REAL TIME SEARCH ALGORITHM FOR OBSERVATION OUTLIERS DURING MONITORING ENGINEERING CONSTRUCTIONS

Dorota Latos, Bogdan Kolanowski, Wojciech Pachelski, Ryszard Sołoducha

Military University of Technology in Warsaw,
The Faculty of Civil Engineering and Geodesy, Poland

Abstract

Real time monitoring of engineering structures in case of an emergency of disaster requires collection of a large amount of data to be processed by specific analytical techniques. A quick and accurate assessment of the state of the object is crucial for a probable rescue action. One of the more significant evaluation methods of large sets of data, either collected during a specified interval of time or permanently, is the time series analysis. In this paper presented is a search algorithm for those time series elements which deviate from their values expected during monitoring. Quick and proper detection of observations indicating anomalous behavior of the structure allows to take a variety of preventive actions. In the algorithm, the mathematical formulae used provide maximal sensitivity to detect even minimal changes in the object's behavior.
The sensitivity analyses were conducted for the algorithm of moving average as well as for the Douglas-Peucker algorithm used in generalization of linear objects in GIS. In addition to determining the size of deviations from the average it was used the so-called Hausdorff distance. The carried out simulation and verification of laboratory survey data showed that the approach provides sufficient sensitivity for automatic real time analysis of large amount of data obtained from different and various sensors (total stations, leveling, camera, radar).

Keywords: *time series analysis, construction monitoring, automation of data analysis*

1. Introduction

One of the important issues of today's activities in various fields of science, engineering, economy, etc., is data collection, compilation and presentation of results. It is mathematical statistics which deals with these problems at the most common. In the modern sense mathematical statistics allows to analyze the quality of data collected, as well as to evaluate them in conditions of uncertainty. One of the

common types of measurements are measurements of the same quantity made within a specified period or made permanently. The obtained data, especially at their large amount, are subjected to an analysis known as the analysis of time series as has been shown (Chatfield,2013; Machura, 2012).

Any engineering structure is subject to various deflections both during the construction and during exploitation. Besides of normal factors (static forces), such as gravity load, natural factors (wind, water pressure, the instability of the ground, and others), mostly depending on the type of construction, there are specific situations in which there are unpredictable factors, which lead to rapid and catastrophic consequences (such as floods, fires, terrorist attacks, etc.) that may require specified reaction within crisis management (Skłodowski, 2009).

A significant factor of modern measuring instruments is their possibility of numerical control of the measurements and of processing their results. This is important for monitoring engineering structures, especially during the emergency response actions in real time. This aspect of data processing is considered very essentially in the presented here system for monitoring engineering structures (Baszkiewicz, Kolanowski et al., 2014). The algorithm is a part of a more general system for monitoring changes of structures and buildings in quasi real time. It consists of three parts: the measuring system, the evaluation of the measurement results and distortion analysis segment, and an expert segment.

Today, many implemented in our country construction investments and quite commonly occurring emergencies of engineering structures and other objects, caused both by human activities and natural factors, claim a need for proper methods for measuring the status of buildings, as well as for evaluation of the measurement results in almost real time. This is of vital importance in crisis management, particularly in cases of disasters or other cases where there is a need to quickly assess the status of an engineering structure (flood, fire, etc.).

The correct identification of the threats to object security requires information about its current state and about its behavior under the impacts of the factors. The measurements providing the information can be run either periodically or continuously. Skłodowski (2009) notes that adequate monitoring, i.e. the continuous measurement, allows to review changes of the structure necessary for the risk identification. Monitoring, unlike individual measurements providing information only on the current state, allows to keep track of changes. A significant difference between periodic measurements and monitoring is also the amount of recorded data and the time needed for their collection, as well as the time required for their processing. The amount of information recorded during monitoring requires the use of automated systems for registering observations and their evaluation.

2. General description of the problem

In the process of monitoring of engineering structures, except of the essential surveying issues (such as methods, measurements, measuring equipment, the accuracy of measurements and of the networks), it is also essential the question of the frequency and the number of measurements. Within the being designed at the Institute of Geodesy of the Military University of Technology the system for monitoring engineering structures it is assumed running automated, frequent measurements of positions of the building located markers (Baszkiewicz, Kolanowski et al., 2014). The system is designed to support decision making based on real time analyses of changes of object's geometry and on the knowledge base in the course

of the emergency disaster. One of the components of the system is the measuring module to collect data from different sensors such as total station, digital camera, scanner, etc. Very important element is also a module to analyze any offset or deformation of the structure. They are to be determined through analyses of the measured displacement vector components of the marker. To research the required accuracy of measurements at the assumed detection sensitivity of changes it is designed and constructed a simulator called Monitoring Structural Objects Simulator, with the working name DRAGON (Baszkiewicz, Kolanowski et al., 2014).

The main objective of the research is to formulate a computer algorithm to be able, when processing measurement data, to identify trend of changes of point positions, as well as to detect any of its possible rapid changes that would indicate an abnormal behavior of the monitored structure.

3. Basic assumptions

The basic idea of the search algorithm is the requirement of automatic warning about the occurrence of values larger than expected. The algorithm is considered as a component the system consisting of a number of sensors, each measuring positions of several markers placed on the monitored structure, a software to control measurements, and a database to collect both the measurement data and the results of calculations. An interactive real time operator control (while performing a large number of measurements) is probably not effective enough to avoid oversights of significant data values. Therefore, the system is to provide automatic checking in such cases. Due to the nature of the method the marker position changes are not to be determined in common, but rather they are to come from separate sets of data processed individually by means of a single point method (Glass, Willson et al., 2008).

Within the algorithm there is also a problem of the "level of sensitivity" of the detected changes. It consists in the relationship between measurement and calculation errors, a change of the marker position and the reliability of the detected change as a real change. Therefore, the approach applied in that case uses some features of different methods of statistical analysis in processing data and measurements. The algorithm is based on the properties of the method of moving average and on the Douglas-Peucker algorithm (or Ramer-Douglas-Peucker algorithm), applied for cartographic generalization of linear objects (Douglas & Peucker, 1973). Its main task is to check whether the last observation made in the monitoring or the result of calculation is beyond the limit values, and, if so, whether it is a separate case or is the beginning of an observation subseries relevant to the new marker position. The range of the subseries determines the number of points to be used to calculate the moving average.

4. Basics of the research methodology

The algorithm was developed in a way strictly accounting for the specifics of the project. The algorithm itself, as well as the relevant application, are to be parts of the monitoring system. Therefore, at the first stage of the studies the methods of mathematical modeling and computer simulation were used. To assess the quality of the algorithm a mathematical model of the system described in (Baszkiewicz, Kolanowski et al., 2014) was constructed. The model was then used to create the following:

- A mathematical model of the building, where the geometric parameters can be modified according to the simulated changes;
- Computer programs to simulate:
  - photogrammetric stereo measurements;
  - total - station measurements of deformations;
- Application to carry out real time searching for outliers;

The computer programs are written in Borland C++ language, while computer application in VBA. These tools allowed for a wide range of simulated displacements and modelled object's deformations, as well as of the measurement accuracies. To simulate measurement errors the reverse cumulative distribution function of the normal distribution was used.

The second part of the study consisted in the analysis of the laboratory total-station measurements.

4.1. The moving average

The moving average method (*moving average* - MA) is a statistical method used in analyses of time series, mainly for economic and sociological applications. It is primarily used to smooth out data, which is the first step in the process of identifying the trend. An important feature of this method is its sensibility to outliers, which causes the desired trend line to become uneven or rough (Heckert & Filliben, 2003/2013). So, we can make use of this feature to find such data in the observation series, which differ more than the estimated or assumed observation noise from the data average. This needs, however, to check if the found outlier is a real displacement of the marker, as a beginning of the new observation subseries, or a random and isolated case. As a data average it is used the method of moving average according to (Box & Jenkins, 1976) and (Smith, 2015) with the method of linear regression. For a sequence of data there are determined parameters of the linear regression equation (Fig. 1). The analyzed parameter is a slope factor of the regression line. The relationship between the change of the regression line slope and the numeric values used to its determination is shown on Fig. 4. In addition to the method of moving average the algorithm uses also elements of the Douglas-Peucker algorithm. Its essential feature is the use of the Hausdorff metric to determine the distance between two geometric elements, forming collections of points in a metric space (Douglas & Peucker, 1976).
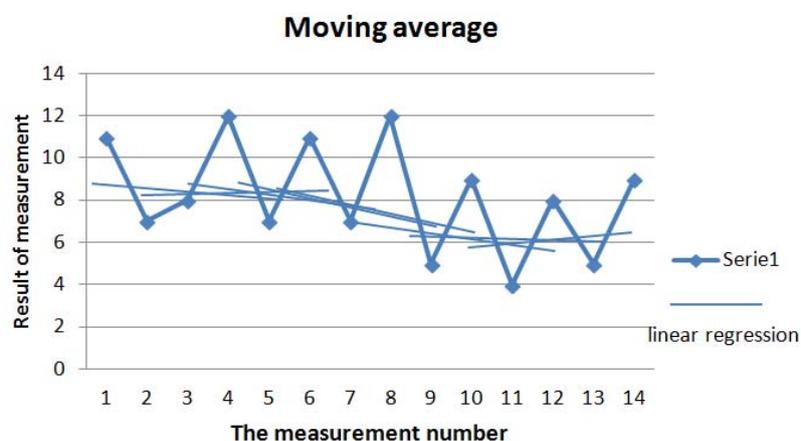

Fig. 1. Graph of simple linear regression moving average

4.2. Haudsorff's metric and Douglas-Peuker's algorithm

If we assume that A and B are non-empty compact subsets of R^n, then the Hausdorff distance between A and B is given by the formula (Herburt & Moszyńska, 2014):

$$dH\,(A,\,B):\, = \inf \{t > 0: A \subset Bt \text{ and } B \subset At\}. \tag{1}$$

It's easy to see that dH is in compliance with all properties of the metric. In addition, the border of a convex set (a convex symmetric set) in terms of the Hausdorff metric is a convex set (a convex symmetric set). Convex bodies limit must not to be a body, as it can have empty interior. If as a set A we consider a collection of points that make up the line from V_0 to V_1, and as a set B is a point outside the line V_0-V_1, then the Hausdorff distance is, according to its definition, the shortest distance between A and B, which is the distance perpendicular to the line of V_0-V_1 (Fig. 2.)
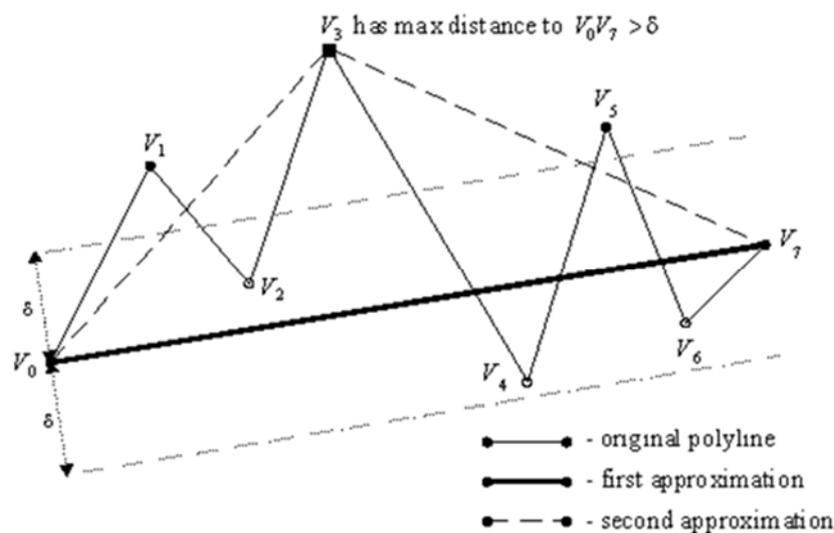

Fig. 2. The idea of the method of Douglas-Peukera (Sunday, 2012)

In the algorithm in (Douglas & Peucker, 1973) another assumption is adopted. It is specified the maximum Hausdorff distance of a set of points from a set of breakdown points of the line forming the iterative reference for their reduction.

In the first step (Fig. 3) as the end points of the line, actually the line segment being reduced, the starting and ending point of the polyline chain are assumed. The algorithm checks if within the assumed Hausdorff distance there is at least one of the vertexes, unless the polyline chain intersects the line segment being reduced. If this condition is not fulfilled the central point is taken as the end point of the new segment and the above condition is checked again.

These steps are repeated until both conditions are satisfied. As soon as this occurs, which corresponds to all points being located between the starting and ending point of the line segment being reduced are at the distance less than the assumed Hausdorff distance, they are removed from the set of vertexes of the polyline. The polyline is then approximated with the new reduced set of breakdown points.
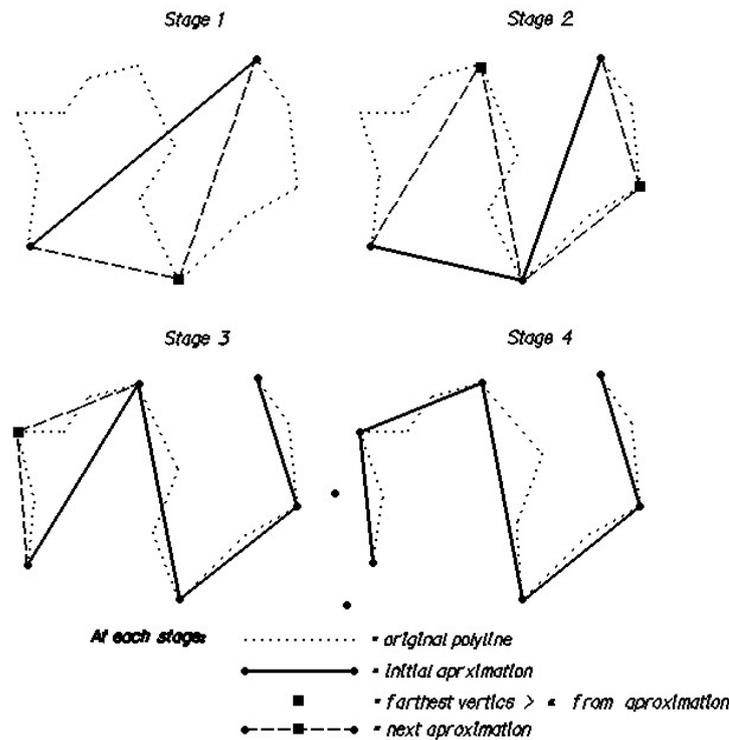
Fig. 3. The next steps of the Douglas-Peukera algorithm
(Sunday, 2012)

5. Algorithm of real time search of deviations

Both the differences in regression slope coefficient and in distance of the analyzed point from the straight line are used as search parameters for the outliers. An essential requirement for the algorithm is correct and reliable detection of the marker position change at the level of three times the mean measurement error. It is important that the algorithm should work differently than the generally used methods for processing time-series. In contrary to those of them, it is not to smooth the data, but rather to look for and to identify the outliers. Additionally, this has to be done in quasi real-time, but not in post processing.

The concept of the algorithm is shown on Fig. 4. Applying the moving average to calculate the expected "temporary" value makes it possible to analyze changes of marker position, which can be compared to the possible measurement error. This makes the algorithm to detect changes without a clear trend, but rather of incidental nature.

The Fig. 5 shows the flowchart and the pseudocode procedure of the algorithm. It comes with control parameters such as:

* *Dim_of_average* – the number of measurement results used to calculate the moving average
* *Hausdorf_distance* – the limit value for the distance between the sets of measurements results used to calculate the moving average and the simple linear regression
* *Diff_of_angel* – the limit value of a change of the simple linear regression gradient coefficient (expressed in degrees);
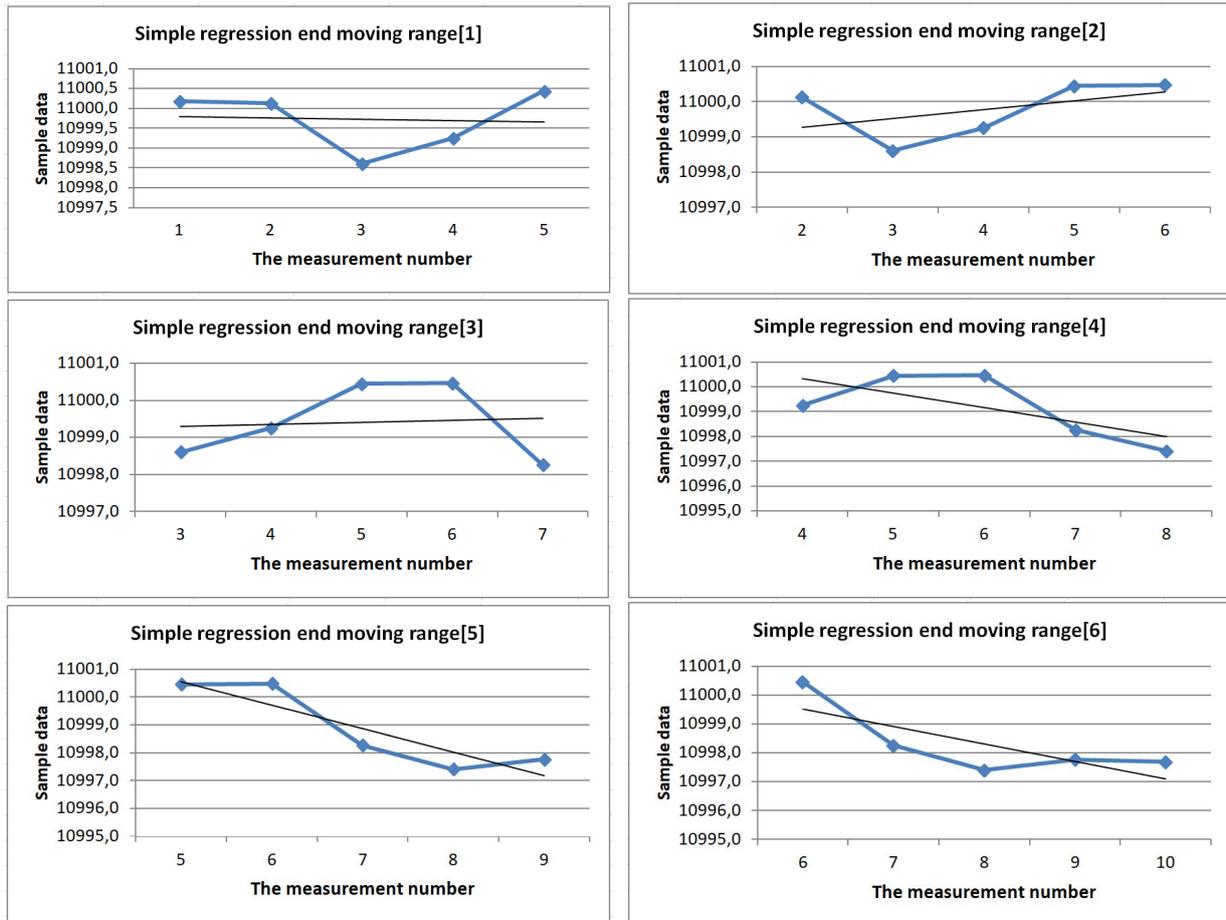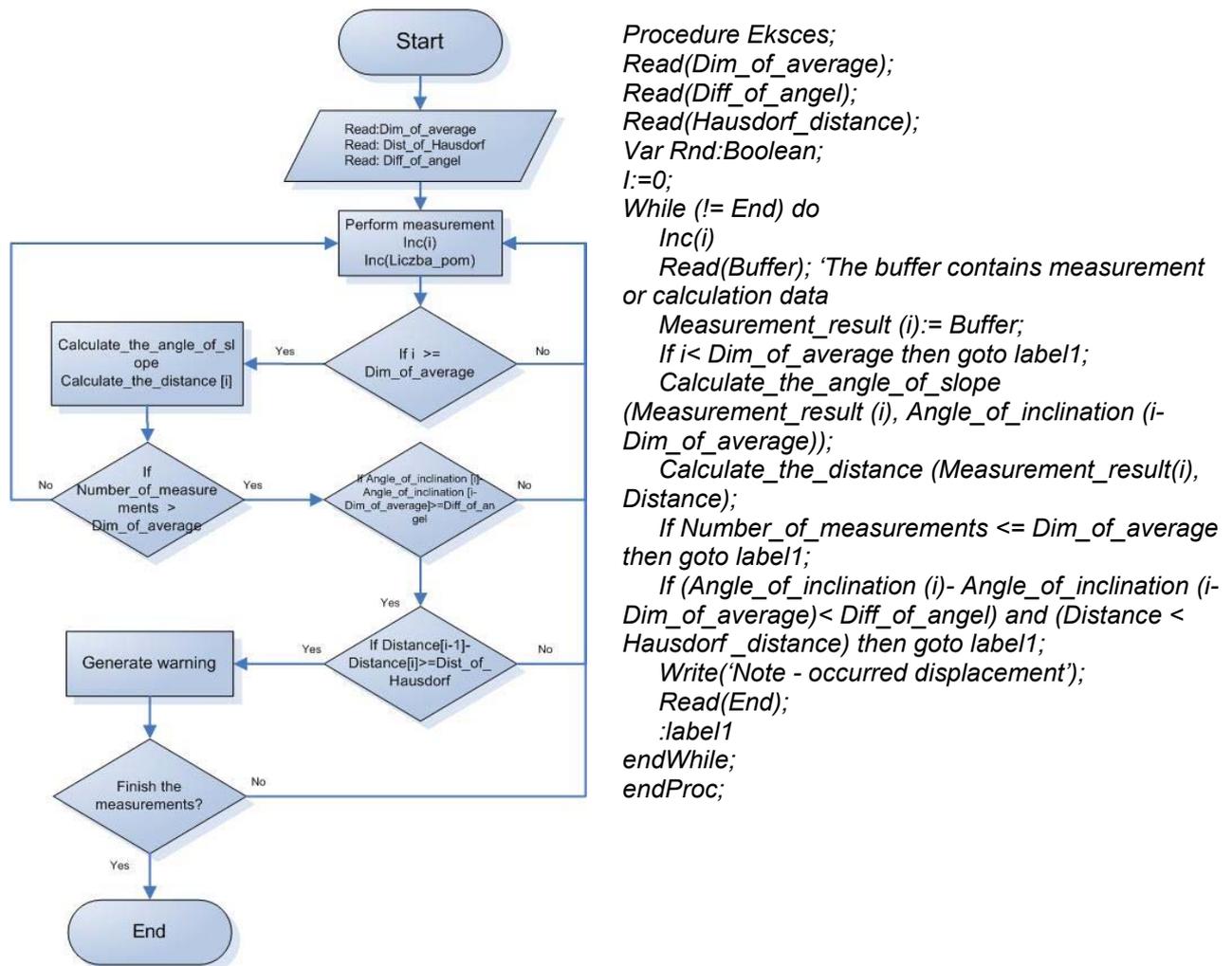
Fig.4. Changing the direction of the regression line based on the following arguments moving average dimension of 5

The flowchart contains the following nodes: Start; Read: Dim_of_average, Read: Dist_of_Hausdorf, Read: Diff_of_angel; Perform measurement Inc(i) Inc(Liczba_pom); If i >= Dim_of_average (Yes/No); Calculate_the_angle_of_slope Calculate_the_distance [i]; If Number_of_measurements > Dim_of_average (Yes/No); If Angle_of_inclination [i]- Angle_of_inclination [i-Dim_of_average]>=Diff_of_angel (Yes/No); If Distance[i-1]-Distance[i]>=Dist_of_Hausdorf (Yes/No); Generate warning; Finish the measurements? (Yes/No); End.

```
Procedure Eksces;
Read(Dim_of_average);
Read(Diff_of_angel);
Read(Hausdorf_distance);
Var Rnd:Boolean;
I:=0;
While (!= End) do
    Inc(i)
    Read(Buffer); 'The buffer contains measurement
or calculation data
    Measurement_result (i):= Buffer;
    If i< Dim_of_average then goto label1;
    Calculate_the_angle_of_slope
(Measurement_result (i), Angle_of_inclination (i-
Dim_of_average));
    Calculate_the_distance (Measurement_result(i),
Distance);
    If Number_of_measurements <= Dim_of_average
then goto label1;
    If (Angle_of_inclination (i)- Angle_of_inclination (i-
Dim_of_average)< Diff_of_angel) and (Distance <
Hausdorf _distance) then goto label1;
    Write('Note - occurred displacement');
    Read(End);
    :label1
endWhile;
endProc;
```

**Fig. 5.** The block diagram of the algorithm

6. Studies and tests of the algorithm

The functioning of the algorithm was tested on sets of data generated by the Simulator DRAGON (Baszkiewicz, Kolanowski et al., 2014). These data related measurements modelled on normalized images visualizations.

The Simulator generates measurement data like those made by a modelled set of digital cameras with uniquely determined common points (up to a negligible error). Generated errors contain the combined impact of instrumental errors and influences of the environment.

The modelled data of the building (the set of coordinates of the construction nodes (Baszkiewicz, Kolanowski et al., 2014)) contain information about the occurrence and the displacement values. The simulation results of 100 measurements at the specific location include information transformed by the Simulator. The goal of the application is to detect this fact and to indicate the measurement relevant to the displacement. The application is able to control the following input parameters:

- dimension of the moving average (the adopted standard is 5),
- the limit value of the Hausdorf's distance;
- the limit change of the slope angle of the regression line;

Changing of the regresion slope line

Changing of the Hausdorf's distance
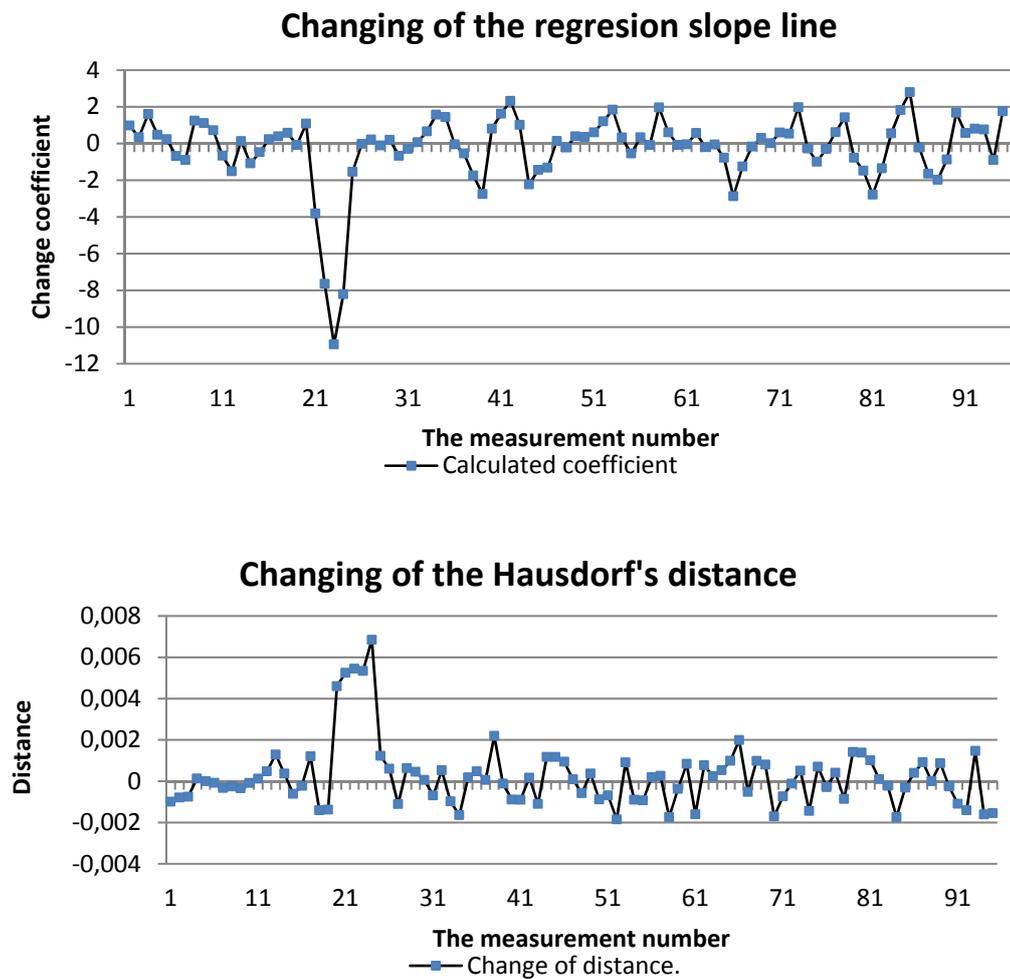
Displacement:	0,010 m
Mean error of a coordinate:	<=±0,004m
Size of the moving average:	5
Hausdorf's distance:	<=0,020m
Change of the regression line slope:	<=20°

Fig. 6. Results of simulation and analysis

While running the tests the effectiveness of the algorithm was checked (detection rate of the displacements) in relation to the control parameters.

Charts on Fig. 6 show functioning of the algorithm detecting displacements at 0.010m resolution at the generated position marker error ±0.004m. A negative value of the change of the slope angle of the regression line indicate negative change with respect to the previous value of the regression coefficient. In the presented case (as well as in many others) the algorithm pointed out a corresponding item on the list of results.

7. Initial laboratory tests

The functioning of the algorithm was initially verified through the measurements made in the Department laboratory. To perform displacement measurements it was

used the total station SOKKIA Net set on a concrete pillar of the lab. The target was indicated by a mirror magnetically attached to a mechanical device that made precise changes of the mirror position on the other pillar. The distance between the total station and the mirror was 23 m. The device and the electronic-mechanical sensor allow to change the mirror position to within ± 0.01 mm along each of the three Cartesian coordinate axes.

Under above conditions a number of measurements were made, from which a representative sample was treated by the described application. The results are illustrated in Fig. 7 and 8. Each setting of the mirror was observed four times, then the mirror position was changed by the desired interval by means of the screw micrometer. The result of the application functioning in the range of 0.12 mm of the mirror position change is shown on Fig. 7. Fig. 8 shows results of detection of position changes by 1, 2 and 3 mm. It should be noted that Any mirror position was changed along the X axis, that is perpendicularly to the instrument's line of sight while the application initial parameters were set the same as in the case of simulated measurements. Similar results were obtained when mirror position was changed along the axis Z, that is vertically. Changes along the Y axis, that is along the line of sight, were recorded using the total station. All the characteristics of the position changes and the way of their detection were the same as for the X axis.
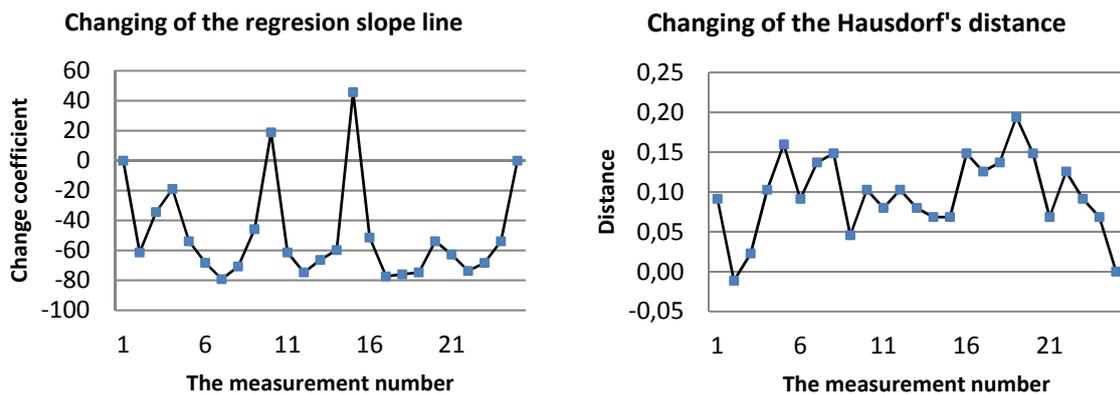

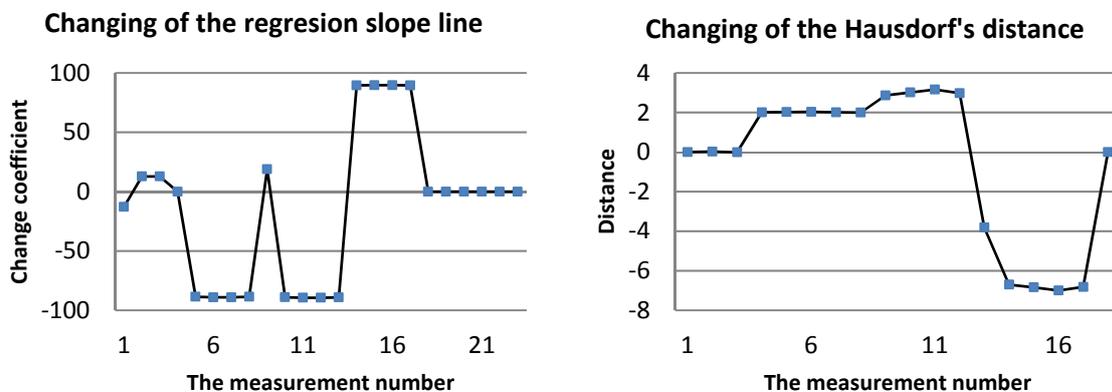**Fig. 7.** Moving the 0,12 mm at 4 measurement perpendicular to the target

Fig. 8. Offsets: I - 1,0mm; II - 2,0mm; III - 3,0mm
every 4 measurement perpendicular to the target

8. Summary

The algorithm is designed to evaluate results of the displacement measurements of engineering structures in quasi real time during monitoring in a threat situation. It was tested on data sets of stereo photogrammetry and total station measurements. By means of computer simulated measurements, as well as on the ground of laboratory measurements, the functioning of algorithm was tested against various cases of data sets. There were analyzed data from the following sensors: precision total station, simulated digital camera and simulated electronic total station. The results of the measurements and calculations are illustrated on Figs. 6, 7 and 8, which demonstrate the algorithm effectivity. The analysis was carried out in the post processing mode. The monitoring system will provide the real time detection of the displacements so that this will allow immediate automatic warning. They will also make a database element, which, together with the base of knowledge of the structures, will make the basis for an expert system to support decision making.

References

Baszkiewicz, K., B. Kolanowski, D. Latos, D. Krembuszewski, W. Pachelski & R. Sołoducha (2014). The selected methods of measuring deformations as a part of a system for monitoring structures, in *Selected papers the 9th International Conference on Environmental Engineering*, Vilnius, 2014.

Baszkiewicz, K., B. Kolanowski, D. Latos, W. Pachelski & R. Sołoducha (2014). Application of mathematical modeling and computer simulation to assess the technical parameters of sensors for measurement, *Theoretical foundations of civil engineering*, volume VI, Surveying measurement systems, Warsaw, Poland, 2014.

Baszkiewicz, K., B. Kolanowski, D. Latos, R. Sołoducha & W. Pachelski (2014). The concept of the system monitoring the State of the object and to assist the conduct of the rescue operation, *Logistics* 5/2014, 2014.

Box G. E. P. & G.M. Jenkins (1976). *Time series analysis: Forecasting and control*, Holden-Day, San Francisco, 1976.

Chatfield, C. (2013). *The Analysis of Time Series: An Introduction*, Sixth Edition, CRC Press, 2013.

Douglas, D. & T. Peucker (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *The Canadian Cartographer*, pp. 112-122, 1973.

Glass, G. V., V. I. Willson & J.M. Gottman (2008). *Design and Analysis of Time-series Experiments*, The Information Age Publishing, Inc., 2008.

Heckert, N. A. & J.J. Filliben (2003). *NIST/SEMATECH e-Handbook of Statistical Methods*, 2003/2013.

Herburt, I. & M. Moszyńska (2014). *Tool Geometry*, Center for the study of Advanced at the Warsaw University of Technology, 2014.

Liu, L. M. & G.B. Hudak (1992). *Forecasting and time series analysis using the SCA statistical system*, Scientific Computing Associates Corp., Chicago, Illinois 60607-3528, 1992-1994.

Machura, Ł. (2012). *Time-series Analysis*, University of Silesia, Katowice, 2012.

Ramer, U. (1972). An iterative procedure for the polygonal approximation of plane curves, *Computer Graphics and Image Processing*, p. 244-256, 1972.

Revel, M. (2012). *Statistics in Bayesian terms*, Katowice: UPGOW-Silesian University, 2012.

Skłodowski, M. (2009). Contemporary monitoring of construction, *Construction Review*, pp. 37-46, # 3 2009.

Smith, de M. J. (2015). *Statistical Analysis Handbook* http://www.statsref.com/HTML/index.html, 2015.

Sunday, D. (2012). Polyline decimation. Available online: http://geomalgorithms.com/a16-_decimate-1.html. (Date of access: 12 01 2016).

The National Institute of Standards and Technology, http://www.nist.gov/public_affairs/nandyou.cfm, 19.08.2009. (Online). Available: http://www.itl.nist.gov/div898/handbook/index.htm. (Date of access: 8.9.2015).

Authors:

Dorota Latos [1], dorota.latos@wat.edu.pl
Bogdan Kolanowski [1], bogdan.kolanowski@wat.edu.pl
Wojciech Pachelski [1], wojciech.pachelski@wat.edu.pl
Ryszard Sołoducha [1], ryszard.soloducha@wat.edu.pl
[1]Faculty of Civil Engineering and Geodesy,
Military University of Technology
Kaliskiego St. 2, 00-908, Warsaw, Poland