Microchemical Journal 129 (2016) 173–183

Contents lists available at ScienceDirect

Microchemical Journal

journal homepage: www.elsevier.com/locate/microc

# Forensic analysis of automotive paints using a pattern recognition assisted infrared library searching system: Ford (2000–2006)

Barry K. Lavine *, Collin White, Matthew Allen

*Department of Chemistry, Oklahoma State University, Stillwater, OK 74078, United States*

ARTICLE INFO

*Article history:*
Received 16 June 2016
Accepted 23 June 2016
Available online 25 June 2016

*Keywords:*
Forensic automotive paint analysis
Spectral library matching
Wavelets
Genetic algorithm
Pattern recognition
Cross correlation
Search prefilters
Forward and backward library searching
Variable selection

ABSTRACT

A prototype infrared (IR) library search system for the paint data query (PDQ) database has been further developed to determine the make, line and model of an automotive vehicle from the clear coat, surfacer-primer, and e-coat layers in an effort to improve discrimination capability in automotive paint comparisons involving intact paint chips. Search prefilters for the IR spectral library of PDQ were developed from the clear coat, surfacer-primer and e-coat layers for 1179 manufacturer paint systems within a limited production year range (2000–2006) to identify vehicle manufacturer (Ford, Chrysler, and General Motors). For each make (i.e., manufacturer), search prefilters were developed to identify the assembly plant of the vehicle using a hierarchical classification scheme. A cross correlation library search algorithm that performed both forward and backward searching was then used to identify the line and model of the vehicle from the truncated IR spectral library of PDQ identified by the search prefilters. Samples assigned to the same line and model by both a forward and backward search of the IR spectral data were always correctly matched, always correlated well on an individual basis to a specific library sample and were well represented in the truncated PDQ spectral library identified by the search prefilters. The performance of the prototype IR library searching system (search prefilters and cross-correlation library search algorithms) for the PDQ database was benchmarked against commercial library searching algorithms. Only the results for Ford are reported here.



## 1. Introduction

Multi-layered paint fragments (i.e., paint chips) are often recovered from a crime scene involving a hit-and-run where damage to a vehicle or injury and/or death to a pedestrian has occurred. In these situations, the task confronting a forensic paint examiner is to identify the make, line, model, and production year of automotive vehicles with paint of similar composition to that of the recovered paint fragment (chip). To make this identification possible, the Royal Canadian Mounted Police (RCMP) have developed a comprehensive forensic automotive paint database known as the paint data query (PDQ) database [1,2]. PDQ is a database of the physical attributes, the chemical composition, and the infrared (IR) spectrum of each layer of the original manufacturer's paint system for over 21,000 individual samples representing paint systems used in domestic and foreign vehicles sold in North America. If the original manufacturer's paint system is present in a recovered paint chip, PDQ can assist in identifying the automotive vehicle. Currently, PDQ is used by forensic scientists in the US (local, state, and federal crime laboratories including the FBI Laboratory), Canada, Australia, New Zealand, Singapore, South Africa, United Arab Emirates, and 19 European countries.

Modern automotive paints [3] have a typical layer sequence of clear coat, color coat, surfacer-primer and e-coat. Each layer (with the exception of the clear coat) contains pigments and fillers, and all layers contain binders. Each automotive manufacturer employs a unique combination of pigments, fillers, binders, and layers, and it is this unique combination that allows forensic scientists to determine the possible make, line, and model of a vehicle within a limited production year range from a paint chip recovered at a crime scene.

In order for a forensic scientist to use PDQ, the color and chemical formulation of each layer must be translated into specific text codes based on the IR spectrum of the layer and the guidelines established for the database. The text based retrieval system of PDQ then searches the database to identify the make (i.e., manufacturer), line, model, and production year of vehicles whose paint systems correspond to the coded information provided by the user. However, the use of text to encode the chemistry of each layer is sometimes problematic as coding is generic and can lead to nonspecific search criteria resulting in a very large number of hits that a forensic scientist must then work through and eliminate.

* Corresponding author.
*E-mail address:* bklab@chem.okstate.edu (B.K. Lavine).

http://dx.doi.org/10.1016/j.microc.2016.06.024
0026-265X/

To address this problem, prefilters had been previously developed to search the IR spectral library of the PDQ database to identify the assembly plant(s) of the vehicle whose paint system corresponds to the sample recovered at the crime scene in an effort to improve the discrimination capability for automotive paint comparisons involving the original equipment manufacturer. However, these search prefilters [4–10] were limited to a single manufacturer (General Motors or Chrysler) and to a single layer of automotive paint (clear coat layer) as the focus of these published studies was clear coat paint smears, not automotive paint chips. For this reason, the methodology used to develop the prefilters for the clear coat layer has been modified in this study to tackle samples containing multiple layers of automotive paint (clear coat, surfacer-primer, and e-coat layers). Furthermore, a new search prefilter to identify the manufacturer of the vehicle has been developed using the clear coat, surfacer-primer and e-coat layers to direct each unknown to the appropriate assembly plant search prefilter (Ford, Chrysler or General Motors).

The prototype pattern recognition library search system developed in this study for samples containing multiple layers of automotive paint consists of two separate but interrelated components: search prefilters developed from the clear coat, surfacer-primer and e-coat layers to reduce the size of the PDQ spectral library to a specific assembly plant and a cross correlation library search algorithm that utilizes both forward and backward searching to identify PDQ library spectra most similar to the unknown in the truncated IR spectral library identified by the search prefilters. Although the advantages associated with using multiple layers instead of a single layer of automotive paint to develop search prefilters were demonstrated in a previous study involving 89 automotive paint samples from a single manufacturer [11], the current study, which involves 1303 samples from three automotive manufacturers (Ford, Chrysler, and General Motors), will serve as a more rigorous test. Previously published studies undertaken in our laboratory using the cross correlation library search algorithm have been limited to a forward search and a single layer of automotive paint [9,10]. The advantages of using both a forward and backward search as well as multiple layers of automotive paint to identify the make, line, and model of an automotive vehicle are demonstrated in the current study.

To ascertain the line and model of a vehicle from an intact paint chip, search prefilters have been developed from the clear coat, surfacer-primer and e-coat layers for 1179 manufacturer paint systems within a limited production year range (2000–2006) to differentiate automotive paint samples by manufacturer. For each manufacturer, search prefilters were formulated to identify the assembly plant of the vehicle. A cross correlation library search algorithm that performs both forward and backward IR searching was used to identify the line and model of the vehicle from the truncated IR spectral library identified by the assembly plant search prefilters. Only the results for Ford are reported here.

To perform spectral library searching, the cross correlated spectra were divided into windows. For the forward search, the original spectra were also divided into three regions. The top five hits identified in each search window and each region were compiled, and a histogram was computed that summarized the frequency of occurrence (which was weighted based on the average similarity index of the spectra across all windows and all regions) for each selected library sample. The five library samples with the highest frequency of occurrence across all windows and regions comprised the final hit list.

A similar procedure for windowing was used to perform a backward search except that the original spectra were divided into twenty-two equally spaced intervals and the frequency of occurrence was computed for the vehicle line and model without regard to the identity of the individual spectra. Only those lines and models with a frequency of occurrence greater than or equal to 20% were included in the final hit list. When there was agreement between the two searches, the specific line and model common to both hit lists was always the correct assignment. Samples assigned to the same line and model by both searches correlated well on an individual basis to a specific library sample and were well represented in the truncated PDQ spectral library identified by the search prefilters. To benchmark the performance of the prototype library search system (modified to accommodate multiple layers of automotive paint), we compared the prototype system to a commercial library search algorithm.

## 2. Experimental

IR transmission spectra of 1303 manufacturer's automotive paint systems from the PDQ database were collected using a Bio-Rad 40A, Bio-Rad 60A or Thermo-Nicolet 6700 FTIR spectrometers. All FTIR spectrometers were equipped with a DTGS detector. The FTIR spectrometers were operated at 4 $cm^{-1}$ resolution. Each spectrometer employed a beam condenser: a Harrick 4× beam condenser for the two BioRad instruments and a Harrick 6× beam condenser for the two Thermo Nicolet 6700 FTIR spectrometers.

The paint systems of the three major manufacturers (clear coat, surfacer-primer and e-coat layers of Ford, General Motors and Chrysler vehicles) spanned 54 assembly plants located in North America. For each PDQ library sample, 3 or 4 micrograms of the clear coat, surfacer-primer or e-coat layers were compressed between two diamond anvils. The helium neon laser frequency assigned to the Thermo Nicolet 6700 FTIR spectrometer was 15,798.0 $cm^{-1}$ whereas the frequency assigned to the BioRad 40A or BioRad 60A instruments was 15,798.3 $cm^{-1}$. Differences in the helium neon laser frequencies assigned to each spectrometer led to differences in the number of data points in each FTIR spectrum. For this reason, all FTIR spectra were normalized to the helium neon laser frequency of 15,798.0 $cm^{-1}$ using OMNIC (Thermo-Nicolet). After normalization, each IR spectrum (400 $cm^{-1}$ to 4000 $cm^{-1}$) consisted of 1869 points. Further details about the collection of the FTIR transmission spectra in PDQ using diamond anvil cells can be found elsewhere [12].

## 3. Pattern recognition methodology

### 3.1. Search prefilters

Classifiers were developed using IR transmittance spectra of the fingerprint region (1640–667 $cm^{-1}$) of the clear coat and the two undercoat layers (surfacer-primer and e-coat). Each layer was vector normalized to unit length, and the discrete wavelet transform [13] was applied to each vector normalized IR spectrum using the Symlet 6 mother wavelet at the 8th level of decomposition (8Sym6). For each sample, the wavelet transformed spectra of the three layers were horizontally concatenating together into a single data vector in the order of clear coat, surfacer-primer, and e-coat. For each layer, the wavelet coefficients were ordered (A1, D1, A2, D2, …, A8, D8) where A1 is the vector of first level approximations and D1 is the vector of first level details and so forth.

Wavelet coefficients characteristic of the manufacturer or assembly plant of the vehicle were identified by a genetic algorithm (GA) for pattern recognition analysis [14,15]. The fitness function of the pattern recognition GA emulates human pattern recognition through machine learning to identify a set of wavelet coefficients that optimize the separation of the classes (e.g., assembly plants) in a plot of the two or three largest principal components of the wavelet transformed spectral data. The pattern recognition GA is able to focus on those classes and/or samples that are difficult to classify as it trains by providing this information to a perceptron algorithm which increases the relative importance (i.e., weights) of the classes/samples which are consistently misclassified. Over time, the pattern recognition GA learns its optimal parameters in a manner similar to a neural network. Further details about the fitness function of the pattern recognition GA used in this study can be found elsewhere [16–21].