STATE OF THE ART HYPERSPECTRAL DATA ANALYSIS AND PROSPECTS WITH RESPECT TO AI

ROSUS 2019, 21/03/19 Maribor Dr. Matthias Kerschhaggl
1. Presentation EVK

2. State of the art HSI data analysis

3. AI Prospects

4. Industrial apps using chemical imaging

5. Conclusions
WHO WE ARE

EXPERT COMPANY
Industrial imaging solutions for sorting and inspection

CORE EXPERTISE
Hyperspectral, color and conductivity imaging technologies

COMPLETE SOLUTIONS
From raw data acquisition to decision making
EVK AT A GLANCE

TARGET MEKETS

DEPLOYED PRODUCTS
(kumuliert)

EVK TEAM

Management & Administration

Research & Development

Production

Marketing & Sales

40% F&E
17% M&V
23% Fertigung
20% G&V

TURNOVER / REGION

94% EMEA
5% Americas
2% APAC

© EVK DI Kerschhaggl GmbH
INDUSTRIAL SOLUTIONS

PROCESS ANALYTICS

SMART SORTING

MONITORING AND INSPECTION
HSI – HOW DOES IT WORK

WAVELENGTH (m)

10^{-11}  10^{-9}  10^{-7}  10^{-5}  10^{-3}  10^{-1}  10^{1}  10^{3}

GAMMA RAYS  X-RAYS  ULTRA VIOLET  INFRARED  MICROWAVES  RADIO FREQUENCY

FREQUENCY (s^{-1})

10^{20}  10^{18}  10^{16}  10^{14}  10^{12}  10^{10}  10^{8}  10^{6}  10^{4}

1.0 \mu m  2.3 \mu m

HELIOS NIR/SWIR
HSI – HOW DOES IT WORK

C-O-H-N molecular combinations

Absorption spectra according to the molecular structure

stretching  bending  rocking

stretching  twisting  wagging
NIR Reflektionsspektren unterschiedlicher Kunststoffe

HSI – HOW DOES IT WORK

Seeing Spectroscopy Clearly™

© EVK DI Kerschhaggl GmbH
HSI – HOW DOES IT WORK

Camera internal classification and RGB visualization of chemical differences for real-time analysis
Hyperspectral data and its challenges in industrial applications

- High dimensionality ("curse of dimensionality")
- Limited training samples
- Mixed spectral signatures
- Pre-Processing
- Real-time and inline suitable classification algorithms
Curse of dimensionality

- Data coverage decreases with dimensionality ("empty space")
- Data distributed mainly on the outskirts of data space
- Differences in min and max distances of data sets become marginal, i.e. distance measures lose meaning (e.g. knn search)
- Subspace methods (projections) can help!
- Manifold hypothesis
STATE OF THE ART HSI DATA ANALYSIS

Mixed spectral signatures

<table>
<thead>
<tr>
<th>Material</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.50</td>
</tr>
<tr>
<td>B</td>
<td>0.25</td>
</tr>
<tr>
<td>C</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Methods: MCR, LU, ICA

M. E. Klein et al

M. Kamal et al

© EVK DI Kerschhaggl GmbH

www.evk.biz
STATE OF THE ART HSI DATA ANALYSIS

Spectral Preprocessing

- Spectral Scatter: Norm, SNV, MSC
- High Frequency Noise: Savitzky-Golay Filtering
- Baseline correction
- Spectroscopic Feature Localisation: Derivative Spectroscopy

Ä. Rinnan et al
Real-time and inline suitable classification algorithms

- Subspace Methods:
  - PCA, PLS, ICA, LDA
- Cluster Methods:
  - k-means, kNN, RDF
- And many more (SVM etc.)…
STATE OF THE ART HSI DATA ANALYSIS

Real-time and inline suitable classification algorithms

- **Subspace Methods:**
  - **PCA**, PLS, ICA, LDA
- **Cluster Methods:**
  - k-means, kNN, RDF

© EVK DI Kerschhaggl GmbH

17 www.evk.biz
Real-time and inline suitable classification algorithms

- **Subspace Methods:**
  - PCA, PLS, ICA, LDA
- **Cluster Methods:**
  - k-means, kNN, RDF

ICA finds directions which maximize independence (using higher order statistics)

www.sci.utah.edu
Real-time and inline suitable classification algorithms

- **Subspace Methods:** PCA, PLS, ICA, **LDA**
- **Cluster Methods:** k-means, kNN, RDF

\[
J(w) = \frac{\|\mu_1 - \mu_2\|^2}{\bar{s}_1^2 + \bar{s}_2^2} = \frac{w^T S_B w}{w^T S_W w}
\]

Hence \(J(w)\) is a measure of the difference between class means (encoded in the between-class scatter matrix) normalized by a measure of the within-class scatter matrix.
Real-time and inline suitable classification algorithms

- **Subspace Methods:** PCA, PLS, ICA, LDA
- **Cluster Methods:** *k*-means, kNN, RDF
Real-time and inline suitable classification algorithms

- **Subspace Methods:** PCA, PLS, ICA, LDA
- **Cluster Methods:** k-means, kNN, RDF

Problematic for high dimensional spaces!
Real-time and inline suitable classification algorithms

- Subspace Methods:
  PCA, PLS, ICA, LDA
- Cluster Methods:
  k-means, kNN, RDF
Problems to be solved:

- HSI-data are high dimensional
- Usually few labelled data
- Feature extraction is often difficult, tedious and operator dependent.
- Vital information is lost in current data reduction methods due to neglection of image features.

http://robertsy.github.io
The promise of deep learning approach:

- Combination of spectral and spatial domains addressed in a single approach.
- Powerful and reliable computing hardware now available.
- Strong scientific evidence and technology penetration (e.g. CNN, SAE etc.) in leading high-tech companies.

Chen et al.
**DEEP LEARNING – A PROMISING APPROACH FOR HSI**

- **HSI cube**
- **DEEP LEARNING architecture**
- **embedded hardware**
- **classified objects**

**HELIOS**

http://visal.cs.cityu.edu.hk
DEEP LEARNING – A PROMISING APPROACH FOR HSI

Benefits:

- Best performing classification results in current benchmark computer vision problems -> application efficiency in food apps?
- Seamless classification model generation without operator variance -> cost reduction, process reliability
- New customer benefits via big data (mining) and predictive analysis/maintenance.

http://konwersatorium1-ms-pjwstk.blogspot.co.at
Ongoing activities using deep neural network classification with HELIOS data

Current results demonstrate a strong potential of this technology (reduced test times, superior classification power)

Further R&D projects related to deep learning are planned.
WE BRING THE LAB INTO YOUR LINE

From at-line to in-line quality control
INDUSTRIAL APPS

GROUND TRUTH UNDERSTANDING

DATA ACQUISITION

DATA ANALYSIS

WHAT DOES IT TAKE

TESTING AND EVALUATION

SYSTEM INTEGRATION & DEPLOYMENT

IN-FIELD EVALUATION
Refuse Derived Fuels

STATUS: FIELD TESTS
Humidity vs. Time in RDF stream: Lab vs. QCI HELIOS.
DRY MATTER DETERMINATION

STATUS: FULL PRODUCTION
DRY MATTER DETERMINATION

Dry substance model

\[ R = 0.94929 \]
\[ RMS = 5.4129 \]
RANCID WALNUTS

STATUS: MARKET INTRODUCTION
RANCID WALNUTS

Heat map walnuts

Graphical user interface
IN-LINE MEASUREMENT OF AMYGDALIN IN BITTER ALMONDS

STATUS: MARKET INTRODUCTION

### IN-LINE MEASUREMENT OF AMYGDALIN IN BITTER ALMONDS

#### Phenotypes in raw almonds:

**Non-bitter**

Sweet snacking almonds (nutty flavor)

**Bitter**

3-5% amygdalin content, cyanide aroma, hazardous to health

<table>
<thead>
<tr>
<th>Sample#</th>
<th>Amygdalin [mg/g]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBA 1</td>
<td>4,6</td>
</tr>
<tr>
<td>SBA 2</td>
<td>3,7</td>
</tr>
<tr>
<td>SBA 3</td>
<td>3,9</td>
</tr>
<tr>
<td>SBA 4</td>
<td>63,3</td>
</tr>
<tr>
<td>SBA 5</td>
<td>58,6</td>
</tr>
<tr>
<td>SBA 6</td>
<td>57,2</td>
</tr>
<tr>
<td>SBA 7</td>
<td>66,3</td>
</tr>
<tr>
<td>SBA 8</td>
<td>55,5</td>
</tr>
<tr>
<td>SBA 9</td>
<td>4,8</td>
</tr>
<tr>
<td>SBA 10</td>
<td>4,1</td>
</tr>
</tbody>
</table>
IDENTIFICATION OF MYCOTOXINS IN FIGS

PRELIMINARY
PARTNER PROJECT

- Aflatoxins are toxic carcinogenic secondary metabolites produced by fungal species
- Great variety of crops is affected
- They are an increasing problem due to climate change
- This leads to huge economic losses

PRELIMINARY
IDENTIFICATION OF MYCOTOXINS IN FIGS

Results

Tests

PRELIMINARY
DETECTION OF WOODEN BREAST DEFECT

- Increasing prevalence in broiler production
- Considerable economic losses
- Currently manual inspection/hand sorting

Ref: http://www.helsinki.fi/food-and-environment/research/groups/meat.html

STATUS: FULL PRODUCTION
DETECTION OF WOODEN BREAST DEFECT

Bad product: Wooden breast filets

Good product: Ordinary chicken filets

Chemical colour classified stream:
API CONCENTRATIONS

STATUS: FULL PRODUCTION
At-line quality control of pharmaceutical products

Chemical concentrations traced by spectral signatures, spatially resolved.

In-situ inference of quantitative values possible at bulk flow velocities of several m/s without destroying the product for every camera image pixel.

Figure 2: Left: pharmaceutical sponge with active substance on its surface. Centre: quantitative ‘chemical colour’ representation of concentration. Right: zones (squares) with sufficient and insufficient concentrations.
EVK Software Tools for multivariate measurements of CHEMICAL COMPOSITIONS

Brings the LAB INTO your LINE

For use IN-LINE in combination with EVK Helios camera systems
Currently used HSI data analysis methods are spectroscopy motivated and for the most part ignore imaging information.

AI seems to be a good approach to overcome both the limitations of high dimensionality and the somewhat historical segregation of spectral and spatial information.

Quantitative Chemical Imaging (QCI) using modern algorithms opens a new window into the world of hyperspectral applications.
REFERENCES

- H. Bischof, TU Graz, “Machine Learning and Hyperspectral Imaging” (talk chii 2016)