Standoff detection and classification procedure for bioorganic compounds by hyperspectral laser-induced fluorescence

Thomas Fischbach[1], Frank Duschek[1], Anita Hausmann[1], Carsten Pargmann[1], Valeri Aleksejev[2], Larisa Poryvkina[2], Innokenti Sobolev[2], Sergey Babichenko[2], Jürgen Handke[1]

[1]Institute of Technical Physics, German Aerospace Center, Langer Grund, 74239 Hardthausen, Germany;
[2]LDI Innovation OÜ, 12 Lohu, 12618 Tallinn, Estonia

ABSTRACT

The high and still increasing number of attacks by hazardous bioorganic materials makes enormous demands on their detection. A very high detection sensitivity and differentiability are essential, as well as a rapid identification with low false alarm rates. One single technology can hardly achieve this. Point sensors can collect and identify materials, but finding an appropriate position is time consuming and involves several risks. Laser based standoff detection, however, can immediately provide information on propagation and compound type of a released hazardous material. The coupling of both methods may illustrate a solution to optimize the acquisition and detection of hazardous substances.

At DLR Lampoldshausen, bioorganic substances are measured, based on laser induced fluorescence (LIF), and subsequently classified. In this work, a procedure is presented, which utilizes lots of information (time-dependent spectral data, local information) and predicts the presence of hazardous substances by statistical data analysis. For that purpose, studies are carried out on a free transmission range at a distance of 22m at two different excitation wavelengths alternating between 280 nm and 355 nm. Time-dependent fluorescence spectra are recorded by a gated intensified CCD camera (iCCD). An automated signal processing allows fast and deterministic data collection and a direct subsequent classification of the detected substances. The variation of the substance parameters (physical state, concentration) is included within this method.

Keywords: Biological sensing and sensors; Fluorescence, laser-induced; Spectroscopy, ultraviolet; Standoff detection; Hazardous material detection; Classification procedure

1. INTRODUCTION

An increasing amount of attacks in public of varying dimensions like e.g. the disposal of the neurotoxin sarin in a subway in Tokyo 1995[1] or the transmission of anthrax to american government officials in 2001[2] show the need for an effective and fast solution to detect chemical and biological (CB) hazardous substances. Also unintended releases of such material, caused by earthquakes, industrial accidents, or floods may lead to a high risk for the public. A quick identification of the released material and the tracking of a potential aerosol cloud is essential to initiate the right counteractions and to minimize the risk for the public and emergency services. The potential risk of self-replication of bioorganic substances also demands a high sensitivity for low amounts of such aerosol particles.

Laser based techniques are applicable for standoff detection of hazardous material from secure areas with a target distance up to the kilometer range. Laser induced fluorescence (LIF) is one of the possible methods and provides the ability to map and classify aerosol clouds. In the case LIF cannot identify the substance, the gained cloud distribution parameters can be taken for an optimized positioning of point sensors, which may identify the substance.

The LIF technique is based on the fluorescence behaviour of molecules after being excited by laserlight. Electromagnetic waves within specific wavelength ranges (in our case UV) excite the molecule to higher energy levels.

As the molecule relaxes to lower energy states, fluorescence light is emitted in a broad spectral range with lower photon energies than those of the excitation photons. The fluorescence lifetimes vary from a few Nanoseconds up to the magnitude of 50 ns. The broad spectra can lead to ambiguous results because of the potential indistinguishability of the spectra. Therefore, additional discrimination features like spectra optained from different excitation wavelengths together with time-resolved measurements (fluorescence lifetime measurements) can be applied.[3–6]

In order to operate a LIF system under real outdoor conditions several effects have to be taken into account, such as natural surroundings like pollen, dust, and diesel, which can interfere with the substance to be investigated. Also different weather conditions (solar radiation, fog, rain) affect the measurements by affecting the laser light propagation and interfering with the fluorescence light. The system has to be robust, compact and eye-safe for operations within public areas (laser wavelength below 400 nm).[7,8]

To satisfy these requirements the LIF system is operated on a free space optical testrange which offers the possibility to measure CBE substances at distances from 20 m up to 135 m under different weather conditions.[3,9] Laser pulses with two eye-safe wavelengths at 280 nm and 355 nm are used to excite the target molecules. Time-resolved spectra of different chemical and biological substances (in fluid and aerosol form) are captured by a gated intensified CCD (iCCD) camera. After background correction further analysis is done by a pattern recognition software which classifies the substances into discrete classes (chemical, oil, plant, biological). The following work describes the technical details of the LIF system, discusses measured spectra and outlines the classification process.
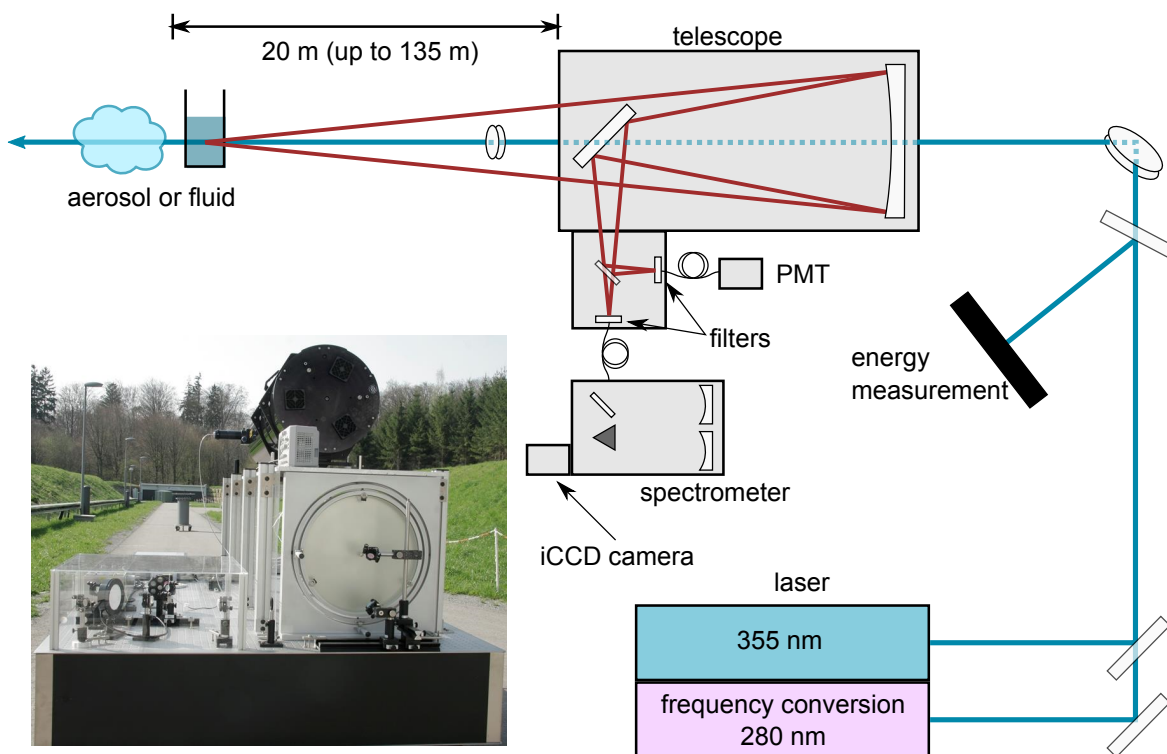
2. EXPERIMENTAL SETUP

Figure 1: Schematic draft of the optical setup. Details are described in the text. The left side shows a picture of the free laser transmission range with a view from the laser system in the direction of the target. The laser transmission range has a total length of 135 m.

The standoff detection system is operated on an outdoor free transmission range at the German Aerospace Center Lampoldshausen to simulate realistic conditions. Measurements can be performed from a distance of
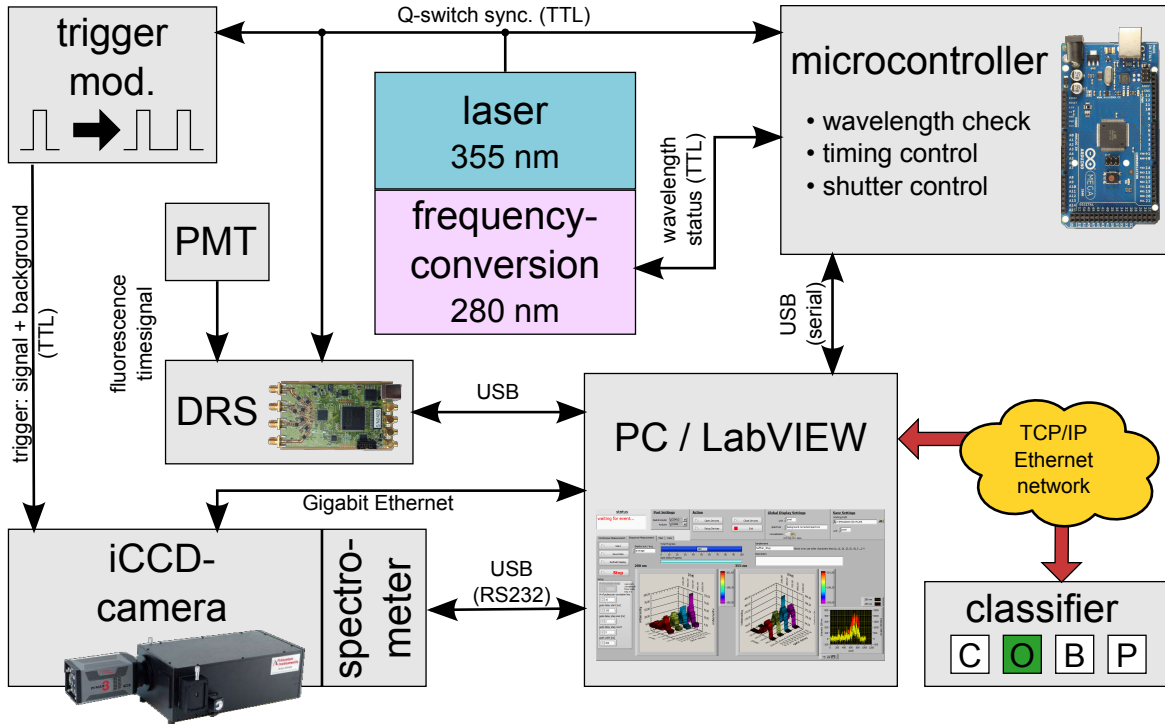
Figure 2: Schematic draft of the electronic setup. Details are described in the text.

20 m up to 135 m. The laser system is operated indoor whereas the target consists of a sprayed aerosol or a solution in a cuvette and is positioned outside along the free transmission range.

2.1 Optics

Fig. 1 represents the schematic optical setup of the system. The picture in the left bottom shows a view from the laser system in the direction of the target. 7 ns laser pulses are emitted by a Nd:YAG laser with a repetition rate of 10 Hz, a pulse energy of about 10 mJ and a wavelength of 355 nm. Each second laser pulse is frequency converted to 280 nm, so that each wavelength is emitted with an effective repetition rate of 5 Hz. The laser pulses of both wavelengths are guided through a set of mirrors to the target placed at a distance of 22 m. The isotropic emitted fluorescence light is collected by a Newton telescope with an optical diameter of 400 mm. The collected light is spectrally filtered to supress the laser lines and splitted into two parts. One part is fiber-coupled into a spectrometer with a resolution of 1 nm and a spectral range from 300 nm to 600 nm for the spectral analysis of the fluorescence light. The spectrum is captured by a gated iCCD camera. The second part of the fluorescence light is fiber-coupled into a photomultiplier tube (PMT) to get a wavelength-independent time signal of the fluorescence pulse. Details of the data acquisition are described in Section 2.3.

2.2 Electronics

The electronics of the system is shown in Fig. 2 as a schematic draft. The laser acts as a master trigger for the whole system. Each trigger pulse opens the capture gate of the iCCD camera for the response fluorescence signal of the target. To be able to substract a background spectrum from each fluorescence spectrum a second trigger pulse is generated ("trigger mod.") approximately 50 ms after each laser pulse. The iCCD camera and the spectromenter are controlled by a LabVIEW[10] program via Gigabit ethernet and USB/RS232 respectively. The time-resolved fluorescence signal which is captured by the PMT is analysed and integrated by a digital oscilloscope (DRS[11]), which is connected to the LabVIEW PC via USB. For a proper timing of all signals a microcontroller provides information about the currrent timestamp within the period between two laser pulses.
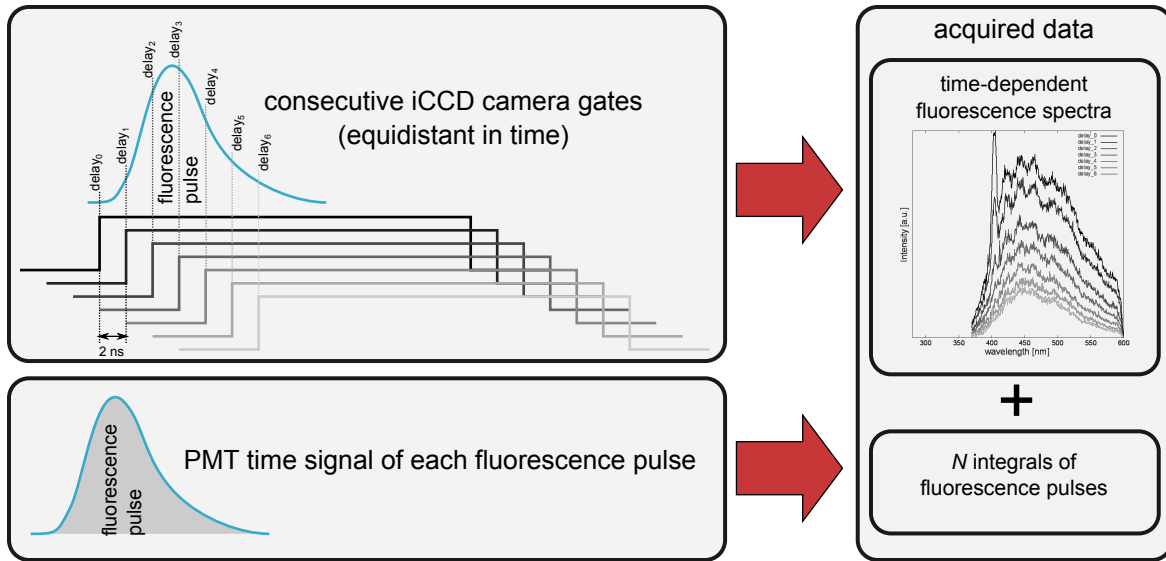
Figure 3: Data acquired during a measurement. Spectra are captured with consecutive iCCD camera gates with linear increasing gate delay. A dataset consists of N subsets containing a fluorescence spectrum, a background spectrum and a fluorescence pulse integral.
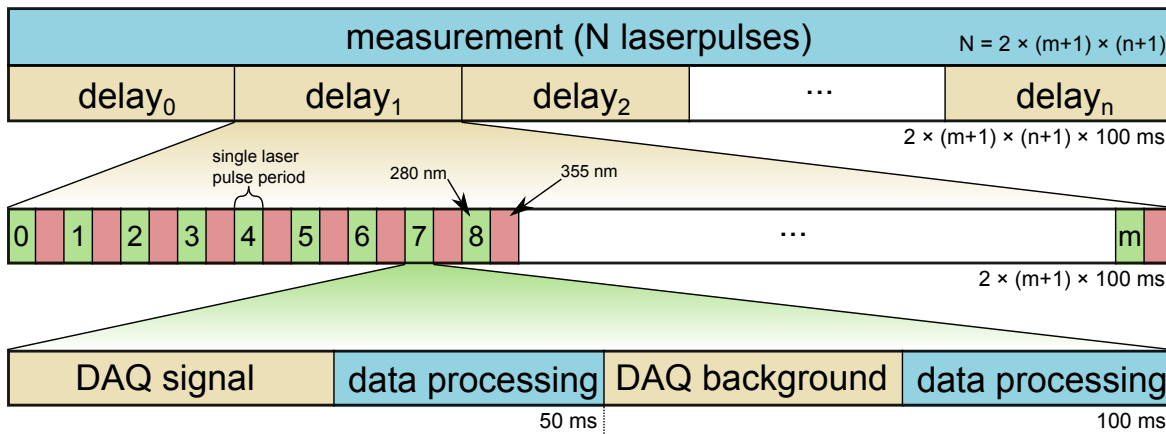

Figure 4: Measurement timing diagram. One measurement consists of $n + 1$ camera gate delays. Each delay contains $2 \cdot (m + 1)$ laser pulses and the respective fluorescence spectra. Because of a laser repetition rate of 10 Hz, each period between two laser pulses has a duration of 100 ms, which is splitted into two equally sized parts: fluorescence spectrum and background spectrum acquisition.

It also monitors the current wavelength state of the laser system and controls a shutter which is only open during the measurement to preserve the target sample from unnecessary bleaching effects due to laser irradiation. The LabVIEW PC is connnected to the classifier software via Ethernet (TCP/IP) to be able to classify the substances online and location-independent.

2.3 Acquired data and measurement timing

Fig. 3 shows the acquired data considered for later classification, whereas Fig. 4 represents the timing structure of a measurement. Fluorescence spectra are captured in consecutive time shifted iCCD camera gates. This additional time information (depending on the fluorescence lifetime of the substance) provides another important feature for the substance discrimination. The first gate delay starts at an offset to the laser Q-switch, placed at

the fluorescence pulse leading edge. The offset depends on the propagation time of the light (target distance) and the electronic signals. The following camera gate delays are shifted one by one by 2 ns to provide the fluorescence lifetime feature. Parallel to the acquisition of the spectrum the complete fluorescence pulse is captured in time and integrated, which represents the signal strength of each spectrum and can be used for a later normalization of the spectra. The individual normalization with a time-independent value helps to reproduce signal strengths depending on the camera gate delay. For each delay several $(1 - 100)$ single spectra are captured to supress statistical effects. Each single spectrum is linked to an individual background spectrum which is captured directly after the corresponding spectrum acquisition and within the second half of the period between two laser pulses which is 100 ms. The substraction of this directly following background spectrum makes each measurement independent of the background radiation which - especially in free atmosphere - may fluctuate on time scales of seconds.

Finally the measurement dataset contains

$$N = 2 \cdot (m+1) \cdot (n+1) \tag{1}$$

subsets, each consisting of a fluorescence spectrum, a background spectrum, the integral value of the fluorescence pulse time signal and meta data like the current excitation wavelength, where $n + 1$ is the camera gate delay count, $m + 1$ is the spectrum accumulation within each delay and the factor 2 represents the two different excitation wavelengths. The duration of one measurement is

$$t_{\text{meas}} = N \cdot 100 \, \text{ms}, \tag{2}$$

or e.g. $t_{\text{meas}} = 8\,\text{s}$ for 4 camera gate delays and an accumulation of 10 spectra. An overhead of appoximately $2\,\text{s}$, caused by post-data processing, data communication and classification, has to be added to get the effective classification time.
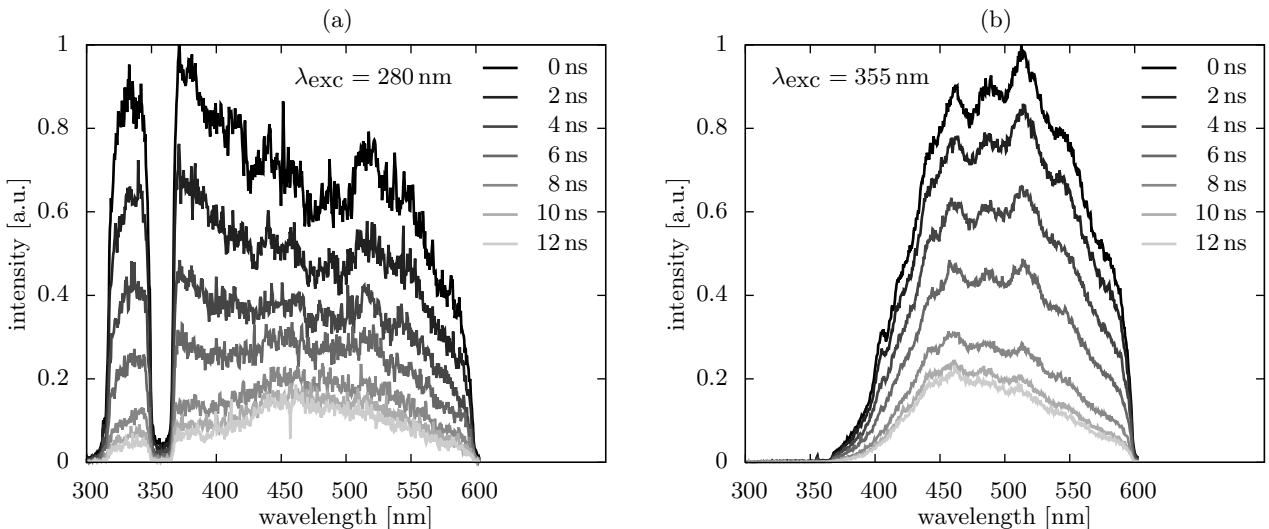
3. MEASUREMENTS

Figure 5: Fluorescence spectra of yeast (solution with a concentration of $1\,\text{mg/ml}$) with an excitation wavelength of 280 nm (a) and 355 nm (b). The different spectra in one plot represent different camera gate delays measured relatively to the offset. The dip between 350 nm and 370 nm is caused by a notch filter which blocks the 355 nm light from the laser. The 280 nm laser light is outside the bandwidth of the captured spectra (300 nm to 600 nm).

For training the system and building up a spectral database substances contained in the groups fungi, bacteria, vitamins, enzymes and aromatic amino acids are measured in liquid solutions. To include interfering natural
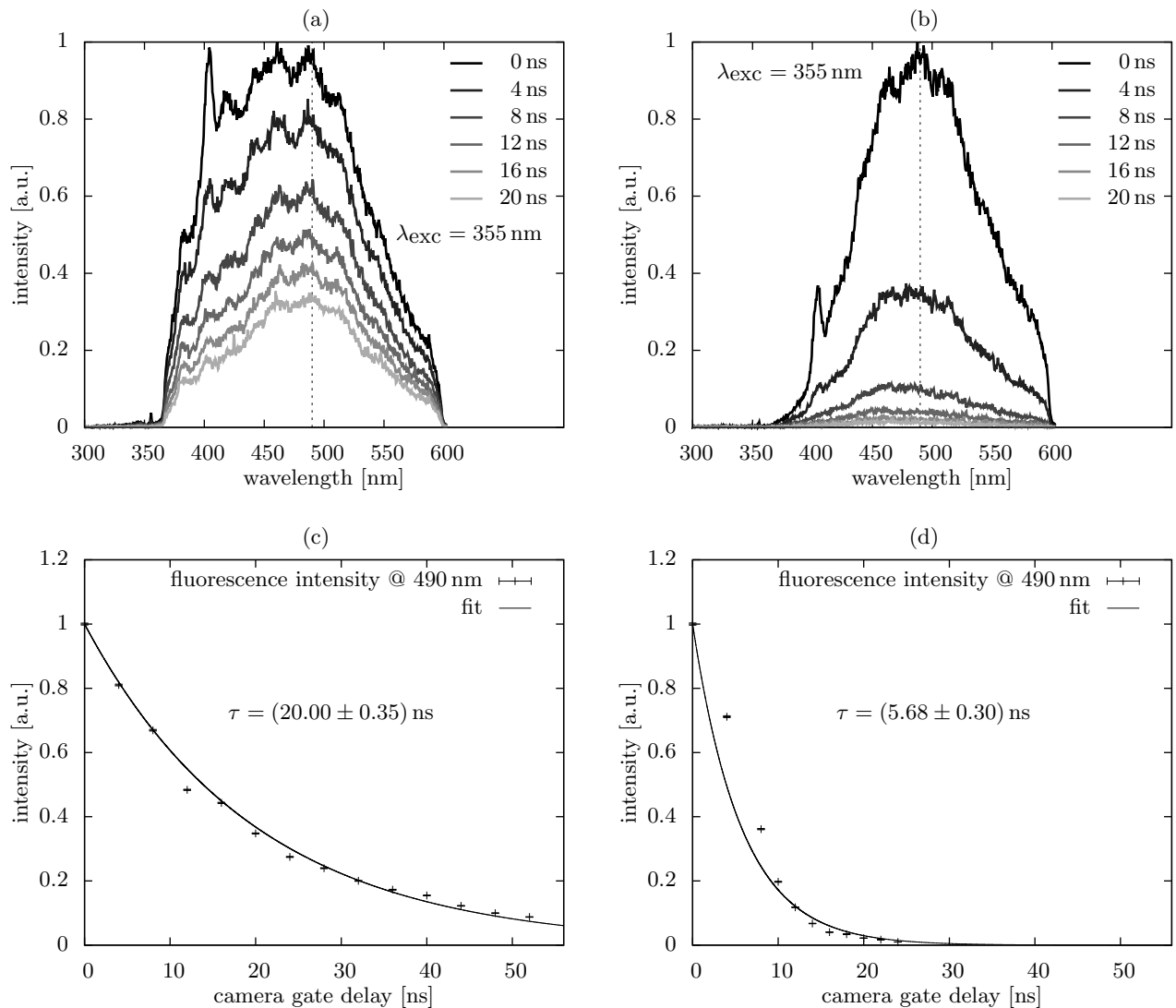
Figure 6: Fluorescence spectra of diesel (a) and dandelion (b) (solutions with a concentration of $2.5\,^{\mu l}/_{ml}$ and $500\,^{\mu g}/_{ml}$ respectively) excited by a wavelength of $355\,nm$. The spectra represent different camera gate delays measured relatively to the offset. Plot (c) and (d) show the lifetime of diesel and dandelion respectively at $490\,nm$.

background substances different plants like dandelion or saffran are also measured. Oils like diesel or petrol complete the set of substance groups as non-dangerous anthropogenic substances. As solvent deionized water is used in different concentrations. Some substances are not (completely) soluble in water. To keep the solutions homogeneous every substance is stirred during the measurements. For the database each measurement consists of 100 spectrum accumulations to supress statistical effects. To check the reproducibility each sample of a substance is measured three times. The following spectra are background-corrected but not unfolded from the spectral properties of the optical parts of the system. So the real spectra of the substances may look significantly different. For the substance classification and discrimination on this special apparatus this circumstance is negligible.

Fig. 5 shows the fluorescence spectra of $1\,^{mg}/_{ml}$ yeast in deionized water with the excitation wavelengths $280\,nm$ (Fig. 5a) and $355\,nm$ (Fig. 5b) and camera gate delays from offset $+\,0\,ns$ to offset $+\,12\,ns$. The

different spectra of the different excitation wavelengths show the improvement of the ability to discriminate different substances in contrast to the use of only one excitation wavelength. Fig. 6 shows an example of the discimination of substances by their fluorescence lifetime. Diesel and dandelion do not have big differences in their fluorescence spectrum shapes. But diesel (Fig. 6a and Fig. 6c) has a lifetime of about 20 ns whearas dandelion (Fig. 6b and Fig. 6d) has a lifetime of about 6 ns which is a big difference. So lifetime - which may also be dependent on the observed wavelength - can be a good feature to discriminate substances. The laser pulse width is about 7 ns, which limits the lifetime method. So the real lifetime of dandelion at 490 nm differs from the calculated value.

4. CLASSIFICATION AND RESULTS

The main objective of the spectral and temporal analysis is the creation of an algorithmic system which is able to strongly discriminate substances into disjunct classes. Currently the system provides four classes: chemicals, plants (natural surroundings), oil, and bacteria (living material). The last class represents potential hazardous bioorganic material. There are different methods, which can be used for the classification. In our current system we are using a structural extraction combined with statistical classification for the spectral data analysis. Another method is the Principle Component Analysis (PCA) which has been e.g. implemented by the Swedish Defence Research Agency (FOI).[4]
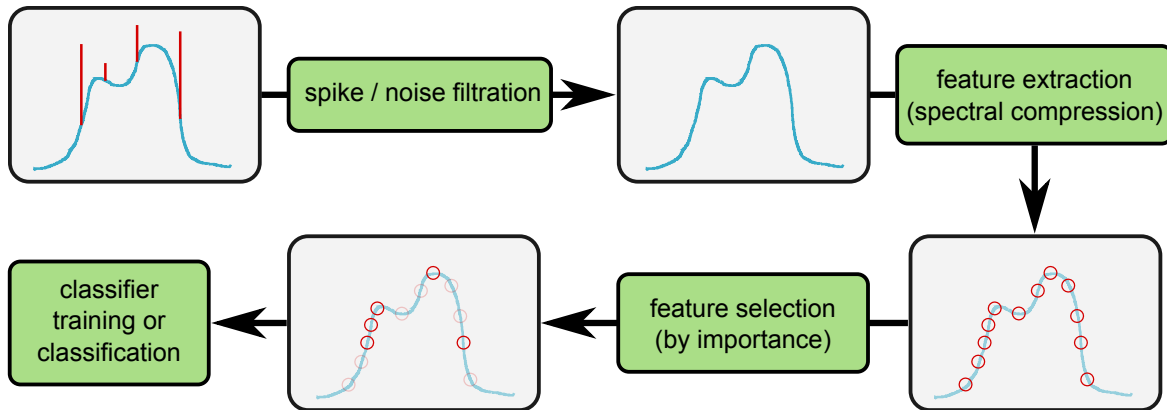

Figure 7: Analysis/Classification process.

In Fig. 7 the analysis/classification process is shown. First of all artificial spikes and noise is filtered out of the spectrum. The original spectrum contains 720 datapoints. This dimensionality (number of datapoints or features) is redundant for an efficient classification process. Therefore a reduction of the dimensionality to a few significant features is done by structural and statistical methods (called "feature extraction").[12–14] After this process the data is compressed down by a factor of ≈ 60. Each remaining feature can be seen as a representative of a spectral region within the original spectrum. It is possible to reconstruct the original spectrum from the compressed dataset with good accordance. An excellent performance of such a hyperspectral feature exctraction was already demonstrated.[15] Subsequently important features, which contribute significantly to the discrimination process, are selected.

The next goal in the classification process is to generate many weak or "okay" classifiers, which can be combined to one strong classifier.[16] A classifier is created by growing a decision tree on a set of features. Each node of a decision tree is a binary decision on the value of a feature. The leafs represent the substance classes in which the analyzed spectrum is put depending on the decisions which are made on its feature values. To be able to grow many decision trees (classifiers) substantially more sets of features are required than were provided by the measured spectra. The so called "bootstrap aggregation" method (bagging[17]) helps to produce different replicas from one feature set. Each set is generated by randomly picking N features from a set of the size N with replacement. The strong classifier is now built by voting among all decission trees which are grown on the

substance	concentration	expected class	correct classified	correct and unique classified
NADH	$30\,\mu\text{g}/\text{ml}$	Bacteria	100 %	100 %
NADH	$60\,\mu\text{g}/\text{ml}$	Bacteria	100 %	100 %
NADH	$125\,\mu\text{g}/\text{ml}$	Bacteria	100 %	100 %
Tryptophan	$30\,\mu\text{g}/\text{ml}$	Bacteria	100 %	100 %
Tryptophan	$60\,\mu\text{g}/\text{ml}$	Bacteria	100 %	100 %
Tryptophan	$125\,\mu\text{g}/\text{ml}$	Bacteria	100 %	100 %
Yeast	$250\,\mu\text{g}/\text{ml}$	Bacteria	50 %	30 %
Yeast	$500\,\mu\text{g}/\text{ml}$	Bacteria	90 %	90 %
Yeast	$1000\,\mu\text{g}/\text{ml}$	Bacteria	70 %	70 %
DEET	$0.5\,\mu\text{l}/\text{ml}$	Chemical	70 %	60 %
DEET	$1.0\,\mu\text{l}/\text{ml}$	Chemical	90 %	80 %
DEET	$2.5\,\mu\text{l}/\text{ml}$	Chemical	90 %	90 %
RAID	$1.0\,\mu\text{l}/\text{ml}$	Chemical	100 %	90 %
RAID	$5.0\,\mu\text{l}/\text{ml}$	Chemical	100 %	100 %
Diesel	$2.5\,\mu\text{l}/\text{ml}$	Oil	100 %	100 %
Diesel	$5.0\,\mu\text{l}/\text{ml}$	Oil	100 %	100 %
Diesel	$10.0\,\mu\text{l}/\text{ml}$	Oil	100 %	100 %
Petrol	pure	Oil	100 %	100 %
Curcumin	$2.1\,\mu\text{g}/\text{ml}$	Plant	100 %	100 %
Curcumin	$4.2\,\mu\text{g}/\text{ml}$	Plant	100 %	100 %
Curcumin	$8.3\,\mu\text{g}/\text{ml}$	Plant	100 %	100 %
Dandelion	$250\,\mu\text{g}/\text{ml}$	Plant	90 %	70 %
Dandelion	$500\,\mu\text{g}/\text{ml}$	Plant	100 %	40 %
Saffran	$30\,\mu\text{g}/\text{ml}$	Plant	100 %	100 %
Saffran	$60\,\mu\text{g}/\text{ml}$	Plant	100 %	100 %
Saffran	$125\,\mu\text{g}/\text{ml}$	Plant	100 %	100 %
Mean			$(94 \pm 12)\,\%$	$(89 \pm 19)\,\%$

Table 1: Classification of some substances with different concentrations. Each sample was measured 10 times. The percentage values show how often the sample was classified "correct" or "correct and unique", respectively.

bootstraped datasets. The bagging method is implemented by the "Framework for Ensemble Learning" from the "Matlab Statistics Toolbox".[18]

In Tab. 1 classification results of some test samples are listed. Each sample was measured 10 times to get the success rate (in percentage) of a correct classification. "correct classified" means a correct classification but the result may be ambiguous, whereas "correct and unique classified" represents the rate of nonambiguous results. The classification algorithm yields a classification confidence value for each measurement and class. Confidence values can be in the interval $[0, 1]$. A sample is regarded as classified if the classification confidence value is larger than the defined threshold value of 0.4. Most samples are classified correctly with an average rate of $(94 \pm 12)\,\%$. Yet, more improvements have to be done on the classification process to increase the overall success rate. For a correct classification, confidence values have been found to be typically larger than 0.6.

5. CONCLUSION AND OUTLOOK

To be able to classify CB substances by using a laser based standoff technique a LIF system has been developed which is optimized to operate under realistic outdoor conditions. A broad set of CB substances (soluted in deionized water with different concentrations) are measured to build up a database for training the system's pattern recognition algorithm. For the excitation process the samples are repetitively illuminated by two different eye-safe wavelengths at $280\,\text{nm}$ and $355\,\text{nm}$. Time resolved measurements of the fluorescence spectra increase

the dimensionality of information for the substance discrimination. A pattern recognition software classifies the material into four disjunct classes by applying binary decision trees on automatically selected important features of the measurement data. First realistic evaluations show promising results and prove the qualification of the LIF technique for the classification of (hazardous) CB substances. LIF can provide useful information for subsequent identification techniques and counteractions.

As an outlook an extension of the substance portfolio is planned, which demands for a new training of the pattern recognition algorithm. Furthermore with a larger database the granularity of the classes can be increased, e.g. by splitting them up into different living organisms. Higher standoff distances are also aimed in the near future.

REFERENCES

[1] Tu, A. T., "Overview of sarin terrorist attacks in japan," *Natural and Selected Synthetic Toxins* **745**, 304–317 (2000).

[2] National Research Council, [*Review of the Scientific Approaches Used During the FBI's Investigation of the 2001 Anthrax Letters*], The National Academies Press, Washington, DC (2011).

[3] Hausmann, A., Duschek, F., Fischbach, T., Pargmann, C., Aleksejev, V., Poryvkina, L., Sobolev, I., Babichenko, S., and Handke, J., "Standoff detection: classification of biological aerosols using laser induced fluorescence (LIF) technique," *Proc. SPIE* **9073**, 90730Z–90730Z–8 (2014).

[4] Meyer, O., Jacquelard, C., Melkonian, J., Chardard, P., Lanson, P., and Petitgas, D., "Stand-off biological detection by LIF (laser induced fluorescence) LIDAR," *Optronics in Defence and Security, 4th International Symposium* (2010).

[5] Sivaprakasam, V., Lin, H.-B., Huston, A. L., and Eversole, J. D., "Spectral characterization of biological aerosol particles using two-wavelength excited laser-induced fluorescence and elastic scattering measurements," *Opt. Express* **19**(7), 6191–6208 (2011).

[6] Mierczyk, Z., Kopczyński, K., Zygmunt, M., Wojtanowski, J., Młynczak, J., Gawlikowski, A., Młodzianko, A., Piotrowski, W., Gietka, A., Knysak, P., Drozd, T., Muzal, M., Kaszczuk, M., Ostrowski, R., and Jakubaszek, M., "Fluorescence/depolarization lidar for mid-range stand-off detection of biological agents," *Proc. SPIE* **8037**, 80371J–80371J–9 (2011).

[7] Handke, J., Duschek, F., and Pargmann, C., "Gefahrenerkennung mit Lichtgeschwindigkeit," *Laser Technik Journal* **7**(4), 43–46 (2010).

[8] Wallin, S., Pettersson, A., Östmark, H., and Hobro, A., "Laser-based standoff detection of explosives: a critical review," *Analytical and Bioanalytical Chemistry* **395**(2), 259–274 (2009).

[9] Duschek, F., Pargmann, C., Grünewald, K., and Handke, J., "Stand-off detection at the DLR laser test range applying laser-induced breakdown spectroscopy," *Proc. SPIE* **7838**, 78380I–78380I–6 (2010).

[10] LabVIEW, [*version 2012*], National Instruments Germany GmbH, Munich, Germany (2012).

[11] Ritt, S., *DRS4 Evaluation Board, User's Manual* (2014).

[12] Bruce, L., Koger, C., and Li, J., "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction," *Geoscience and Remote Sensing, IEEE Transactions on* **40**(10), 2331–2338 (2002).

[13] Mojaradi, B., Abrishami-Moghaddam, H., Zoej, M., and Duin, R., "Dimensionality reduction of hyperspectral data via spectral feature extraction," *Geoscience and Remote Sensing, IEEE Transactions on* **47**(7), 2091–2105 (2009).

[14] Wu, K. and Hsieh, P., "Empirical mode decomposition for dimensionality reduction of hyperspectral data," *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International* **2**, 1241–1244 (2005).

[15] Sobolev, I. and Babichenko, S., "Analysis of the performances of hyperspectral lidar for water pollution diagnostics," *EARSeL eProceedings* **12**(2), 113–123 (2013).

[16] Breiman, L., "Random forests," *Machine Learning* **45**(1), 5–32 (2001).

[17] Breiman, L., "Bagging predictors," *Machine Learning* **24**(2), 123–140 (1996).

[18] MATLAB, [*version 7.15 (R2011a)*], The MathWorks Inc., Natick, Massachusetts (2011).