Advanced Data Analysis and Machine Learning: Applications in Gem Identification

Matthew F. Hardman

Gemological Institute of America, Carlsbad, California

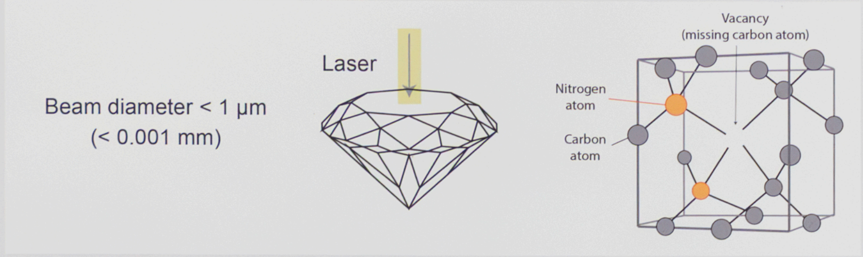
1. Introduction

- Natural and laboratory-grown diamonds can be indistinguishable to the naked eye.
- Growth and/or treatment history of diamond must be determined using non-destructive analytical techniques.
- Photoluminescence (PL) spectroscopy is used to characterize the atomic structure of diamond, which can vary based on growth, and post-growth treatment.
- Spectral data can be simplified and evaluated using statistics and machine learning (ML).

2. Samples and Data

- PL spectra for 1,121 natural diamonds, measured with 457 and 514 nm lasers are compiled. Also compiled are PL spectra for 1,178 diamonds grown by chemical vapor deposition (CVD). Of the latter, 578 are concluded to be untreated, and 600 are concluded to have undergone post-growth treatment under high-pressure and high-temperature (HPHT) conditions to improve their color. All diamonds in this study are D-color.

Analysis of diamond with PL spectroscopy

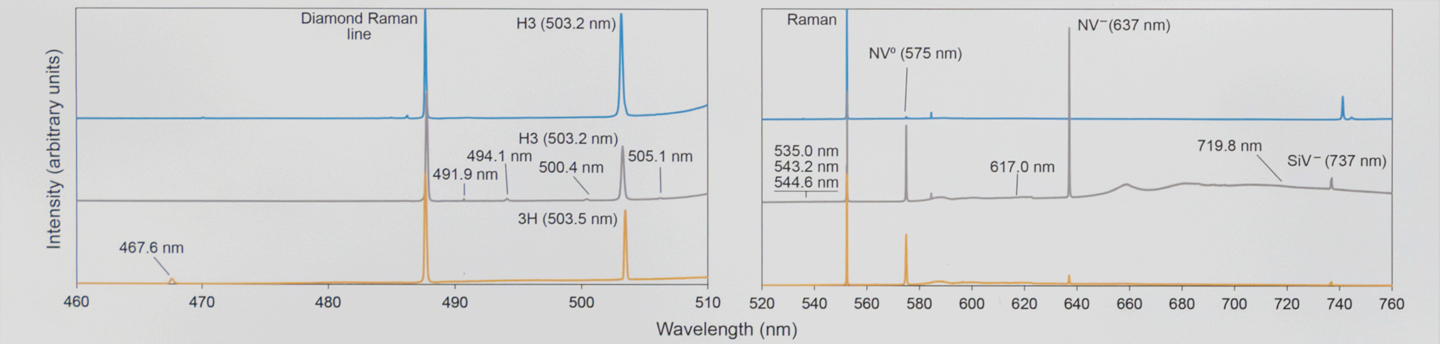


Diamond is analyzed using a laser beam with a particular wavelength

The laser interacts with the diamond lattice.

The combination of nitrogen-vacancy-nitrogen in this example defines the H3 defect. H3 is detected as the 503.2 nm peak by PL using a 457 nm laser (below)

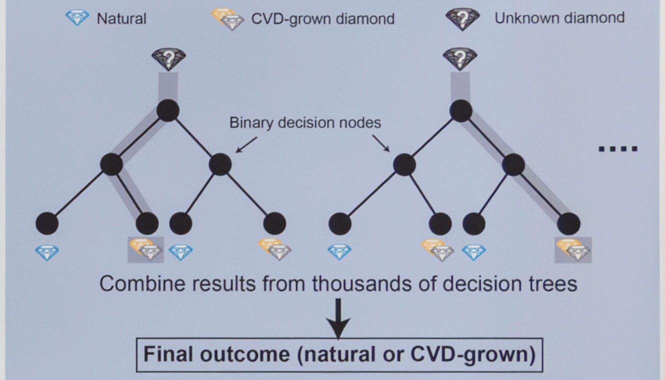
Representative PL spectroscopic data for diamonds



Representative PL spectra for three diamonds, acquired using 457 nm and 514 nm lasers. Spectra were acquired at liquid nitrogen temperatures.

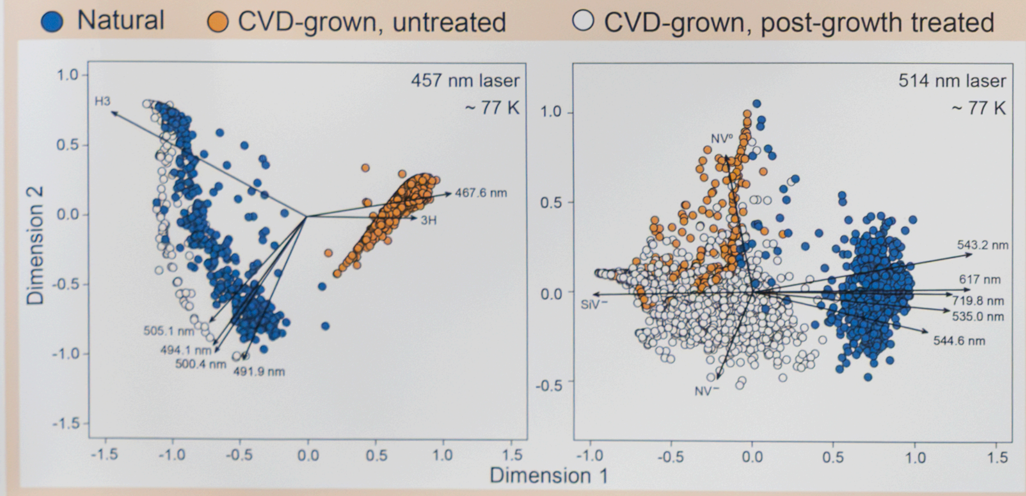
3. Methods

- In PL spectra, the areas of different peaks or bands can be used to evaluate the presence and/or concentration of different atomic defects at the measured position.
- To compare peak areas in PL from different diamonds, the areas of the peaks are normalized to the area of the diamond Raman line, a feature that is intrinsic to diamond.
- The peak areas for seven peaks in spectra measured by a 457 nm and eight measured by a 514 nm laser are Raman-normalized.
- Multidimensional scaling (MDS) transforms the complex data into single data points that can be compared to each other.
- The random forest (RF) machine learning model differentiates diamonds using their spectroscopic data as inputs.



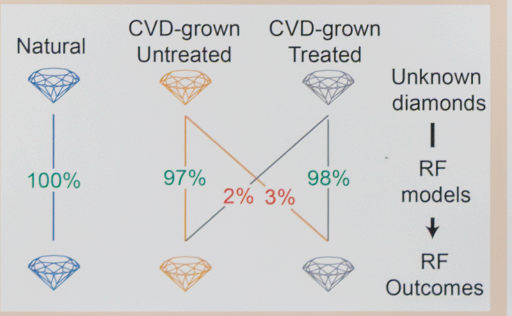
Example random forest approach to diamond classification, based on combination of multiple decision trees

4. Results



Raman-normalized peak areas from PL spectra in this study. Values are transformed using MDS and plotted in two dimensions. Points close together represent diamonds with PL spectra that are similar in all variables simultaneously. Black arrows indicate relationships between variables. Arrows that are close to parallel indicate features that are broadly correlated.

Random forest: Classification results



By applying machine learning to the transformed spectroscopic data measured by both 457 and 514 nm lasers, no natural diamond is misclassified as CVD-grown (and vice-versa).

5. Discussion

- PL spectra can be used to characterize diamond defects, which can help to evaluate diamond origin and/or treatment history.
- Machine learning and statistical methods can reduce the complexity of large spectroscopic datasets.
- Untreated and HPHT-processed CVD-grown diamonds tend to have very different 457 nm spectra: the latter typically have stronger H3, and the former stronger 3H and 467.6 nm peaks, leading to visual separation (See Results). These features can also occur in natural diamonds, but the unique combination of relative intensities enables them to be distinguished.
- PL spectra measured with a 514 nm laser can readily separate natural from CVD-grown diamonds, as the latter commonly have very strong silicon-vacancy defects (737 nm in PL).
- These data are combined with other data - gemological, infrared absorption and ultraviolet-visible absorption spectroscopy, and fluorescence imaging - to render judgments on the origin and treatment history of diamonds.
- HPHT-grown diamonds can also be evaluated.