Training and Validation of a Wide-angle Optical Scattering (TAOS) Pattern Classifier

Giovanni Franco Crosta[1], **Yong-Le Pan**[2], **and Gustavo Eddino Fernandes**[2]

[1]Dipartimento di Scienze dell'Ambiente & Territorio, University of Milan-Bicocca, Milan, Italy
[2]Department of Applied Physics, Yale University, New Haven, CT 06520, USA

Abstract— *TAOS* (two-angle optical scattering) is a well-known experimental technique which records the intensity patterns of laser light scattered by single aerosol particles over an extended range of the scattering angles $\{\theta, \phi\}$ [1]. In the absence of a method which solves the inverse obstacle problem from intensity data, patterns have to be classified by artificial intelligence techniques. The classifier described herewith extracts features from *TAOS* patterns by the spectrum enhancement algorithm [2], which is controlled by a few parameters (the n-tuple ψ), not described herewith. Training corresponds to finding a ψ which maximizes a suitable figure of merit (F). The newly developed training-validation scheme is illustrated by the following example. **1)** Out of 100 *TAOS* patterns produced by clusters of polystyrene spheres of controlled size (material $a6$, **class 1**) and 100 patterns from single spores of *Bacillus globigii* (material bg, **class 2**), form e.g., 10 training (T) sets of 50 patterns from each class, which differ by at least $5 + 5$ patterns from one another. Two sample patterns are shown by Figures 1 and 2 below.

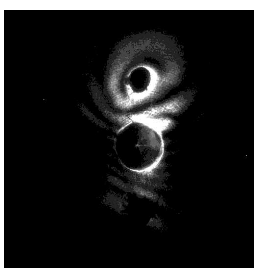

Figure 1: Scattering pattern of a polystyrene (PS) sphere aggregate (**class 1**).

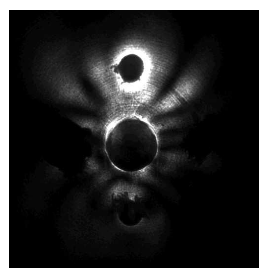

Figure 2: Scattering pattern of Bacillus globigii (Bg) spores. (**class 2**).

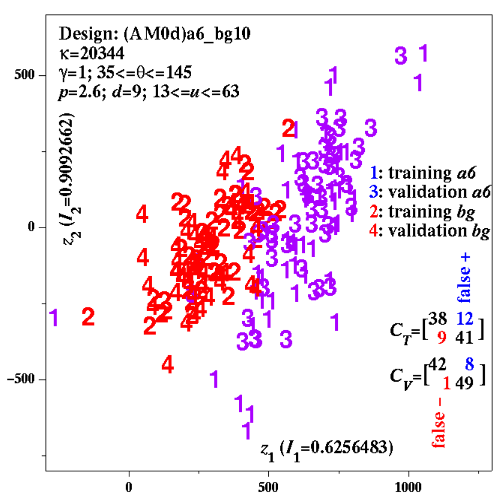

Figure 3: Classifier output: representation of *TAOS* patterns on the plane of the first two principal components z_1, z_2. To each pattern there corresponds a point on $\{z_1, z_2\}$. Patterns are labeled after their class of belonging and role in either T- or V-mode. The parameters γ, p, d and the endvalues of both θ and u form the n-tuple ψ; u is the spatial wavenumber on which enhanced spectra depend. I_1, I_2 are the fractions of sample variance explained by z_1 alone and by z_1 and z_2 together. C_T and C_V are the T- and V-mode confusion matrices. A "false-" (false negative) is a misclassified bg pattern.

2) Evaluate $F[\psi]$ for all available ψ and all T-sets, keep those ψ such that $F[\psi] > \tau$, a threshold value. **3)** Select a ψ which minimizes $H[\psi] := -\Sigma_\psi \ln[f[\psi]]F[\psi]$, where $f[\psi]$ is the relative frequency by which $F[\psi]$ occurs. **4)** Use the leftover patterns for classifier validation (V).

Figure 3 displays the classification result which corresponds to $F = 0.824$ in training (T) and $F = 0.883$ in validation (V) mode. The eventual goal is to implement a realtime TAOS pattern classifier.

REFERENCES

1. Fernandes, G. E., Y.-L. Pan, R. K. Chang, K. Aptowicz, and R. G. Pinnick, "Simultaneous forward- and backward-hemisphere elastic-light-scattering patterns of respirable-size aerosols," *Opt. Lett.*, Vol. 31, 3034–3036, 2006.
2. Crosta, G. F., C. Urani, and L. Fumarola, "Classifying structural alterations of the cytoskeleton by spectrum enhancement & descriptor fusion," *J. Biomed. Optics*, Vol. 11, No. 2, 024020:1–18, 2006.