Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-20 (2016) 055–060

Big Data Solution for Quality Monitoring and Improvement on Flat Steel Production ⋆

Jens Brandenburger * **Valentina Colla** ** **Gianluca Nastasi** **
Floriano Ferro *** **Christoph Schirm** **** **Josef Melcher** ****

* *VDEh-Betriebsforschungsinstitut GmbH, BFI, Düsseldorf, Germany (e-mail: jens.brandenburger@bfi.de).*
** *Scuola Superiore Sant'Anna, SSSA, Pisa, Italy (e-mail: colla {nastasi}@sssup.it)*
*** *ILVA S.p.A, Novi Ligure, Italy (e-mail: floriano.ferro@gruppoilva.com)*
**** *thyssenkrupp Rasselstein GmbH, Andernach, Germany (e-mail: christoph.schirm@thyssenkrupp.com)*

Abstract: A new big-data handling paradigm for quality monitoring and improvement in the flat steel production is presented based on the efficient exploitation of high resolution (HR) measuring data. Based on a multi-scale data representation over multiple production stages this new concept provides simple and fast HR data access. Realized as a three-tier software architecture including a web-service for a standardized data access the potential of this new concept could be remarkably proven within two industrial pilot applications. This includes an application for data visualisation that enables the in-coil aggregation of millions of quality and process measures within seconds and an application for advanced cause-and-effect analysis based on HR data.

Keywords: Data mining, Artificial intelligence, Hot/cold rolling, Database structures, Information technology, Quality control, Plantwide, Big Data, Industry 4.0, Material Tracking

1. INTRODUCTION

Promoted by the telecommunication and commercial market analysis the "Big-Data" paradigm has gained a lot of popularity within the last years. From the technology point of view Big-Data is commonly defined by 4 Vs: **V**olume, **V**elocity, **V**ariability and **V**eracity.

However from the viewpoint of manufacturing industries it is more important how to combine data with production knowledge instead of processing huge amounts of data in real-time. In this context the focus must be set on the usability of Big-Data and not only on the technological limits of data processing. Therefore in (Freytag, 2014) the concept of the domain expert is introduced who demands 3 Fs to Big-Data applications: **F**ast, **F**lexible and **F**ocused.

Also the German ITC industry association BITKOM states that the aim of Big-Data is not to create vast data pools, but to generate economic benefit (Bartel et al., 2012).

Therefore within this paper a solution is presented that tries to maximize data usability already before storage by means of a multi-scale data representation. Based on this new approach two industrial application were realized and the results are given in this paper.

⋆ The research leading to these results has received funding from the European Community's Research Fund for Coal and Steel (RFCS) under grant agreement n° RFSR-CT-2012-00040.

2. PROBLEM DEFINITION

Today the production of high quality steel is supported by modern measuring systems gathering an increasing amount of high resolution (HR) quality and process data along the complete flat steel production chain. All these measurements can be uniquely assigned to a position on the coil surface and categorized into 3 different types of data:

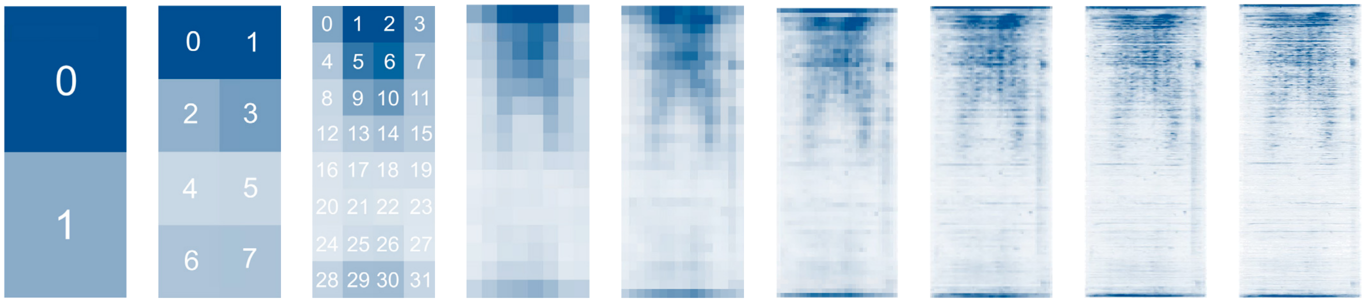
- **1D-continuous** - pairs $(p_{md}, v) \in \mathbb{R}^2$ with coil length-position p_{md} and measurement value v (e.g. strip tension, width, speed, etc)
- **2D-continuous** - triples $(p_{cd}, p_{md}, v) \in \mathbb{R}^3$ with coil width-position p_{cd}, length-position p_{md} and measurement value v (e.g. thickness, flatness, temperature, coating layer, etc)
- **Event-based** - 5-tuples $(p_{cd}, p_{md}, l, w, c) \in \mathbb{R}^4 \times \mathbb{N}$ defining a rectangular regions at position (p_{cd}, p_{md}) with length l and width w and an assigned class c (e.g. Surface defects, Internal defects, Manual inputs, etc)

Due to the continuously increasing resolution of the installed measuring devices a relevant volume of data per coil is generated during the full flat-steel production process.

Usually for further processing this kind of information is aggregated based on constant length segments (e.g. 1 m) and stored in a factory-wide quality database (QDB). Thus

10.1016/j.ifacol.2016.10.096

Table 1. Multigrid representation of "number of surface defects" over 7913 coils



Stage 0	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7	Stage 8
$n = 2$	$n = 8$	$n = 32$	$n = 128$	$n = 512$	$n = 2048$	$n = 8192$	$n = 32768$	$n = 131072$

already today common applications based on the available data in a QDB can monitor the current production, support quality decisions or allow dedicated investigations in case of customer claims (Brandenburger et al., 2014).

However for quality monitoring and improvement of the production processes a more statistical view on the data is mandatory. The absolute coil-position is of minor importance and a normalized view on the data is needed to compare not only single coils but full production cycles and material groups regarding suspicious data distributions.

Furthermore if the cause of a problem and the problem detection itself are located at different production steps (as often the case for surface defects) it has to be possible to cross borders and combine data of subsequent production steps including all transformations applied to the coil positions within the production process (rotation, lengthening, cutting, welding). Presuming adequate performance this is not feasible using classical QDB data models as they are optimized for a fast per-coil access to the data.

3. HR DATA MODEL

To find a suitable data model for efficient HR data access it is useful to address the problem from the visualisation point of view. To create a 2-dimensional HR data visualisation of one single coil the HR data has to be represented on the screen as an image with pixel matrix $I := [0, N_x] \times [0, N_y]$. Furthermore as well N_x as N_y should be chosen independently of the real coil dimensions to present each coil to the user in the same way.

The Image I can be seen as a constant grid and the pixels $p_{x,y} \in I$ as disjoint grid cells, each of them representing a rectangle $R_{x,y}$ on the coil that is located relative to each pixel position in I. For HR data visualisation the color (resp. value) of each pixel represents an aggregation of the available HR data over $R_{x,y}$.

It is easy to see that a data model that stores only the position x, y and the aggregated value $p_{x,y}$ for a given grid I provides any information that can be displayed by the described setup. On the other hand this data model is independent of type and resolution of the measurement and e.g. the amount of per-coil surface defects detected by an automatic surface inspection system (ASIS).

By means of a bijective function

$$\mu : [0, N_x] \times [0, N_y] \rightarrow [0, N_x N_y] \quad (1)$$

a unique `TileID` $:= \mu(x, y) \in [0, N_x N_y]$ can be assigned to each grid cell. This unique `TileID` again can be used to aggregate grid data over multiple coils by simply aggregating the data of equal `TileID`s.

The approach described above is fast as it is exactly fitting the requirements emerging from the data visualisation. On the other hand this causes a lack of flexibility as some applications might get along with less resolution. Furthermore for the user-acceptance a solution is preferable that makes the user feel that the system is reacting instantaneously leading to a constraint of less than 0.1 seconds response time (Nielsen, 1993) which is difficult to achieve using one fixed high-resolution grid.

One existing approach for a fast interactive access to huge amounts of data can be found in the context of geographic mapping services as described in (Tanner et al., 1998) and similar used by modern rendering engines as applied by the virtual globe "Google Earth". The idea is to store the data in different resolutions simultaneously to enable fast access to any required resolution. Such an approach is highly efficient and parallelizable for multiple tiles. Each tile processes the requested data simultaneously and shows the coarsest available resolution first before waiting for the desired resolution. Therefore each tile displays immediately something before the full resolution appears. This leads to a high user acceptance of such an approach since the time to wait for a response is very short.

A combination of both above principals leads to a multiscale grid representation of measurement data as shown in Table 1. It shows a visualisation of the number of surface defects of a certain class for multiple resolution stages. In the first three stages additionally the `TileID` is shown whereas for the higher resolution grids only the total number of grid cells is given.

Using (1) each tile of each stage is addressable by a pair (`Stage`, `TileID`) leading to a trivial synchronization of data coming from different coils and different measuring systems as again simply the data of equal `TileID`s has to be aggregated for each stage. Hence this aggregation can be realized using only one SQL-query for a complete grid making this approach outstanding fast.

Table 2 shows the final grid definition chosen for the full data model. Grid resolutions of adjacent stages are multipliers of 2 leading to unique tile associations across different stages. Because of unequal length to width ratio