none

Asuka Sawamoto[a] and Shigeki Nakauchi[b]

[a,b]Department of Computer Science and Engineering, Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi, Japan 441-8580

ABSTRACT

Keywords: Microplastics, Fluorescence imaging, UV excitation, Machine learning, Spectral optimization, High-throughput screening

1. INTRODUCTION

The widespread accumulation of microplastics (MPs) in the natural environment has become an increasingly serious international issue [1, 2]. To understand the actual state of MP pollution and identify its sources, it is necessary to comprehensively quantify the distribution and flux of MPs across various media [3].

However, while conventional analytical methods such as Fourier Transform Infrared Spectroscopy (FTIR) and Raman spectroscopy excel at identifying the chemical structure of MPs, they rely on point-scanning, which requires an enormous amount of time for analysis [4, 5]. Therefore, for in-process applications, there is a strong demand for the development of high-throughput solutions that achieve both practical classification accuracy and rapid analysis speeds [6].

In this study, we propose a multispectral imaging method specialized for polymer identification based on the Excitation-Emission Matrix (EEM). This method utilizes the inherent auto-fluorescence emitted by polymers under UV excitation to capture the spectral features necessary for classification as images. First, toward the future construction of a high-throughput system, we conducted a feasibility study using a benchtop system equipped with a broadband light source. Furthermore, we identified a combination of wavelength bands that minimizes measurement time while maintaining classification accuracy.

2. PRINCIPLE

2.1 Fluorescence Characteristics

Most petroleum-based synthetic polymers emit specific auto-fluorescence under ultraviolet excitation due to their chemical structures, such as electron conjugated systems, and additives. This fluorescence characteristic depends on the EEM and serves as a unique fingerprint for each resin type. In this study, we use these EEM characteristics as the physical basis for MP classification.

Generally, the fluorescence intensity F at an excitation wavelength λ_{ex} and an emission wavelength λ_{em} is described by the following simplified model:

$$F(\lambda_{ex}, \lambda_{em}) = K \cdot I_0(\lambda_{ex}) \cdot (1 - 10^{-\varepsilon(\lambda_{ex})cL}) \cdot \Phi(\lambda_{ex}, \lambda_{em}) \tag{1}$$

where K is the instrument constant (geometric factors and detection efficiency), I_0 is the excitation light intensity, ε is the molar absorption coefficient, c is the concentration, L is the optical path length, and Φ is the fluorescence quantum yield. In the dilute solution approximation or thin film approximation, Equation (1) can be considered linear with respect to the material-intrinsic parameters ε and Φ.

Since the spectral profiles of ε (absorption characteristics) and Φ (emission efficiency) differ depending on the polymer type, measuring these at multiple wavelengths enables material identification. As a preliminary experiment, we measured the EEMs of the nine target polymers used in this study (HDPE, LDPE, PP, PS, PVC, PC, PMMA, PET, ABS) using a fluorescence spectrophotometer (F-7000, Hitachi High-Tech). The obtained EEM spectra are shown in Figure 1. These results confirmed that each polymer possesses a specific fluorescence spectrum under UV excitation.

Further author information: (Send correspondence to A.S.)
A.S.: E-mail: sawamoto.asuka.us@tut.jp, Telephone: +81 90 7647 0330

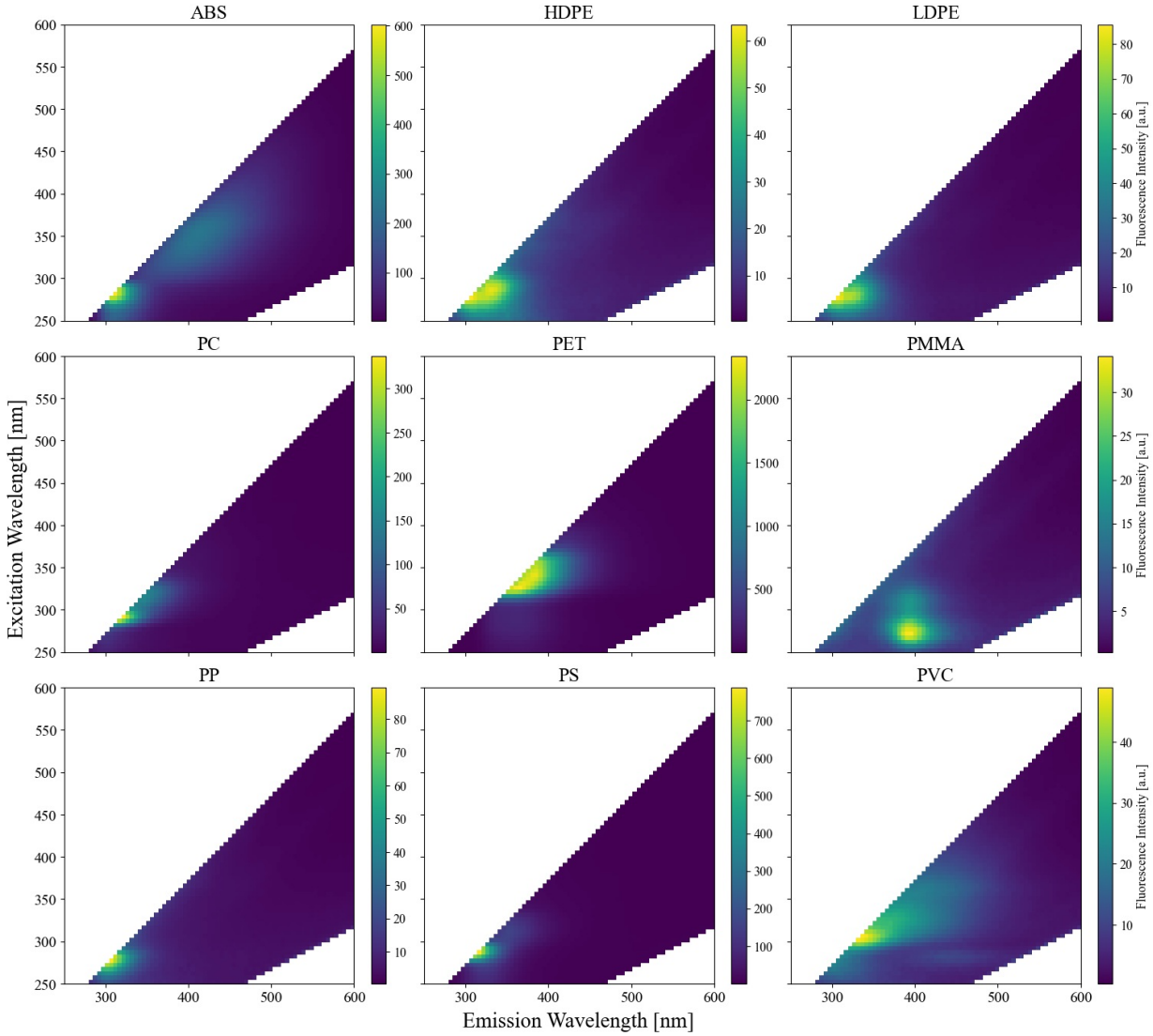


Figure 1: Excitation-Emission Matrices (EEMs) of the nine types of virgin polymer pellets used in this study. The contour plots represent the fluorescence intensity [a.u.], with resin names indicated above each plot. The distinct spectral patterns (fingerprints) observed for each polymer type demonstrate the feasibility of classification based on UV-excited fluorescence. Note: The diagonal regions corresponding to Rayleigh and Raman scattering have been computationally removed to enhance the visibility of fluorescence features.

2.2 Data-Driven Band Selection

To achieve the high-throughput imaging essential for effective on-site environmental monitoring, minimization of data acquisition time is required. Acquiring and processing spectral data for the entire wavelength range (Datacube) is inefficient for real-time processing due to the massive data volume and spectral redundancy. Therefore, an approach is needed to selectively acquire only the minimal wavelength bands that have a high contribution to discrimination. However, since spectral data possess strong correlations between bands and contain non-linear relationships, it is difficult to determine the optimal subset using simple linear models.

In this study, we employed Random Forest (RF), which can account for non-linear interactions between features. In RF, the importance (Gini Importance) of a feature (wavelength band) X_j is defined as the sum of the decrease in impurity at the nodes where that feature was used for splitting, as shown in the following equation:

$$I_G(X_j) = \frac{1}{N_T} \sum_{T} \sum_{t \in T : v(s_t) = X_j} p(t) \Delta i(s_t, t) \tag{2}$$

where N_T is the total number of trees constituting the forest, $v(s_t)$ is the variable used for splitting at node t, $p(t)$ is the proportion of samples reaching node t, and $\Delta i(s_t, t)$ indicates the Gini Impurity Decrease due to the split.

Based on this $I_G(X_j)$, we applied a stepwise method in which bands with high importance are sequentially added and evaluated. This allowed us to determine the optimal band configuration that minimizes imaging time while maintaining classification accuracy.

3. EXPERIMENTAL SETUP

3.1 Optical Configuration

To acquire comprehensive data for spectral optimization, we used a benchtop evaluation system equipped with a broadband light source [7]. The schematic diagram of the experimental apparatus is shown in Figure 2. A broadband xenon lamp (MAX-303, ASAHI-SPECTRA) was used as the light source. The excitation wavelength was selected via bandpass filters within a filter wheel and irradiated the sample through an optical fiber. The emitted light from the sample passed through an emission-side bandpass filter, which removed the reflection component (scattered light) of the excitation light, allowing only the fluorescence signal transmitted through the filter to be detected by a UV-sensitive camera (BU-56DUV, BITRAN).

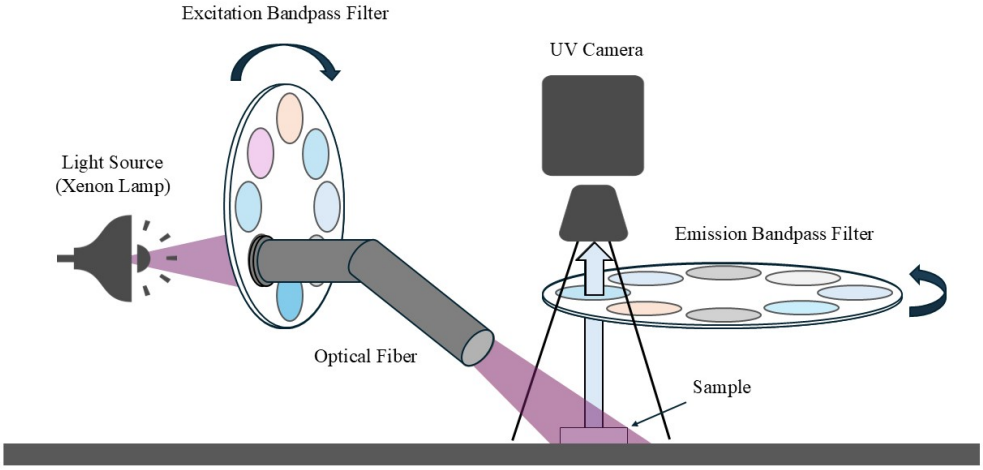


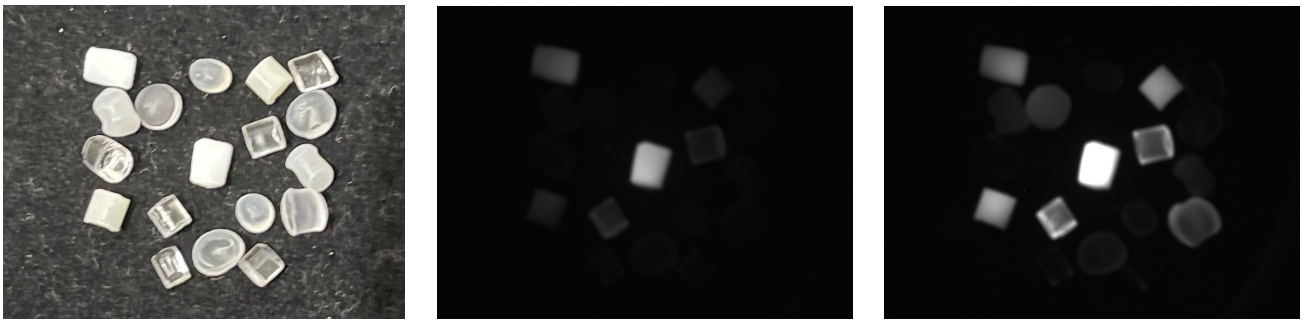
Figure 2: Schematic diagram of the benchtop evaluation system used for hyperspectral data acquisition. A broadband xenon lamp and interchangeable bandpass filters were used to excite the samples, and the emitted fluorescence was captured by a UV-sensitive camera to construct a comprehensive dataset for spectral optimization.

3.2 Samples and Data Acquisition

We used nine types of virgin resin pellets (HDPE, LDPE, PP, PS, PVC, PC, PMMA, PET, ABS), which account for the majority of global plastic production, as experimental samples. The size of each pellet is approximately 3-5 mm (Figure 3a). While these are transparent or white under visible light and difficult to distinguish, it was confirmed that they exhibit characteristic fluorescence intensity differences under UV excitation (Figure 3b, c). To improve data acquisition efficiency, the nine types of pellets were arranged within the same field of view

during imaging. The datasets used for model training and evaluation were acquired in independent imaging sessions with different sample arrangement conditions.

- **Training Set:** To cover spectral variations depending on the arrangement state and learn robust features, we integrated images acquired in the following two sections:
  1. *Adjacent Arrangement:* Images where pellets of the same type are arranged to touch and cluster together.
  2. *Dispersed Arrangement:* Images where pellets are randomly dispersed so that they do not touch each other.
- **External Validation Set:** A dataset used solely for the final performance evaluation of the model, not used in the training process. It consists only of images with Dispersed Arrangement. This verified the pure identification capability for unknown data.



(a) Visible Light (b) Ex 320 / Em 400 (c) Ex 360 / Em 480

Figure 3: Images of the experimental samples consisting of nine types of virgin polymer pellets. (a) Visible light image; most samples appear transparent or white. (b, c) Representative fluorescence images acquired at (b) Ex 320 nm / Em 400 nm and (c) Ex 360 nm / Em 480 nm. Note: The sample arrangement differs between visible and fluorescence images.

4. METHOD

4.1 Pre-processing

The pixel intensity of the acquired fluorescence images strongly depends on the geometric shape of the sample, the distance from the light source, and the non-uniformity of illumination. These variations in absolute intensity values hinder accurate polymer identification. Therefore, in this study, to achieve robust identification based solely on the polymer-specific "spectral profile" rather than absolute intensity, we applied the following minimum value subtraction and L2 normalization to the spectral data vector $\mathbf{x}$ of each pixel:

$$\mathbf{x}' = \frac{\mathbf{x} - \min(\mathbf{x})}{\|\mathbf{x} - \min(\mathbf{x})\|_2}. \tag{3}$$

This process generates a normalized feature vector $\mathbf{x}'$ that is independent of lighting conditions and sample arrangement.

4.2 Feature Selection Strategy

To determine the optimal subset of wavelength bands, we implemented the step-wise method based on the Random Forest Gini Importance described in Section 2.2. The specific procedure is as follows:

1. **Ranking (Importance Calculation):** An RF model was trained using the pre-processed training set, and the importance $I_G(X_j)$ of each band was calculated based on Equation (2).
2. **Iterative Evaluation:** Features (bands) were added to the model one by one in descending order of calculated importance, and discrimination performance was evaluated each time.
3. **Determination of Optimal Subset:** To consider the balance of the entire system rather than just specific plastics, the "Weighted Average F1-Score of all 9 classes" was adopted as the evaluation metric. The point where the increase in F1-score saturated with the increase in the number of bands, and the number of images was minimized, was adopted as the optimal band configuration.

5. RESULTS

5.1 Optimization of Spectral Bands

The results of feature selection by the step-wise method using the training dataset are shown in Figure **??**. As the number of wavelength bands k used increased, the weighted F1-score for all 9 classes rose rapidly, exceeding 0.90 at $k = 2$. Subsequently, the improvement in the score became gradual, and it was confirmed that it almost saturated (0.99 or higher) around $k = 5$. Finally, considering the trade-off between slight accuracy improvement and computational cost, we adopted the band configuration of $k = 8$, which showed the highest stability, as the optimal solution. In this internal validation stage, the model showed an extremely high precision rate.

5.2 Performance on External Dataset

Using the selected 8 bands, we classified the external validation dataset. As a result, a high weighted average F1-score (comparable to the training phase) was maintained. As shown in Figure **??**, 7 out of the 9 types were classified with extremely high precision (F1-score ≥ 0.98). The fact that similar trends were observed in the external data as in the training phase is evidence that the constructed model is not overfitting to specific sample arrangements or lighting conditions, but that the selected features function regardless of arrangement.

6. DISCUSSION

6.1 Interpretation of Optimization Behavior

In the step-wise method adopted in this study, a temporary decrease in the score for HDPE was observed (Section 5.1). This is interpreted as a class-specific trade-off resulting from the algorithm prioritizing global optimization to maximize the average performance of all classes. However, for the purpose of this study, which is the simultaneous screening of a wide variety of plastics, it is rational to prioritize the improvement of the system's overall average performance over slight fluctuations in specific classes. As a result, the selected 8 bands showed high generalization performance on the external dataset (Section 5.2), confirming the effectiveness of this optimization strategy.

6.2 High-Throughput Capability

The greatest advantage of this method is the overwhelming throughput achieved by area-scan. The experimental results demonstrate a speedup of several orders of magnitude compared to conventional point-scanning methods.

6.3 Limitations and Potential Improvements

There are several limitations to this study. First, some confusion was observed between PET and HDPE (PET misclassification rate: approx. 35%). Although their EEM spectral shapes are originally different, the reduction of wavelength bands to prioritize data acquisition efficiency caused their feature values to become close in the selected feature space (feature overlap), making the discrimination boundary ambiguous. This issue has the potential to be improved by introducing Deep Learning, which can learn more complex non-linear boundaries, to improve accuracy even with a limited number of bands. Second is the scope of direct accuracy comparison. Since this study focuses on throughput improvement, strict particle-to-particle comparison with conventional methods was not performed. Third, this experiment was conducted using virgin pellets. Since microplastics in actual marine environments may have altered fluorescence characteristics due to biofilm adhesion or weathering, further verification using real environmental samples is necessary.

7. CONCLUSION

In this study, we demonstrated the feasibility of a high-throughput microplastic sorting method using UV excitation fluorescence imaging. Through band selection using the step-wise method, we revealed that an average classification accuracy of 89% can be obtained with only 8 wavelength bands. The proposed method achieves high throughput compared to conventional methods and has high industrial practicality. In the future, we aim for the field deployment of this method and will transition to the verification phase in real environments.

8. FUTURE WORK: DEVELOPMENT OF A COMPACT DEDICATED SYSTEM

As mentioned in the Introduction, this study was conducted as a proof-of-principle for the future construction of a high-throughput system. Based on the verification results in this paper, we designed a small dedicated system for on-site implementation. Figure 4 shows the block diagram of the proposed system. To replace the large benchtop equipment, high-power UV-LEDs were adopted as the light source. On the other hand, commercially available LEDs have a wider Full Width at Half Maximum (FWHM) compared to bandpass filters, and the selectable center wavelengths are discrete. In addition, important wavelength bands contributing to identification may change due to system-specific factors such as the camera's spectral sensitivity characteristics and light source intensity. Therefore, in this design, we selected a group of commercially available LEDs that approximate the spectral regions suggested to be effective in the benchtop experiment. In the future, using the compact dedicated system, we plan to perform re-optimization of features based on actual hardware characteristics and proceed with demonstration experiments using actual sea area samples. A single-board computer (SBC, e.g., Raspberry Pi 5) will be adopted for the control system, aiming to realize a low-cost, standalone monitoring solution.

REFERENCES

[1] Hale, R. C., Seeley, M. E., La Guardia, M. J., Mai, L., and Zeng, E. Y., "A global perspective on microplastics," *J. Geophys. Res. Oceans* **125** (Jan. 2020).

[2] Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., Narayan, R., and Law, K. L., "Marine pollution. plastic waste inputs from land into the ocean," *Science* **347**, 768–771 (Feb. 2015).

[3] Rocha-Santos, T. and Duarte, A. C., "A critical overview of the analytical approaches to the occurrence, the fate and the behavior of microplastics in the environment," *Trends Analyt. Chem.* **65**, 47–53 (Feb. 2015).

[4] Prata, J. C., da Costa, J. P., Duarte, A. C., and Rocha-Santos, T., "Methods for sampling and detection of microplastics in water and sediment: A critical review," *Trends Analyt. Chem.* **110**, 150–159 (Jan. 2019).

[5] Shim, W. J., Hong, S. H., and Eo, S. E., "Identification methods in microplastic analysis: a review," *Anal. Methods* **9**(9), 1384–1391 (2017).

[6] Serranti, S., Capobianco, G., Cucuzza, P., and Bonifazi, G., "Efficient microplastic identification by hyperspectral imaging: A comparative study of spatial resolutions, spectral ranges and classification models to define an optimal analytical protocol," *Sci. Total Environ.* **954**, 176630 (Dec. 2024).

[7] Bui, M. V., Rahman, M. M., Nakazawa, N., Okazaki, E., and Nakauchi, S., "Visualize the quality of frozen fish using fluorescence imaging aided with excitation-emission matrix," *Optics Express* **26**(18), 22954–22964 (2018).

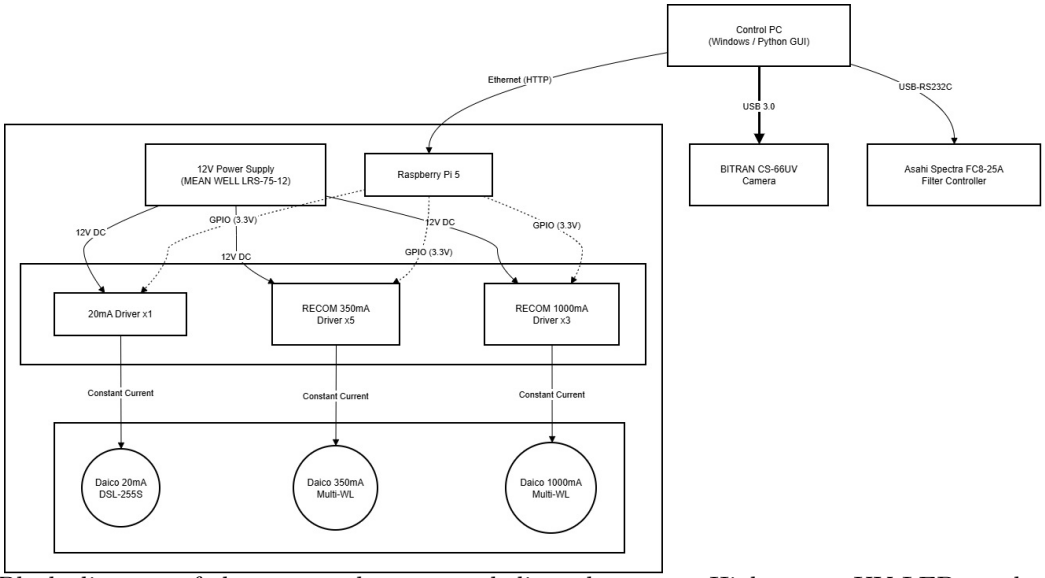


Figure 4: Block diagram of the proposed compact dedicated system. High-power UV-LEDs and an SBC are used to achieve a standalone, low-cost solution.