a problem formulation, after which the proposed solution is clarified. Simulation results are displayed and discussed in Section 3 and lastly the conclusions are presented in Section 4.

E-mail: j.noom@tudelft.nl

Real-Time Image Processing and Deep Learning 2022, edited by Nasser Kehtarnavaz, Matthias F. Carlsohn, Proc. of SPIE Vol. 12102, 1210203 · © 2022 SPIE
0277-786X · doi: 10.1117/12.2618750

2. METHODS

2.1 Problem Formulation

Consider the objects to be imaged x_k for each time step k. The objects x_k can represent one of the N_M classes, therefore N_M hypotheses are defined:

$$M_i : \ x_k \in \mathcal{X}_i \quad \forall k \tag{1}$$

with $i \in \{0, 1, \ldots, N_M - 1\}$ and $\mathcal{X}_i$ the set of images corresponding to class i. The images are formed using

$$y_k = x_k \cdot u_k + v_k, \tag{2}$$

with $u_k \in \mathbb{R}_+^{m \times m}$ the illumination input, $y_k \in \mathbb{R}_+^{m \times m}$ the output intensity, $v_k \in \mathbb{R}^{m \times m}$ the noise at time step k, and the symbol $\cdot$ represents element-wise multiplication. With all measurements available up to (and including) y_{k-1}, the hypothesis probabilities $P_{k-1}(M_i)$ evolve according to the Bayesian update rule

$$P_{k-1}(M_i) = \frac{p(y_{k-1}|M_i, u_{k-1}) P_{k-2}(M_i)}{p(y_{k-1}|u_{k-1})}, \tag{3}$$

where $p(y_{k-1}|M_i, u_{k-1}) \in \mathbb{R}_+$ is the probability density function (PDF) of output image y_{k-1}, conditioned on hypothesis M_i and input u_{k-1}. The initial conditions $P_0(M_i)$ can be set to any prior probabilities.

The goal is to minimize the probability misdiagnosis after the next measurement y_k, which is defined as[10]

$$P_{\mathrm{e}}(u_k) = \sum_i \sum_{j \neq i} \int_{\mathfrak{R}_j} p(y_k|M_i, u_k) P_{k-1}(M_i) \, \mathrm{d}y_k \tag{4}$$

where

$$\mathfrak{R}_j = \Big\{ y_k \quad \Big| \quad p(y_k|M_j, u_k) P_{k-1}(M_j) > p(y_k|M_i, u_k) P_{k-1}(M_i) \quad \forall i \neq j \Big\}.$$

In the current application we limit the so-called input energy

$$\mathrm{vec}(u_k)^\top \mathrm{vec}(u_k) \leq \varepsilon. \tag{5}$$

This input will be evaluated in the "Controller"-block in Figure 1. The closed-loop procedure has the purpose of reducing phototoxicity while increasing the reliability of diagnosis. One can for instance limit the total energy $\sum_k \varepsilon$ by limiting the number of measurements, or alternatively predefine a desired probability of misdiagnosis and iterate the scheme until this confidence is achieved.

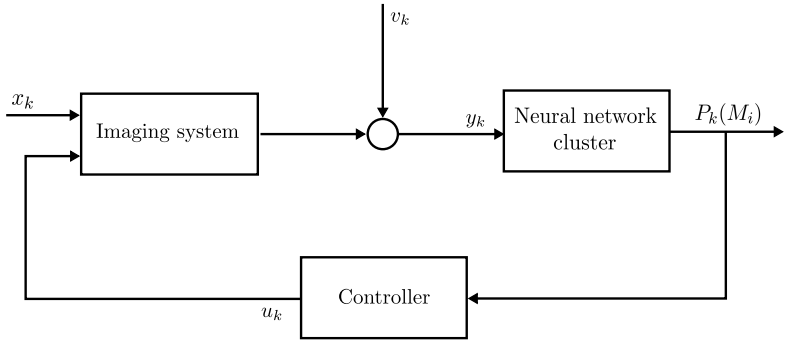

Figure 1. Control scheme. The (unknown) object x_k is fed to the imaging system together with illumination input u_k. Its output is corrupted with noise v_k, after which y_k is measured. The neural network cluster updates the probabilities of each hypothesis, after which the controller determines the input for next measurement.

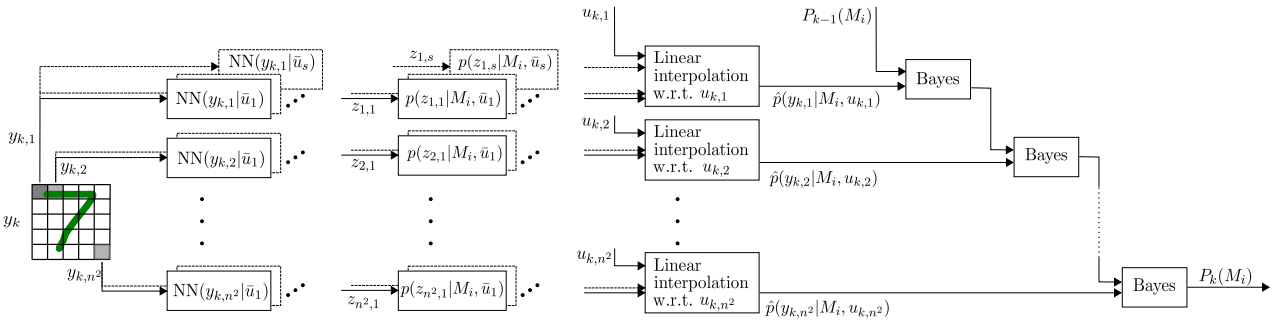

Figure 2. Neural network cluster from Figure 1 unfolded. The images y_k are split up in n^2 subimages. Each subimage is fed to s different neural networks which are trained for varying prospected inputs $\bar{u}$. The neural network scalar outputs z are converted to their corresponding relative likelihood $p(z|M_i,\bar{u})$, using the probability density functions obtained in the training phase of the neural networks. Regarding the actual input u_k which was fed to the system to obtain y_k, the relative likelihoods are then linearly interpolated. Finally, the results of the n^2 estimated relative likelihoods are combined by successive Bayesian updates as in (3).

2.2 Multiple-Input Inference Implementation

For simplification of the problem, the images are split up in n^2 subimages with uniform input $u_{k,\ell} \in \mathbb{R}$ for $\ell \in \{1,2,\ldots,n^2\}$. For each subimage, s neural networks are trained with varying prospected inputs. As depicted in Figure 2, the relative probabilities of the neural network outputs are evaluated using the accompanying PDFs for each hypothesis M_i. These PDFs were constructed by fitting Gaussian kernel density estimates to the neural network output distributions. Since multiple neural networks with different uniform illumination intensities are used for one single subimage, the outcome of the PDFs are linearly interpolated with respect to the actual illumination input. For combining the outcomes all subimages, the Bayesian update rule (3) is applied repetitively and for each hypothesis.

2.3 Closed-Loop Input Determination

The probability of misdiagnosis in (4) is bounded by[10]

$$P_{\mathrm{e}}(u_k) \leq \sum_i \sum_{j>i} \sqrt{P_{k-1}(M_i)P_{k-1}(M_j)}\,\mathfrak{B}_{ij}(u_k) \tag{6}$$

for which the Bhattacharyya coefficient is defined as

$$\mathfrak{B}_{ij}(u_k) = \int \sqrt{p(y_k|M_i,u_k)p(y_k|M_j,u_k)}\;\mathrm{d}y_k. \tag{7}$$

Furthermore, this inequality also holds for the subimages:

$$P_{\mathrm{e},\ell}(u_{k,\ell}) \leq \sum_i \sum_{j>i} \sqrt{P_{k-1}(M_i)P_{k-1}(M_j)}\,\mathfrak{B}_{ij,\ell}(u_{k,\ell}). \tag{8}$$

The total error probability is then obtained by applying the Bayesian update rule (3) to the hypothesis probabilities of all individual subimages, assumed independence of the measurements. The task now is to distribute the inputs $u_{k,\ell}$ such that the total error probability is minimized. The next step approximates the Bhattacharyya coefficient with an affine function.

2.3.1 Least-squares approximation of Bhattacharyya coefficient

A least-squares fit to the Bhattacharyya coefficient can be performed by fitting the function

$$\hat{\mathfrak{B}}_{ij,\ell}(u_{k,\ell}) = 1 + a_{ij,\ell}u_{k,\ell} \tag{9}$$

to the data gathered with the s training instances. Its solution is

$$a_{ij,\ell} = \frac{\mathfrak{B}_{ij}(\bar{u}_\ell)^\top \bar{u}_\ell - \|\bar{u}_\ell\|_1}{\bar{u}_\ell^\top \bar{u}_\ell} \tag{10}$$

with $\bar{u}_\ell = [\bar{u}_{\ell,1}^\top, \ldots, \bar{u}_{\ell,s}^\top]^\top$ the inputs for which the neural networks are trained.

2.3.2 Input determination using least-squares fit

Using approximation (9), the double sum in (8) can be simplified to

$$\sum_i \sum_{j>i} \sqrt{P_{k-1}(M_i) P_{k-1}(M_j)} \mathfrak{B}_{ij,\ell_1}(u_{k,\ell}) \approx c_k - b_{k,\ell} u_{k,\ell} \tag{11}$$

with

$$b_{k,\ell} = -\sum_i \sum_{j>i} a_{ij,\ell} \sqrt{P_{k-1}(M_i) P_{k-1}(M_j)} \tag{12}$$

$$c_k = \sum_i \sum_{j>i} \sqrt{P_{k-1}(M_i) P_{k-1}(M_j)}. \tag{13}$$

Now, one can see that a large coefficient $b_{k,\ell}$ implies a high degree of distinction between relevant models with high belief states $P_{k-1}(M_i)$, at subimage ℓ. This reasoning suggests that

$$b_{k,\ell_1} > b_{k,\ell_2} \quad \Longrightarrow \quad u_{k,\ell_1} > u_{k,\ell_2} \tag{14}$$

(with $\ell_1 \neq \ell_2$ two different realizations of ℓ) is a sound rule for approximately minimizing the right-hand side of (6), and therefore for minimizing the error probability.

With the energy constraint in (5), a possible input choice is

$$u_{k,\ell} = \frac{n}{m} \sqrt{\frac{b_{k,\ell}}{\sum_{i=1}^{n^2} b_{k,i}} \varepsilon}\,. \tag{15}$$

This solution will not lead to an overall minimum error probability, yet requires low computational effort. Moreover, the Bhattacharyya coefficients and their least-squares approximations can be calculated offline, implying that only Equations (12) and (15) need to be determined online. The next section validates the improvement in recognition performance as compared to using a constant uniform input.

3. SIMULATION RESULTS

For the objects x_k, we used the MNIST handwritten digit dataset,[9] normalized to the segment $[0, 1]$. The noise v_k is Gaussian with zero mean and variance $R_v = 0.04$ and the input energy for each measurement k is limited to $\varepsilon = 64$. The images are split into $n^2 = 49$ subimages. For each subimage, $s = 5$ neural networks are trained using inputs $\bar{u}_\ell = [0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0]^\top$. The neural network internal architecture is similar to a previous contribution,[4] having three layers with 128, 10 and 1 neurons, respectively, with the first two layers a rectified linear unit (ReLU) activation function. It is optimized using the Adam optimizer[11] in the Tensorflow package.[12]

Whereas the open-loop procedure uses constant and uniform illumination over the whole image, the closed-loop input is determined using (15). Two realizations of the closed-loop algorithm for true hypotheses M_5 and M_7 are shown in Figure 3. The challenging noise conditions in the output y_k require multiple measurements in order to find the true hypothesis with high confidence. Taking previous measurements into account, the input energy is distributed efficiently in order to achieve this high confidence in a small number of measurements.

For a Monte-Carlo simulation, 100 realizations are generated for each class, making a total of 1000 realizations. With a desired confidence of $P_k(M_i) \geq 0.98$ for any hypothesis M_i, the distribution of number of measurements before decision is presented in Figure 4. From the left plot it can be seen that the closed-loop

approach requires in general fewer measurements as compared to the open-loop approach. Moreover, on average only 3.22 measurements are required instead of 4.37. The error rates are with 3.6% for open-loop and 2.7% for closed-loop slightly higher than the desired 2%. This is probably due to inaccuracies in obtaining the probability density functions from the neural network outputs. Besides, there might be a dependence between measurements in different subimages, whereas independence was assumed. Interestingly, the error rate was for the open-loop approach higher, while it uses on average more measurements than the closed-loop approach. So in fact, the open-loop approach would require even more measurements to obtain an error rate equivalent to the closed-loop approach.

The right plot in Figure 4 confirms the decrease in average number of measurements for each hypothesis. The digits 0 and 1 are on average diagnosed in fewer measurements than the remaining digits, which is presumably due to their apparent uniqueness in graphical appearance.

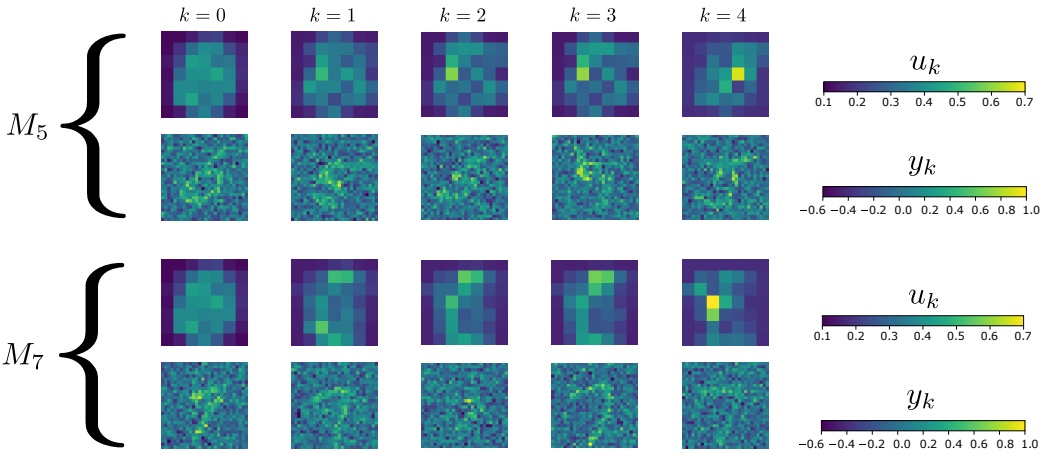

Figure 3. Realizations of the closed-loop algorithm with true hypothesis M_5 (top) and M_7 (bottom).

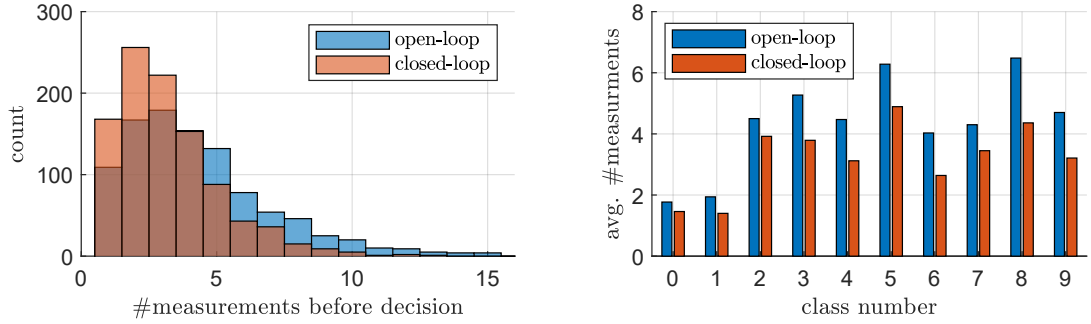

Figure 4. Distribution of number of measurements before a decision with 98% desired confidence is taken, for open-loop (blue) and closed-loop solution (red). Overall distribution (left) and average number of measurements per hypothesis (right). The final error rates of the open- and closed-loop approaches were 3.6% and 2.7%, respectively.

4. CONCLUSIONS

This paper demonstrated the use of CLAMD in reliable object recognition while minimizing phototoxicity. The closed-loop methodology distributes the input illumination based on belief states and spatial model divergences. As compared to open-loop, simulation results show for the closed-loop approach a decrease in average required

number of measurements for a high-confidence decision. Furthermore, the average number of measurements was smaller for each individual hypothesis. By further investigating inputs which are closer to the optimum that minimizes the error probability, a larger improvement can be achieved. Additionally, future developments should aim for practical implementation of closed-loop recognition, for instance in medical imaging systems. This would lead to improved recognition performance with decreased phototoxicity.

ACKNOWLEDGMENTS

This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826589. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Netherlands, Belgium, Germany, France, Italy, Austria, Hungary, Romania, Sweden and Israel.

REFERENCES

[1] Bajcsy, R., Aloimonos, Y., and Tsotsos, J. K., "Revisiting active perception," *Autonomous Robots* **42**, 177–196 (2018).

[2] Bajcsy, R., "Active Perception," *Proceedings of the IEEE* **76**(8), 996–1005 (1988).

[3] Aloimonos, Y., Weiss, I., and Bandyopadhyay, A., "Active Vision," *International Journal of Computer Vision* **1**, 333–356 (1988).

[4] Noom, J., Thao, N. H., Soloviev, O., and Verhaegen, M., "Closed-Loop Active Model Diagnosis Using Bhattacharyya Coefficient: Application to Automated Visual Inspection," in [*Intelligent Systems Design and Applications (ISDA 2020)*], Abraham, A., Piuri, V., Gandhi, N., Siarry, P., Kaklauskas, A., and Madureira, A., eds., 657–667, Springer, Cham (2021).

[5] Noom, J., Soloviev, O., Smith, C., Nguyen, H. T., and Verhaegen, M., "Particle detection using closed-loop active model diagnosis," in [*AI and Optical Data Sciences III*], Jalali, B. and ichi Kitayama, K., eds., **12019**, 106 – 110, International Society for Optics and Photonics, SPIE (2022).

[6] Heirung, T. A. N. and Mesbah, A., "Input design for active fault diagnosis," *Annual Reviews in Control* **47**, 35–50 (2019).

[7] Zhang, X. J. and Zarrop, M. B., "Auxiliary signals for improving on-line fault detection," in [*1988 International Conference on Control - CONTROL 88*], 414–419 (1988).

[8] Zhang, X. J., [*Auxiliary Signal Design in Fault Detection and Diagnosis*], Springer-Verlag, Berlin, Heidelberg (1989).

[9] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE* **86**(11), 2278–2324 (1998).

[10] Blackmore, L. and Williams, B., "Finite horizon control design for optimal discrimination between several models," in [*Proceedings of the 45th IEEE Conference on Decision and Control*], 1147–1152 (2006).

[11] Kingma, D. P. and Ba, J. L., "Adam: A method for stochastic optimization," in [*International Conference on Learning Representations (ICLR)*], (2015).

[12] "TensorFlow: Large-scale machine learning on heterogeneous systems." https://www.tensorflow.org/ (2015).