Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Outlier Detection for 2D Temperature Data

Jani Posio*, Kauko Leiviskä*, Jari Ruuska*, Paavo Ruha**

*University of Oulu, Control Engineering Laboratory, P.O. Box 4300,
FIN-90014 University of Oulu, Finland
(e-mail:kauko.leiviska@oulu.fi)
**Rautaruukki Oyj, Ruukki Production, Raahe, Finland

Abstract: This paper reports the study of using 2D temperature data for analysing the operation of the cooling process in the steel strip mill. Scanning pyrometers are producing data profiles of the strip in longitudinal and transversal directions. Instrument malfunctions, dust and dirt particles on the strip surface and other disturbances make the use of the measurements difficult. This makes the data pre-processing, and especially the outlier detection, of utmost importance for a reliable process and fault analysis.

1. INTRODUCTION

Data pre-processing is perhaps the most important stage in data analysis and it has remarkable effects on the availability and reliability of data. If carelessly done, it can mitigate all the results from the analysis. Knowing the data to be analysed is also the key for understanding the results and their utilisation.

Industrial data is often in large databases, including both direct process measurements and calculated values. Unmeasured disturbances together with process and instrument failures contaminate data, and it represents different operating points that may be difficult to trace afterwards. Timing is also a permanent problem, especially if the data is coming from separate data collection systems. This calls for different data pre-processing methods that in many cases depend on the problems – they require process knowledge.

Outliers are observations that deviate significantly from the majority of observations (Liu et al., 2004), and also effect on statistical properties (the mean and the variance) of the data, if included in calculations. One main aim of data pre-processing is the detection of outliers, removing them and replacing them with better data. This should occur with simultaneously preserving the original data structure and without loosing information as falsely detected outliers. In practice, however, the result is a compromise between the mentioned requirements.

The most detection methods start from the assumption of identically and independently distributed data (i.d.d), where the mean and variance describe the statistics of the data (Liu et al., 2004). The methods are considered later in this paper, but the Hampel identifier is regarded as one of the most efficient methods. According to Liu et al. (2004) there is also a group of methods based on assumed process models utilising maximum likelihood methods and Kalman filter. They also present an on-line filter-cleaner that combines the properties of outlier detection methods and on-line filters. Also wavelet transformation has been used in on-line filtering and outlier detection (Nounou and Bakshi, 1999).

This paper reports the study of using 2D temperature data for analysing the operation of the cooling process in the steel strip mill. Scanning pyrometers are producing data profiles of the strip in longitudinal and transversal directions and the data describes the efficiency of the process and helps in the early detection of possible process faults. Instrument malfunctions, dust and dirt particles on the strip surface and other disturbances make the use of the measurements difficult. This makes the data pre-processing, and especially the outlier detection, of utmost importance for a reliable process and fault analysis.

This paper concentrates on the most difficult and crucial point: the outlier detection using some reliable criteria. The paper gives a comparison of several methods and discusses their advantages and disadvantages.

The outline of the paper is as follows: Next chapter discusses the methods for outlier detection and then follows the introduction to the cooling process in the steel strip mill together with the measurement data and the technical problems in question. Next comes the comparison of these methods and, finally, some conclusions are given.

2. METHODS

The target was to find a method that is sensitive enough, but robust, for finding the outliers. The methods can be classified to global or local, to one- or two-dimensional and adaptive methods. Some of them are classical signal processing methods and some come from the image processing. Some process the two-dimensional profiles as rows or columns, some deal with the whole profile globally or inside a small window, a mask. In two-dimensional algorithms, the windows can be distinct or sliding with one value at time over the whole range. The last alternative causes the biggest computational burden. All the abovementioned options were tested and compared with each other.

The tested signal processing methods were (1) 3-sigma method, (2) a method based on robust estimates, and (3) an outlier detection procedure based on wavelet coefficients. Four image processing methods were also tested, namely: (1) applying the threshold of the one-dimensional histogram of the profile, (2) the two-dimensional clustering of the probability distribution describing the transformations between the data points, (3) the k-means clustering of the profile intensity, and (4) a two-dimensional Hampel filter. For some methods, testing concerned with both global and local versions.

2.1 3-sigma Method

A commonly used method for outlier detection is to look for observations that deviate more than three times the standard deviation from the mean. This is a conventional "3σ edit rule" (Pearson, 2002), which is based on the assumption of the normal distribution. In this case, the probability of point x_i lying more than three standard deviations from the mean is 0.3 %. The algorithm proceeds as follows

- Calculate the mean, m, and the standard deviation, s
- For each observation x_i, calculate the scaled variable

$$z_i = \left| \frac{x_i - m}{s} \right| .$$

- If $z_i>3$, x_i is an outlier.

If outliers are present, this method includes a basic difficulty; outliers lead to biased estimates for both the mean and the standard deviation (Pearson, 2002). Even a single outlier is enough to bias the mean, and the situation gets worse with multiple outliers, especially if they are located on the same side of the mean. Outliers also increase the standard deviation. This means that this algorithm does not find the outliers, but instead of it, robust scaling methods must be used. These methods estimate the parameters of the "normal data", and the problem returns to defining this normal data set. The algorithms resemble the 3-sigma method, but the mean and the standard deviation result from this normal data set.

2.2 *Robust Estimates*

One usual way is to replace the mean with the median and the standard deviation with the median absolute deviation from the median (MAD). The median of a data sequence is obtained by ranking the observations from smallest to largest,

$$x(1) \le x(2) \le \dots x(N)$$

And then taking the middle value (odd case) or the average of the middle two values (even case) (Pearson, 2002). The calculation of MAD estimate proceeds as follows:

• Replace the mean with the median, median(x)

• Calculate the absolute difference for each observation from the median

$$d_i = \left| x_i - median(x) \right|$$

• Estimate the standard deviation using the median of d_i

$$s = 1.4826 median(d_i) ,$$

where the constant 1.4826 is required to make MAD an unbiased estimate of the standard deviation for Gaussian data (Chiang et al., 2004). If 3σ edit rule is used with MAD estimate, the method is usually called the Hampel identifier. This method is robust to multiple outliers and it is also computationally inexpensive. One of its weaknesses is the assumption of symmetrical distributions. This leads to unfavourable results with skewed distributions.

Hampel filter is a simple data cleaning filter which operates inside a moving data window, applies the Hampel identifier and in which the decision threshold is defined as follows (Pearson, 2002)

• If $d_i > ts; 2 \le t \le 5$, x_i is an outlier.

In this case, the outlier can be replaced by the median of the data window. This filter has two tuning parameters: the half width of the moving window, K, and the decision threshold, t. Higher the parameter t, less outliers are detected. If t=0, we have the median filter (Pearson, 2002), i.e. each value is replaced by the corresponding median.

In this paper, testing includes both local and global algorithms in one- and two-dimensional cases. Local version uses a sliding window. In the one-dimensional case, it could slide over the data column by column or row by row. The two-dimensional version uses a mask of a certain size and slides over the whole data one value at time. Global version calculates the median and MAD-estimate for the whole data and uses it in defining the global threshold. The tuning uses the above mentioned parameter, t, and in local version also the size of the window.

Generally speaking, the local version is efficient, when there are a lot of outliers, but their amount in the window does not exceed the breaking point of the median, 50% of the values. Global version works well with many outliers and reasonably stationary data. Testing also includes an adaptive threshold based on which part of the data is processed. This was to minimise the effect of non-stationarity.

2.3 *Wavelets*

A discrete Haar-wavelet was applied to a one-dimensional case. The values of wavelet coefficients are supposed to increase when the outliers exist. When hunting for impulse or pulse changes, thresholds for wavelet coefficients are set to correspond to fast changes in the signal. Equation th = σ(2log(n)1/2, where σ is the standard deviation of wavelet coefficients at the level and n is the number of coefficients, gives the threshold. MAD-estimate is one way to calculate the standard deviation. The algorithm proceeds as follows (Bilen and Huzurbazar, 2002).:

1. Apply the wavelet transform to the data vector

2. Calculate σ for the wavelet coefficients:

σ = mean(| D1(k)-median(D1) |), where D1(k) is the wavelet coefficient at k.

3. Calculate the threshold th = $\sigma(2\log(n))^{1/2}$.

4. Search for indices S = $\{s1,\ldots,s_m\}$, where |D1(k)| > th.

5. Use the indices to locate the outliers in the data vector: Supposing s_k to be the index for the coefficient where |D1(k)| > th, the outlier is located at $2s_k$ or $2s_k - 1$. To find the exact location, calculate the mean of the original data, $mean_{orig}$, without $2s_k$ or $2s_k - 1$. The outlier is at $2s_k$, if the condition, $|origdata(2s_k) - mean_{orig}| > |origdata(2s_k-1) - mean_{orig}|$, holds, otherwise the outlier is at 2_{sk}-1.

2.4 *Histogram threshold*

Histogram threshold is a usual method in image processing to separate image objects and deviating values. Several methods exist to calculate the threshold that divides the frequency distribution in two or more areas. In this case, temperature values were divided into 50 bins. A bimodal distribution follows, where the bigger area corresponds to normal temperature values and the smaller one to outliers. Low-pass filtering and min/max-transformation improve the data and emphasize the bimodality of the distribution (Gonzalez and Woods, 2002).

Triangle-algorithm is the usual way to calculate the threshold for the histogram (Fig. 1) First, the minimum, non-zero frequency point in the histogram is located (point close to 200 °C in the figure) together with the corresponding maximum (close to 600 °C in this case). Next, these points are connected with the line. Finally, the maximum distance, d, from this line to the tops of the histogram is defined. The corresponding point in the temperature axis, TH, is the threshold value. This algorithm works usually well, even with a smaller secondary top in the histogram. This method is global and may give weak results with non-stationary profiles (Zack et al., 1977).

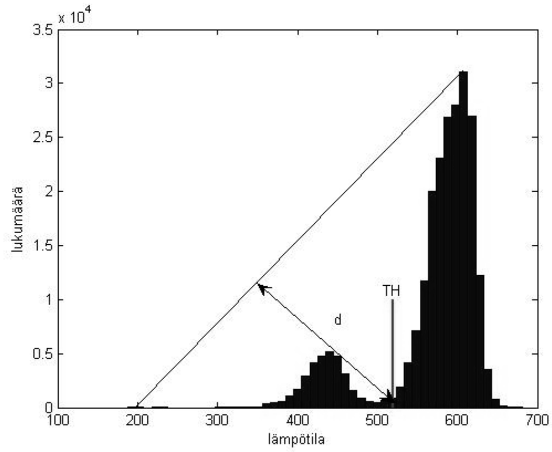

Fig. 1. Basic principle of the Triangle algorithm.

2.5 *Clustering of the two-dimensional frequency distribution*

Image processing uses also a two-dimensional frequency distribution to separate the image from its background. In this case, the two-dimensional frequency distribution describes the transformations in temperature in longitudinal direction for successive measurements in the data. Fig. 2 shows an example, where frequency of transformations between two successive points in the profile is calculated. These frequencies have been clustered with k-means clustering algorithm into four groups. The clusters in the main diagonal represent cases where two successive points have a similar temperature. The clusters in the decreasing diagonal, on the other hand, represent cases, where the temperature between two successive points has strongly changed, i.e. either decreased or increased. The clusters in the main diagonal correspond to areas where the intensity has been constant for a longer time (either high or low). Off-diagonal clusters correspond to cases, where the intensity has changed fast; i.e. the possible outliers. (Haddon and Boyce, 1990, Corneloup et al., 1996),

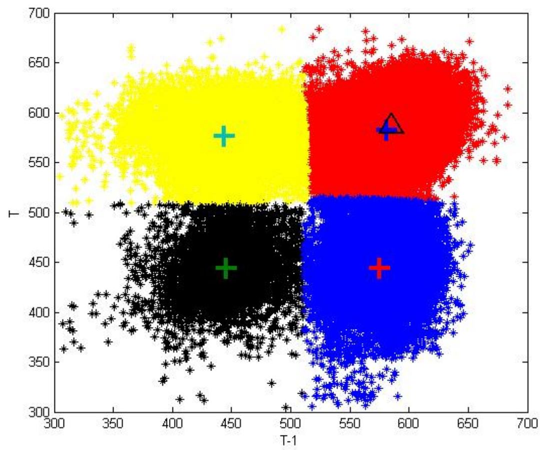

Fig. 2. Clustering of temperature transformations in longitudinal direction.

2.6 *k-means clustering*

Also the common k-means clustering was applied for intensity levels (Shapiro and Stockman, 2001). The number of clusters was 2; corresponding to normal values and outliers. The initial values for cluster centres were the median value for the "normal" cluster and the difference between the median for the maximum deviations and the profile median.

2.7 *Block-wise histogram threshold, (histogram-blockth)*

Histogram threshold does not have to be applied to entire profile but can be used on a region by region basis. A variation was developed in which the M x N profile is divided into non-overlapping regions. The profile was divided in 5 blocks in transversal and 3 in longitudinal direction, 15 blocks per profile. The target was to handle head, middle and tail parts of the strip separately and eliminate the transversal trend. In each block a threshold is calculated using the same Triangle-algorithm as before.

2.8 *Block-wise Hampel filter (MAD-block)*

Also Hampel filter can operate on a block by block basis using the same before mentioned technique for .block division. Parameter t in the before mentioned robust estimate method was adapted according to the MAD estimate of the block in question. Small variance led to a higher parameter value and vice versa. Parameter t varied between 2 and 8.

3. PROBLEM DESCRIPTION

According to Fig. 3 the steel strip is cooled both from above and below after rolling by the cooling system at the run-out table. The cooling system comprises main cooling section with water curtains and trimming section with separate spray nozzles.

The temperature drop in the cooling is from 145 to 585°C. The thickness of the strip varies from 1.4 to 16 mm, its width from 740 to 1860 mm and the average speed from 2.4 to 12 m/s. The length of one strip varies from 65 to 1800 m.

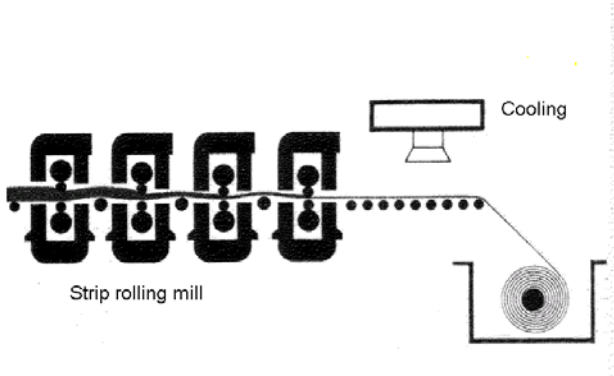

Fig. 3. Position of cooling in the steel mill.

Scanning pyrometers (SCOAP) measure the temperature before and after cooling with the measurement principle shown in Fig. 4. The sensor is above the steel strip and it scans cross the moving strip with the help of a rotating plane mirror. The resolution in the transversal direction is mm. The resolution in the direction where the strip is moving (longitudinal direction) depends, of course, on the speed of the strip and typically varies from 50 mm to 180 mm. This means that the strip with 93 m of length and 1506 mm of width produces the matrix approximately of 172000 temperature readings.

Fig. 5 shows an example of this kind of temperature matrix. The upper figure shows a 3D picture how the temperature varies in both directions. The lower picture is an intensity diagram of the same situation.

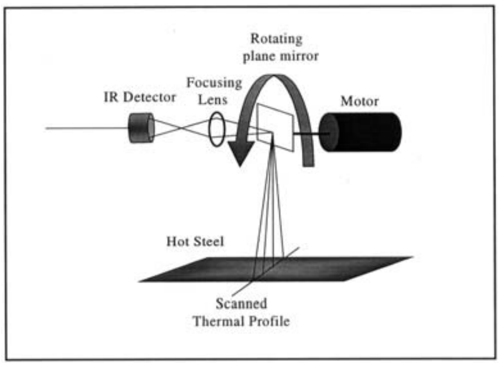

Fig. 4. Principle of scanning IR pyrometer.

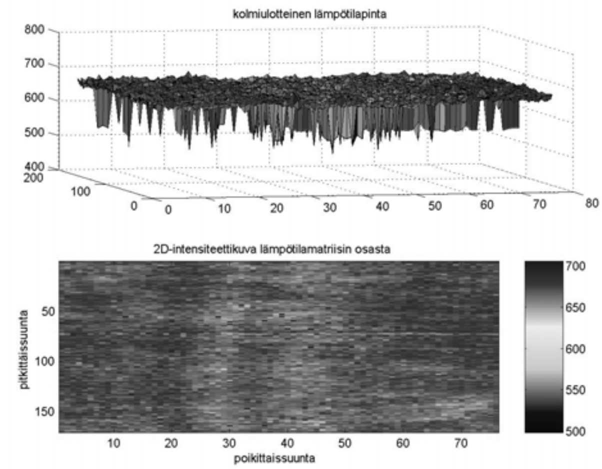

Fig. 5. A sample of the temperature matrix.

False scans cause a partial transversal row in the temperature matrix that shows considerable lower values than the neighbouring rows. Other outliers are single values or small group of values in 2D matrix. Single values could be easily removed, e.g. by median filter, but this does not work with several successive values. Filtering would also unnecessarily change the data structure.

The deviations in the temperature profiles are both transversal lines and single random points. Their frequency does not follow any order. They are short time, usually broadband, disturbances that are usually due to false scans. There is a strong gradient between deviations and normal values that is not even physically possible. There are more deviating values in the beginning (head) of the strip than in other parts. Deviations are nearly always less than the mean, causing negative skew and bimodality in the distributions.

Non-stationary profile is another feature that needs care in pre-processing. In this case, non-stationary profiles result from longitudinal and transversal trends together with random changes in the variance, e.g. in the beginning of the strip. This calls for trend removal or local pre-processing.

The main target of the study was to use the temperature profiles provided with the scanning pyrometers in analysing the efficiency of the cooling process and in diagnosing its operation. This target includes following sub-targets:

• Data pre-processing; outlier detection and straightening the profiles in the beginning and end of the strip and at the strip edges.

• Synchronising the temperature readings before and after the cooling process. This is necessary for building subtracted figure for efficiency analysis.

• Feature detection in order to produce features that best describe the operation of the cooling process.

• Visualisation of the results for the operator use.

• Packing the 2D temperature data.

Only the first point is considered here.

4. RESULTS

The testing used a profile with a big amount of outliers and trends both in transversal and longitudinal direction Also the variance in the head of the strip was bigger than in the rest of it. The accurate amount of outliers was unknown, but the basic value for comparisons came from approximate analysis of the bimodal histogram. The threshold was set to 510 ºC and this led to 28235 outliers, about 12.7 % of all values.

Table 1 compares the performance of different algorithms with the basic case shown above. Manual refers to the basic case and 3-sigma to the method shown in the previous chapter. MAD-global, MAD-block and 2DHampel calculate the robust estimates for median and variance utilizing different parts of the profile. Global algorithms calculate values for the whole profile and block algorithms for the certain blocks. 2DHampel calculates the values in a sliding 2D window for the all values in the profile. MAD-block and 2DHampel use the adaptive feature. Wavelet means the one dimensional wavelet algorithm. Other methods have been also explained before.

Table 1. Comparison of the outlier detection methods.

Method	**Number of outliers**	**Percentage of outliers**	**Mode**
Manual	28235	12,65	Global
3-sigma	3412	1,53	Global
MAD-global	28632	12,83	Global
MAD-block	28063	12,57	Local, adaptive
2Dhampel	20707	9,28	Local, adaptive
Wavelet	21307	9,55	Local
Histogram-th	28924	12,96	Global
Histogram-blockth	29538	13,23	Local
K-means	28599	12,81	Global
2D frequency distribution	28301	12,68	Global

The conventional 3-sigma algorithm does not find the outliers, when they exist in large amounts, because outliers bias the mean and standard deviation too much. Robust estimates work better in this case, both as local and global algorithms, and their performance is very close to the manual. The non-stationary profile, however, degrades their performance. Robust estimates work usually well with large amount of outliers and when data is symmetrically distributed around the median. With fewer outliers, and in non-stationary cases, also normal data can be classified as outliers. Fig. 6 shows the distribution of outliers found using robust estimates, and it is possible that also in this case some normal values have been classified as outliers.

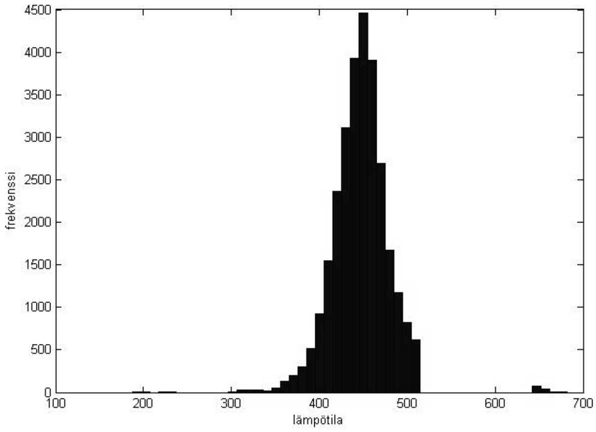

Fig. 6. The histogram of outliers found using global robust estimate's algorithm.

From adaptive algorithms, MAD-block version worked reasonable well, when the parameter t varied from 2 to 4 as a function of the estimated variance. 2DHampel could not find all outliers. In this case, the parameter t got the value 4, when the estimate for the variance was smaller than 20; otherwise it was 3. Also the wavelet algorithm missed a lot of outliers, especially in the case of successive outliers.

The global version of histogram threshold worked well. Strongly non-stationary profile can, however, have a negative effect on its performance. Fig. 7 shows the location of the threshold in this case. The histogram is clearly bimodal, and the definition of the threshold value (about 519 ºC) is easy.

To eliminate the effect of non-stationary profile a local version of the histogram threshold was tested. The algorithm works well, if the distribution is clearly bimodal. It could be improved by adding the bimodality test, which prevents the thresholding, when the distribution is not bimodal.

The method based on the clustering of 2D frequency distributions works also well. The clustering is successful, when there are no trends in data, local variations are small, and the transformations from one cluster to another are clear.

The results were already in Figure 6, where +-signs show the cluster centres and the triangle corresponds to the transformation with the biggest probability.

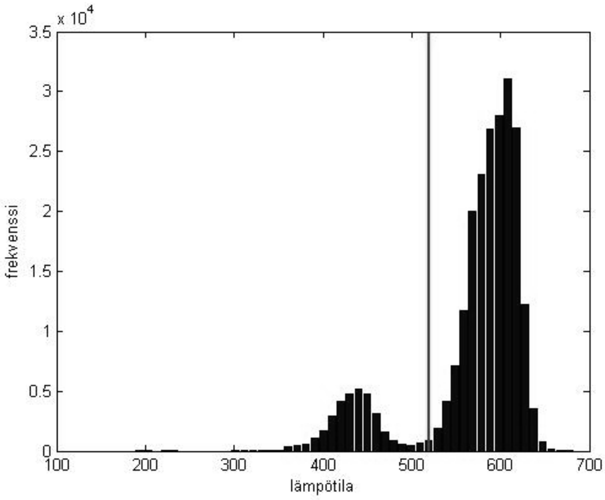

Fig. 7. The location of the threshold, when the Triangle-algorithm is used.

The conventional k-means clustering to two clusters, with good initial guesses, gave also good results. Also here, a stationary data with clearly separable clusters is easier to deal with.

5. CONCLUSIONS

The tested methods fall into several categories: global vs. local, one- or two-dimensional, and adaptive methods. Part of them are conventional signal processing methods – 3-sigma, robust estimates and wavelets – and part is coming from the image processing area – histogram threshold, 2DHampel filtering, profile intensity clustering with k-means algorithm and the method based on clustering of temperature transformations using 2D-histograms.

As a whole, the number of outliers and the possible non-stationary profiles effect on the performance of different algorithms. The breaking point of the median causes problems in robust algorithms; as also the variations of the mean in longitudinal and transversal directions, when global algorithms are used. As a conclusion, adaptation is required to guarantee efficient operation in changing process conditions. This, however, increases the computation load and makes the algorithms slower.

It was clear from the start that there is no single method that is superior in all situations. The selection proceeds case by case. Global robust method, or alternatively global histogram threshold, is the best choice for this case. Both methods are computationally efficient, and their performance is adjustable by the users. The user will have a possibility to check the results from outlier detection and change either the threshold value or the parameter t in the calculation of the MAD-estimate.

REFERENCES

Bilen C., Huzurbazar S., Wavelet-based detection of Outliers in Time Series. Journal Computational and Graphical Statistics, 11(2002)2, 311 327.

Chiang, L.H., Pell R.J., and Seasholtz, M.B., Exploring process data with the use of robust outlier detection algorithms. Journal of Process Control 13(2004) 437-449.

Corneloup G., Moysan J., Magnin I.E.: BSCAN Image Segmentation by Thresholding Using Co-occurrence Matrix Analysis, Pattern Recognition, 29(1996)2, 281-296.

Gonzalez R.C., Woods R.E., Digital Image Processing, Prentice Hall, New Yersey, 2002, 793 p.

Haddon J.F., Boyce J.F.: Image segmentation by Unifying Region and Boundary Information. IEEE Transactions on Pattern and Machine Intelligence, 10(1990)12, 929-948.

Liu, H.C., Shah, S., and Jiang, W., On-line outlier detection and data cleaning. Computers and Chemical Engineering 28(2004) 1635-1647.

Nounou, M.N., and Bakshi, B.R., On-line multilevel filtering of random and gross errors without process models. American Institute of Chemical Engineering Journal 45(1999) 1041-1058.

Pearson R.K.: Outliers in Process Modelling and Identification, IEEE Transactions on Control System Technology, 10(2002)1, 55-63.

Shapiro L.G., Stockman G.C., Computer Vision, Prentice Hall, New Jersey, 2001, 580 p.

Zack G.W., Rogers W.E., Latt S.A.: Automatic Measurement of Sister Chromatid Exchange Frequency, Journal of Histochemistry and Cytochemistry, 25(1977)7, 746-753.