

DATA MINING AND NEURAL NETWORK TECHNIQUES IN STOCK MARKET PREDICTION: A METHODOLOGICAL REVIEW

Debashish Das and Mohammad Shorif Uddin

Department of Computer Science and Engineering, Jahangirnagar University, Dhaka,
Bangladesh

debashish.das@ulab.edu.bd
shorif.uddin@ulab.edu.bd

ABSTRACT

Prediction in any field is a complicated, challenging and daunting process. Employing traditional methods may not ensure the reliability of the prediction. In this paper, we are reviewing the possibility of applying two well-known techniques neural network and data mining in stock market prediction. As neural network is able to extract useful information from a huge data set and data mining is also able to predict future trends and behaviors. Therefore, a combination of both these techniques could make the prediction much reliable.

KEYWORDS

Data Mining, Neural Network, Stock Prediction, Stock Index.

1. INTRODUCTION

Stock market is basically nonlinear in nature. Prediction of stock market plays an important role in stock business. Data mining and neural network can be effectively used to uncover the nonlinearity of the stock market. Several computing techniques need to be combined in order to predict the nature of the stock market. As the time elapsed, Traditional capital market theory has been changed and various methods of financial analysis have been improved [2]. Researchers' attentions are attracted greatly for many years in forecasting stock return or a stock index. It involves an assumption of fundamental information that is publicly available in the past that has some projecting relationships to the future stock returns or indices. The samples of such information include economic variables such as interest rates and exchange rates, industry specific information such as growth rates of industrial production and consumer price, and company specific information such as income statements and dividend yields.

It cannot be assumed that the relationship between the stock returns and the financial and economic variables is perfectly linear. Nonlinear analysis of integrated financial markets needs to be considered instead of considering the modeling techniques that are based on linear assumption.

A large data collection is required for producing information. Only data retrieving is not enough, rather we need a means to automate the aggregation of data, information extraction, and recognize

discovery patterns in the source data. Files, databases and other repositories consists of huge amount of data, hence it is necessary to develop a prevailing tool for analysis and explanation of data and extracting interesting knowledge to facilitate in decision making. Data mining can solve all of the tasks of retrieving data such as mathematical figures and text documents, spatial data, multimedia data and hypertext documents etc.

Data mining is a method of extracting unknown projecting information from large databases which is a widespread technology that helps organizations to focus on the most important information in data repositories with great potential [5],[6],[7],[8]. Data analysis tools predict future trends and behavior, helping organizations in active business solutions to knowledge-driven decisions [6]. Intelligent data analysis tools produce a database to search for hidden patterns, finding projecting information that may be missed due to beyond experts' prediction.

Extraction which was previously unknown, implicit and potentially useful information from data in databases, is an effective way of data mining. It is commonly known as knowledge discovery in databases (KDD) [7], [9]. Although data mining and knowledge discovery in databases (or KDD) both are used as similar often, data mining is actually part of knowledge discovery [5], [7], [9].

On the other hand, neural network is another significant method for stock predictions due to their ability in dealing with fuzzy, uncertain and insufficient data which may fluctuate rapidly in very short period of time [1]. Plentiful research and application of neural network has proved its effectiveness over traditional methods that exclude artificial intelligence. According to Wong and Selvi Bodnovich [3], the most common field of applications of neural networking in production / operations (53.5%) and finance (25.4%) in the last few years. Neural networking is used in financial applications usually for stock and performance prediction. Many articles have addressed about neural networking applications in different predictions, but standard paradigms are not yet addressed that can determine the effectiveness of different methods of neural network in some areas of the problem [4].

This paper will review the main strengths and weaknesses of data mining and neural network in the stock predictions/any other predictions and highlight issues that may be vital for further research in this area. Some of the advantages and limitations will be addressed after performing the comparative analysis, methodology in previous studies with respect to problem domains, the data model and results criteria.

2. DATA MINING

Data mining tasks can be classified as follows (a flow diagram is shown in Fig. 1):

- **Analysis of survey data:** It is an interactive and visual method of easily view data without a clear idea of what we are looking for.
- **Explanatory simulation:** It includes a model for the whole probability information dissemination to express all the data; it divides large dimensional space into groups and models that describes the relationship between variables.
- **Analytical modelling:** This model allows to predicting the value of one variable on the identified values of other variables.

- **Identify patterns and rules:** It refers to the model of detection, recognition and the goal is to identifying fraud in the field of space determined by various types of agreements in which data points are significantly different from others.
- **Acquisition summary:** This task is often used for text and image data sets that find the scheme to extract the data that is parallel to the structure of interest in the data set.

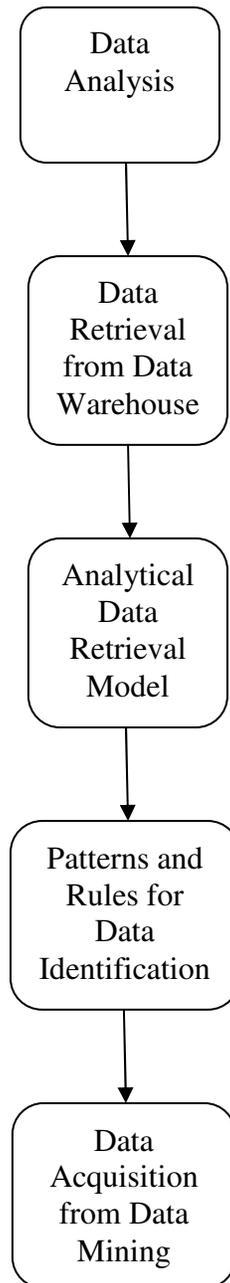


Figure 1. Data mining methodology.

Some of the data mining systems are as follows:

- **Statistical Data Mining:** Statistics provide a useful tool for data mining and they can be used to analyze or make inferences about data to discover useful patterns from a dataset. The database is integrated with statistical functions to draw statistical conclusions about the dataset in the database. A flow diagram is shown in Figure 2.

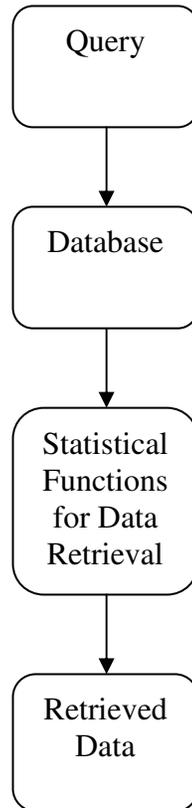


Figure 2. Statistical data mining process.

- **Neural network for Data Mining:** In traditional DBMS, Data is stored as a shape of structured records. When any query is submitted, database system searches for and retrieves records that match user's query criteria. Artificial neural network offers an excellent way for the recognition of intelligent query processing in large databases, especially for data retrieval and knowledge extraction based on partial matches. Neural network uses different methods. It does not need to identify empirical rules in order to make predictions. A neural network generates a network by examining a database and by identifying and mapping all significant patterns and relationships that exist among different attributes. The network then uses a particular pattern to predict an outcome. The neural network tries to identify an individual mix of attributes that reveals a particular pattern. This process is repeated using a lot of training data, consequently making changes to the weights of the data for more accurate pattern matches. The patterns that exist among the attributes in the database can be identified, and the influence of each attribute can be quantified. Neural network concentrate on identifying these patterns

without human guidance. Figure 3 presents a system for neural network-based data processing.

Three different types of datasets used in this process are as follows:

- ❖ **Training set:** It is used for training and for teaching the network to recognize patterns.
- ❖ **Validation set:** A set of examples is used to tune the parameters of a classifier by choosing the number of unknown nodes in a neural network.
- ❖ **Test set:** Neural Network performance is tested using this. It consists of a set of examples used only to assess the performance of a fully specified classifier.

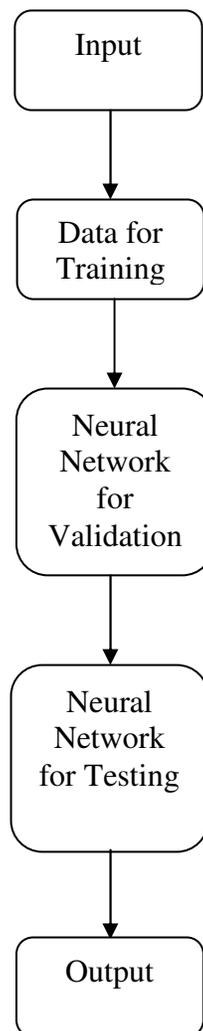


Figure 3. Neural network based data processing.

- **Clustering:** It can be used as a data-mining method to group together items in a database with similar characteristics. It is on how to group data items based on the similarities

among them. A cluster is a set of data items grouped together according to common properties and is considered an entity separate from other clusters. Hence, a database can be viewed as a set of multiple clusters for simplified processing of data analysis. It can be used to achieve objectives such as “identifying critical business values” or “discovering interesting patterns from the database.”

- **Fuzzy sets for data mining:** It is a general case of an arbitrary set. It is a set without a crisp boundary. While conventional sets have only two possible values, 0 and 1, fuzzy sets do not have this arbitrary boundary to separate members from nonmembers. It can be used to describe everyday business applications. In the real world, problems are often vague and imprecise, so they cannot be described in the conventional dual (true or false) logic ways. But, Fuzzy logic allows a continuous gradation of truth values ranging from false to true in the description process of application models.

3. NEURAL NETWORK

Neural network which is also referred as artificial or simulated neural network consisting of computer units (referred as artificial neurons) connected together such that each neuron can transmit and receive signals from each other (illustrated at Figure 4). Neural network can be used in signal processing, speech recognition, financial forecasting, monitoring of process control and monitoring samples for analysis.

The design of neural network is based on the human brain and they get knowledge through learning. Process of neural network learning is to regulate the severity of each of its components with respect to the input of neural network and its expected result. This process requires a set of input data, stock quotes. Figure 4 depicts the architecture of neural network model.

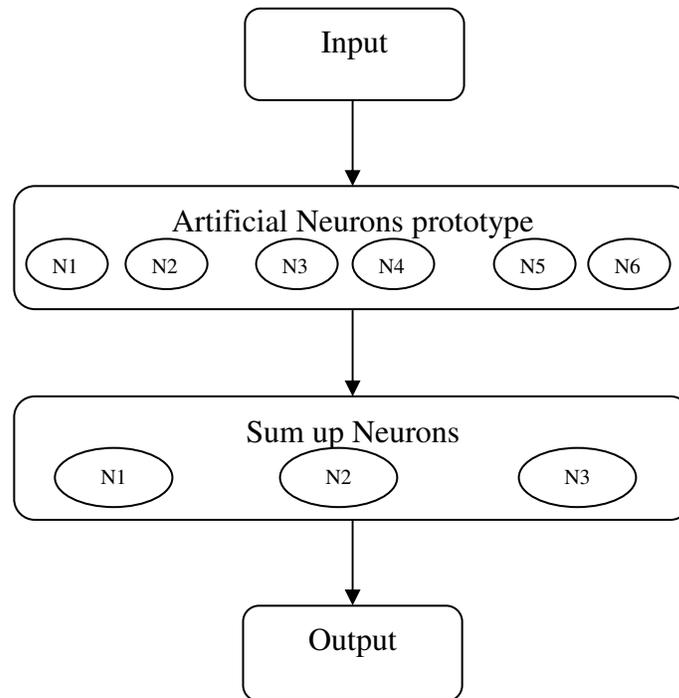


Figure 4. Neural network model architecture.

Neural network calculates and uses nonlinear decision boundaries that approach the optimal function of Bayes' theory. New training data can be used to modify the decision boundaries. Estimating the probability and readability of a classification as well as making decisions are also made [12]. Neural network model is a very useful tool for application into databases and signal processing. The following facilities can be obtained using neural network:

- **High computation:** It takes a lot of time to deal with huge databases and figure out which category a new pattern belongs to, applying this type of neural network can save time and effort, It can also improve the accuracy of the computation results.
- **Learning:** The neural network system can quickly learn the data from the learning data source. In real time using new data the decision boundaries can be modified once they become available and they can be implemented using artificial hardware "neurons" that operate entirely in a parallel processing manner.
- **Fault tolerance:** In a Neural network slight damage to the connections will not result in a serious problem, but it will decrease the functionality slightly. It is due to the fact that in a neural network the information is stored in distributed memory. If the input information is incomplete or has noise, the neural network can still perform proper processing.

4. METHODOLOGICAL REVIEW

Marijana Zekic [13] showed that the neural network (NN) accuracy mostly ranges from 70%-80%. NNs outperform statistical methods for a 5%-20% higher accuracy. The author claimed that if NN is combined with expert system, it will perform higher accuracy rate than only NN.

Bruce J. Vanstone, Gavin Finnie and Clarence Tan [14] demonstrated that artificial neural network (ANN) can be used to identify stocks with a potential to rise significantly on the basis of the stocks fundamental attributes. They have also found that the ANN configurations outperformed their non-neural equivalents. The majority of the trades generated at the signal threshold of 50 or more continued on to achieve their 100% target. The neural signal strength continued to increase also in their research.

Monica Adya and Fred Collopy [15] found that nineteen studies (86%) produced the results favorable to forecasting and prediction through neural network (NN). In those papers NN outperformed alternative approaches.

Mr. S. P. Deshpande, H.V.P.Mandal Amravati and Dr. V. M. Thakare [16] have discussed about the applications of data mining in various field relevant to decision making such as medical science for diagnosis, sports world for player selection, game strategy etc., prediction of personal bankruptcy, for improving product demand forecasting, assorted optimization, product recommendation, assortment comparison across retailers and manufacturers, forecasting production schedules for the manufacturing plants, determining market potential in critical go/no decisions on continuing work, or making financial projections for stock holders and investors.

Xianjun Ni [17] has explained the various neural network methods in data mining. Author has discussed about data mining process based on neural network (NN). He has focused on applying various data mining types based on neural network (NN). Finally, he has explained about key techniques and approaches of implementation regarding data mining based on NN.

Kunwar Singh Vaisla and Dr. Ashutosh Kumar Bhatt [18] proved that neural network (NN) outperform statistical technique in forecasting stock market prices. They have showed it through a method to forecast the daily stock price using neural network and then the result of the neural network forecast is compared with the Statistical forecasting result. They have proved that neural network, when trained with sufficient data, proper inputs and with proper architecture, can predict the stock market prices very well. On the other hand, statistical technique though well built but their forecasting ability is reduced as the series become complex. Therefore, NN can be used as a better alternative technique for forecasting the daily stock market prices.

Dase R.K. and Pawar D.D. [19] tried to sum up the application of Artificial Neural Network for predicting stock market. As per the authors view, in the World Predicting stock market index is a difficult task, but artificial neural network is having ability to predict stock index. They have also included that predicting stock index with traditional time series analysis has proven to be difficult and Artificial Neural network may be suitable for the task. A neural network (NN) has the ability to extract useful information from large set of data. They have presented a review of literature about application of artificial neural network (ANN) for stock market predictions and from this literature they have found that ANN is very useful for predicting world stock markets.

Muhammad A. Razi and Kuriakose Athappilly [20] have proved that NNs and CART models produce better prediction accuracy than non-linear regression model. They have also claimed that it is obvious from the study results that NNs and CART models provide better prediction compared to regression models when the predictor variables are binary or categorical and the dependent variable continuous. However they have also added that neither NNs nor CART model showed clear advantage of one over the other.

Chi-Jie Lu [21] has proved from the experimental results that the integrated independent component analysis (ICA)-based de-noising scheme with neural network proposed for stock price prediction model outperforms the integrated wavelet de-noising technique with BPN model, the BPN model with non-filtered forecasting variables, and a random walk model. According to the experiments, the author has concluded that the proposed method can effectively detect and remove the noise from stock prices/indices and improve the forecasting performance of BPN.

Pratyosh Rai and Kajal Rai [22] have found from the comparison that problem of stock index prediction is one of the most popular targets for various prediction methods in the area of finance and economics. In Their article the researchers have described the comparison of different neural network types for stock prediction. The prediction was carried out by modular neural network, ARIMA-based neural network, Genetic algorithm, Amnestic neural network, Multi-Branch neural network etc. The authors have also performed comparative analysis of all these types of neural network (NN).

5. ANALYSIS

Statistical methods [23] include various techniques such as the root mean square error, the mean absolute error and the mean squared prediction error, statistical indicators like the autocorrelation, the correlation coefficient, the mean absolute deviation, the squared correlation and the standard deviation. However, these methods give only average prediction in handling complex data.

Based on methodological review, non-statistical techniques include measures that are related with the data mining and various neural network techniques such as genetic algorithm, multi-branch

neural network may prove to be good performing methods for correct predictions if the model can be trained with sufficient data, proper inputs and architecture.

On the basis of the current survey analysis, we have found that NN method with expert system is the best performing method. Hence, the stock prediction system using a hybridization of knowledge-based data mining and neural network techniques would be a promising solution. A schematic flow diagram is shown in Figure 5.

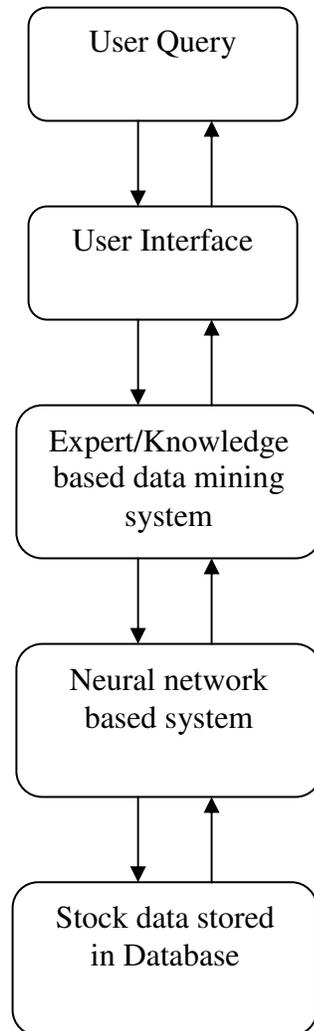


Figure 5. Data mining and neural network based stock predicting system architecture.

6. CONCLUSION

Most of the studies in predicting stocks use variety of techniques of neural network. Only few of them have used data mining. One or two studies have directed towards the combination of various techniques such as data mining, neural network or any other relevant techniques. Selection of data and methods for data mining and neural network is an essential job in this process and it requires

the knowledge of the domain. Numerous attempts have been made to design and develop the standard system for stock prediction but no system found completely generic yet.

From the methodological review, it is quite clear that data mining and neural network is very effective techniques to deal with unpredictable data like stock data or any other data that needs prediction. In future, it can be further investigated with real time stock data from any market.

REFERENCES

- [1] Schoeneburg, E.(1990), "Stock Price Prediction Using Neural Networks: A Project Report", *Neurocomputing*, vol. 2, pp. 17-27.
- [2] Poddig, T., & Rehkugler, H. (1996), "A world of integrated financial markets using artificial neural networks", *Neurocomputing*, 10, pp. 251–273.
- [3] Wong, Bodnovich & Selvi.(1997), "Neural Network application", *Neural Network business*, vol.19, pp. 301-320.
- [4] Li, E.Y.(1994), "Artificial Neural Networks and Their Business Applications", *Information & Management*, vol. 27, pp. 303-313.
- [5] *Introduction to Data Mining and Knowledge Discovery*.(1999), Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.).
- [6] Larose, D. T.(2005), "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2, John Wiley & Sons, Inc.
- [7] Dunham, M. H. & Sridhar S.(2006), "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition.
- [8] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R.(2000), "CRISP-DM 1.0 : Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringen Bank Group B.V (The Netherlands)".
- [9] Bernstein, A. & Provost, F.(2001), "An Intelligent Assistant for the Knowledge Discovery Process", Working Paper of the Center for Digital Economy Research, New York University and also presented at the IJCAI 2001 Workshop on Wrappers for Performance Enhancement in Knowledge Discovery in Databases.
- [10] G. P. Zhang, Ed. NN in business forecasting.
- [11] Sang C. Suh, Ed. Practical Applications of Data Mining.
- [12] Donald Specht.(1990), "Probabilistic Neural Network," *Neural Networks*, Vol. 3, No. 1, pp. 109-118, Elsevier Publishing.
- [13] Marijana Zekic. MS. "Neural Network Applications in Stock Market Predictions-A Methodology Analysis".
- [14] Bruce J. Vanstone, Gavin Finnie & Clarence Tan.(2004), "Applying Fundamental Analysis and Neural Networks in the Australian Stockmarket", Bond University ePublications@bond,.
- [15] Monica Adya & Fred Collopy.(1998), "How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation", *Journal of Forecasting*, J. Forecast, 17, 481 – 495.
- [16] S. P. Deshpande & V. M. Thakare.(2010), "DATA MINING SYSTEM AND APPLICATIONS: A REVIEW", *International Journal of Distributed and Parallel systems (IJDPS)*, Vol.1, No.1.
- [17] Xianjun Ni.(2008), "Research of Data Mining Based on Neural Networks", *World Academy of Science, Engineering and Technology*, 39.

- [18] Kunwar Singh Vaisla & Dr. Ashutosh Kumar Bhatt.(2010), “An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting”, (IJCSSE) International Journal on Computer Science and Engineering, Vol. 02, No. 06, pp. 2104-2109.
- [19] Dase R.K. & Pawar D.D.(2010), “Application of Artificial Neural Network for stock market predictions: A review of literature”, International Journal of Machine Intelligence, ISSN: 0975–2927, Volume 2, Issue 2, pp. 14-17.
- [20] Muhammad A. Razi & Kuriakose Athappilly.(2005), “A comparative predictive analysis of neural networks(NNs), nonlinear regression and classification and regression tree (CART) models”, Expert Systems with Applications, 29, pp. 65–74.
- [21] Chi-Jie Lu, “Integrating independent component analysis-based denoising scheme with neural network for stock price prediction,” Expert Systems with Applications 37(10), pp. 7056-7064.
- [22] Pratyosh Rai & Kajal Rai.(2011), “Comparison of Stock Prediction Using Different Neural Network Types”, International Journal of Advanced Engineering & Application.
- [23] George S. Atsalakis & Kimon P. Valavanis.(2009), “Surveying stock market forecasting techniques – Part II: Soft computing methods”, Expert Systems with Applications 36, pp. 5932–5941.

Authors

Mr. Debashish Das is currently attaining PhD in Computer Science and Engineering from Jahangirnagar University, Dhaka, Bangladesh. He received the Masters in Computer Science degree from University of Pune, India, in 2002. Currently, he is working as an Academic in Faculty of Computing, Engineering and Technology at Asia Pacific University of Technology and Innovation (APU), Malaysia. His interests include Data Mining, Information Systems, Programming, E-Commerce, E-Governance, E-Learning, Networking, Education etc.



Dr. Mohammad Shorif Uddin is currently a Professor in the Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh. He received PhD in Information Science from Kyoto Institute of Technology, Japan, Master in Technology Education from Shiga University, Japan and Bachelor Of Science in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET). He joined in the Department of Computer Science and Engineering, Jahangirnagar University, Dhaka in 1992. He started his teaching career in 1991 as a Lecturer of the Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology, (CUET). Dr. Shorif taught several advanced courses, such as Signal and Image Processing, Pattern Recognition, Computer Graphics, Artificial Intelligence, Neural Networks, and Computer Vision. He also lent his teaching and research experiences in many universities and institutes as a visiting faculty/researcher/guest speaker. He did postdoctoral researches at Toyota Technological Institute, Japan, Kyoto Institute of Technology, Japan and Bioinformatics Institute, Singapore. His research is focused on mainly three areas: scene analysis, development of vision-based navigational aids for blind people and bio imaging. His work is motivated by applications in the fields of computer vision, pattern recognition, blind navigation, medical diagnosis, bioinformatics and disaster prevention. Based on his research findings over 40 papers have been published in peer-reviewed international journals and conference proceedings including well-reputed IEEE Transactions on ITS, British IOP Journal, Japanese IEICE Transactions, Optics Express, Applied Optics (Optical society of America), Elsevier Science Journal. He holds two patents and received the best presenter award in the International Conference on Computer Vision and Graphics (ICCVG 2004), Warsaw, Poland. He is the author of two books and member of IEEE, SPIE, IEB and BCS.

