Fuzzy Clustering Algorithm Visualization Using an Enhanced Graphic Platform

Razvan Tanasie*

**Software Engineering Department, University of Craiova, A.I.Cuza, 13,Craiova, RO-200585, Romania (e-mail: rtanasie@software.ucv.ro)*

Abstract: An enhanced graphics render engine was developed using Microsoft C++ and DirectX API. This engine is part of a multi-purpose platform which, in this case, I used to simulate a two-level fuzzy clustering algorithm. The first level of the fuzzy system receives as inputs economical and financial statistical data provided by Eurostat and computes fuzzy membership values, using the Euclidian distance between clusters. The second level computes the optimal number of clusters. This is done in order to optimize the values of a set of statistic economical pre-selected factors. The final result is rendered using the graphic platform developed.

Keywords: Clustering, graphics engine, fuzzy systems, European Union, simulation.

1. INTRODUCTION

A realistic graphic multimedia platform was developed in order to be able to simulate both economical and other types of applications (Fig. 1) in a virtual environment.

The platform allows a user who is un-familiarized with the advanced graphic techniques to build, at any point, his own visual applications. He can set their constraints and visualize their graphic simulations in a very pleasant and realistic manner. Also, this platform permits a graphic programmer to properly test his algorithms, without being concerned with the environment development.

The platform brings an important innovative and interdisciplinary aspect – there exist simulators for different types of applications, there exist realist render engines, but there is not a platform that contain both. Also, apart from the scientific impact, there is an impact on the educational environment. The graphic simulations represent tools for educational process quality enhancement. A graphic simulator is much more appealing and easier to understand for the students, thus becoming a stimulant in the learning process. Some of the most important aspects of the platform are: portability, modularity, parallelism, and, very important, extensibility.

The graphics engine is built on top of interfaces that abstract the different graphics API's the engine can use. At the moment it is built around the DirectX API and allows to render both static and dynamic scenes (Adams, 2003).

The graphics engine handles the rendering API configuration and initialization, and exposes the interfaces used inside the engine for graphics related resource creation, modification and release.

The graphics engine also implements a lighting manager and frustum culling that work closer to the underlying API for

increased efficiency and uses a shader system.

It implements the pipeline necessary to draw engine objects on the screen using the materials, lights and effects they

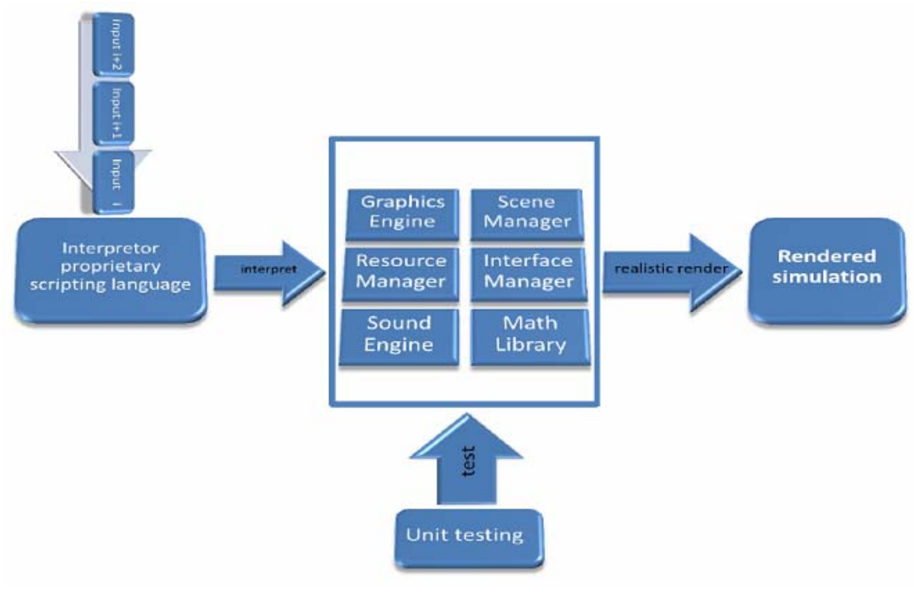

Fig. 1. Graphic simulation platform structure and functionality.

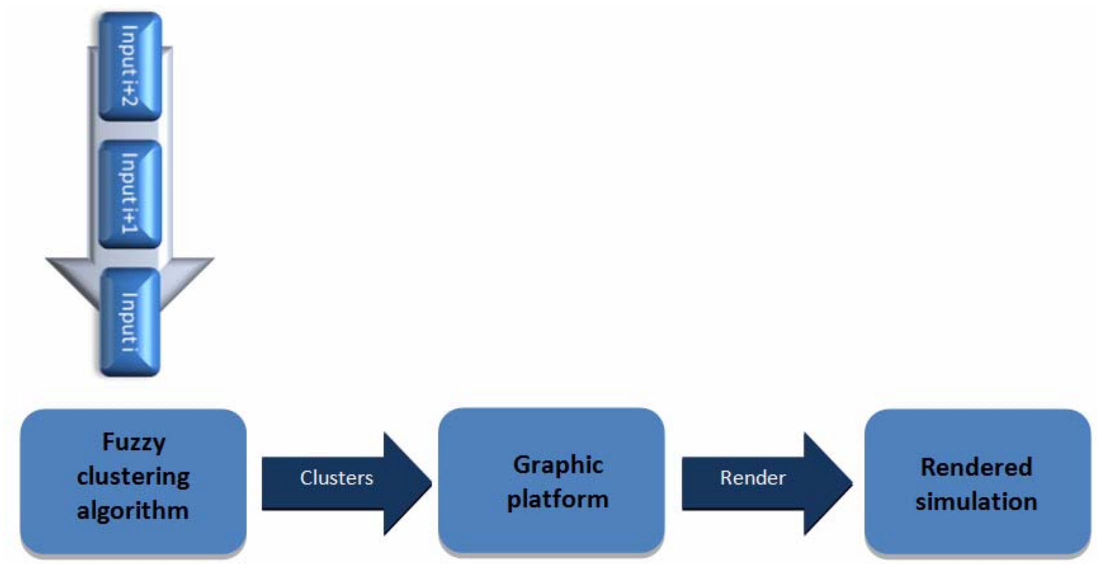

Fig. 2. Fuzzy clustering algorithm visualized using the graphic platform.

specify. It is also used by the interface manager module and the Resource Manager module for graphics resource creation.

This paper will present a set of fuzzy clustering algorithms and their results visualized with the help of the graphically platform previously presented. The economic data represent the input for the platform in Fig. 1, while the algorithms are the "interpreter" for the graphic platform (Fig. 2).

The new and potential members of the EU are all faced with the Euro adoption sooner or later. The enlargement of the Euro-Area will have a considerable impact on both economies and business. Lack of preparation in this respect will determine economic and financial shocks on both sides.

The issue this paper presents is the Euro adoption by the candidates to the European Monetary Union. The essential hypothesis is that the most recent members of the EU do not have the opt-out clause of the monetary union accession.

Thus, the Euro adoption is only a matter of time, and this study offers a method of isolating the sequentiality and the timing of this process. The main question from the basis of this study is "How prepared are the candidates?" and "Where do they gravitate relatively to the Euro-Area?".

2. FUZZY CLUSTERING ALGORITHM

2.1 General Overview

The fuzzy technique offers a new kind of approach, one closer to the economic reality, based on very rigorous mathematical structures that allow the positioning of every candidate in a cluster – a cluster is defined as a group of countries with similar status from the point of view of certain variables or criteria with regard to the studied problem. The economic reality proves that, no matter how the clusters are defined, each country is part of each of them with a certain membership degree. This suits perfectly the fuzzy systems theory (Kaufmann, 1986).

The method is applied in order to point out the similarities of the economic structures leaving from data concerning the candidate states to the euro adoption, and in order to accomplish an optimum grouping into homogenous clusters. This is accomplished using the main sets of data offered by the monetary convergence criteria.

The algorithm will use a two level fuzzy system – one for determining the membership of the countries and one for determining the optimal number of clusters (Fig. 3) (Wang, 1997) (Jungnickel, 2004).

Fig. 3 Two level structure for fuzzy clustering

Each country will be represented as a *c*-dimensional vector (where *c* is the number of criteria used to determine the clusters).

$$\overrightarrow{contry} = \left(criteria_1, \ldots criteria_c\right) \tag{1}$$

2.2 First Level of the Fuzzy Clustering Algorithm

The first level of the fuzzy system is based on the Kaufmann and Rousseuw clustering fuzzy algorithm (Kaufmann, 1990). Consider the data set as a set of n elements (states) and c variables (the monetary convergence criteria), defined as:

$$\{x_{ij}\}, \quad i = \overline{1 \ldots n}, \quad j = \overline{1 \ldots c} \tag{2}$$

where for each *i* constant, you get the values for the vector in Equation 1, defining a country.

Every variable is standardized with median zero and the standard deviation 1 in order to treat them as similar influences in the structure's determination.

The fuzzy algorithm for determining the membership of the countries has the following steps:

1. Set the number of clusters *k*
2. Sort the data set according to average criteria
3. Split the data in *k* partitions
4. Calculate center of partitions
5. Associate states to their closest center
6. Repeat from step 2 until a minimum imposed change is reached.

The dis-similarity coefficient between two countries used to determine whether they are part of the same cluster or not, is defined as the Euclidian distance between two vectors.

$$d\left(\overrightarrow{country}_i, \overrightarrow{country}_j\right) = \sqrt{\sum_{k=1}^{c}\left(country_{ik} - country_{jk}\right)^2} \quad (3)$$

2.3 The Second Level of the Fuzzy Clustering Algorithm

In order to compute the optimal number of clusters, it is required to add a second level in the fuzzy clustering structure that is implemented. For this to be accurate, a set of parameters is required to establish when this "optimal value" is reached. For this paper, two such statistic elements were selected:

- The silhouette;
- The Dunn partitioning coefficient.

The silhouette can be computed either for each cluster or for all the data. The silhouette for each cluster is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)$$

where *a(i)* is the average dissimilarity between all objects of the same cluster, and *b(i)* is the minimum of the average dissimilarities from all the other objects of a cluster, for all clusters.

Based on (4), when *s(i)* is close to 1, the object is correctly classified in the cluster where it can be found. A closer value to 0 indicates the ambiguity concerning the decision to place an object in a certain cluster. The negative values indicate the incorrect classification. The corresponding averages for each cluster, and for all the data considered indicate the accuracy degree of the data partitioning on clusters and overall.

The Dunn partitioning coefficient is computed in order to analyze the accuracy of the data partitioning. The normalized equation for the Dunn coefficient is:

$$F_k = \frac{\frac{k}{n} * \sum_{i=1}^{n}\sum_{v=1}^{k} u_{iv}^2 - 1}{k - 1} \quad (5)$$

where *k* represents the number of clusters and n the number of countries. This coefficient varies from 1 (showing that the data is correctly partitioned) to 0 (showing the exact opposite) (Yen, 2000). The coefficient is 1 only if, for every object, there is a coefficient that equals 1, while the rest are 0 and 0 when all the membership coefficients are *1/k*.

The algorithm for determining the optimal cluster number has the following steps:

1. Set the initial number of clusters *k* to *c*, where *c* represents the number of countries (maximal dissipation)
2. Apply the fuzzy clustering algorithm
3. Compute the statistic elements
4. If the new results are "better" than the previous best, retain the current number of clusters
5. If any clusters have been voided, go back to step 2.

An adaptation and simplification of the entire clustering algorithm would consist of replacing the second level with a random number of clusters, empirically predetermined.

2.4 Implementation

The graphics engine is built on top of interfaces that abstract the different graphics APIs the engine can use. At the moment it is built around the DirectX API (Luna, 2003), (Sanchez, 2000).

The graphics engine handles the rendering API configuration and initialization, and exposes the interfaces used inside the engine for graphics related resource creation, modification and release. It also implements a lighting manager and frustum culling that work closer to the underlying API for increased efficiency and uses a shader system.

The fuzzy clustering algorithm providing the input data is implemented in Microsoft Visual C++. Object oriented programming was used to implement the required data structures.

3. A FUZZY CLUSTERING APPLICATION FOR A MACROECONOMIC CASE

The fuzzy clustering algorithm and graphic platform described above represent a useful tool for macroeconomics, as clustering is a very modern and largely used approached (Gath, 1989). The main applicability domains close to macroeconomics are – monetary integration – with a focus on the development and evolution of the Euro-Area member states, regional development – the EU regional policy and funding, economic convergence of the EU member states, sustainable development etc.

In order to support this new approach and to provide exemplification for these fuzzy clustering tools, this paper aims to develop a past application accomplished by the author, few years ago, in the area of monetary integration (Tanasie, 2008). This is an evolutionary phase, a step forward, an improvement of the past approach based on both the development of the fuzzy clustering algorithm and the graphic platform tested here.

The economic field provides suitable testing data for both the fuzzy algorithm (Boreiko, 2002) and the graphic platform, due to the fact that previous clustering techniques employed in this field generated a comparable set of results.

Comparability and the rigorousity degree of the results obtained represent basis for the validation of the used methodology. This type of validation could extend the use of these instruments even further to engineering or other fields of application. As it was previously performed, this type of testing holds the necessary elements in order in ensure accurate results.

Monetary integration is a suitable testing environment as it uses two sets of criteria (Barr, 2003) – nominal convergence or the so called Maastricht and the real convergence criteria – mainly the GDP (Gross Domestic Product) per capita. These data are neither easy to bring to a common measurement system nor straight forward as far as the combined significance is concerned.

The main indicators for candidate and member states of the Euro Area – the Maastricht nominal convergence criteria include – the inflation rate (generalized increase in the consumer price index – yearly variation - percentage), the budget deficit (negative difference between national budget income and expenditure), the public debt, the interest rate (long term) and the foreign exchange rate variation (yearly percentage change of the national currency against Euro exchange rate).

Each of these indicators should range inside a predefined interval given by the EU Treaty of Maastricht: - inflation +/- 1,5% compared to the first two best performing EU member states, the budget deficit - +/-3% of national GDP, public debt +/-60% of national GDP, the interest rate +/-2% of the three best performing member states and the foreign exchange rate +/-15%.

The central discussion around these figures is that they have proved somewhat imprecise in depicting the real state of an economy, as the recent economic and financial crisis has pointed out.

The fuzzy clustering algorithm combined (Kaufman, 1990), (Mandani, 1974) with the graphic platform provide an instrument for determining the clusters of countries based on these criteria, otherwise not comparable as a combined significance.

It is very important here to present a relative positioning, or performance of member and candidate countries according to these criteria, rather than just determine the bare figures. This way, using this tool, analysts may asses the convergence evolution of a certain country as well as the countries relative positioning compared to the rest of Euro Area member states or candidates.

Even more, this would be a valid tool for the years to come and for all central and eastern European EU member states, including Romania, as they are all under the – *no-opt-out clause*. This treaty clause specifies the obligation to become part of the Euro Area for all existing EU member states still outside the monetary union. The only question remaining to be answered, in this respect, is the moment of the Euro adoption.

Also, from this point of view, of the Euro adoption timing and sequencing, this methodology could represent a useful tool. Based on determined clusters, it could ease the determination of the time interval to adoption and the countries sequencing inside this process, leaving from computed distances to all clusters' centers and the established numeric aims for the convergence criteria.

Adaptability and flexibility are two more features of these instruments useful for this type of applications. They are independent of the type of criteria chosen or given at a certain point in time. Data normalization and employment in the algorithm determine comparability and representation in a single gridline axis.

Test is done using data for the above presented criteria for continuous period of time before the economic and financial crisis provided by Eurostat. These are two main conditions and hypothesis meant to ensure appropriate results from an economic point of view.

As recent crisis years are not a reference period and analysis significance would not be appropriate – 2007 is chosen as reference. This choice is also based on the fact that this has been the moment of the most recent EU enlargement till the previewed Croatia membership from 2013.

Recent economic and financial crisis has raised even more questions and problems surrounding the Euro Area enlargement process, and the accuracy of convergence assessment.

4. RESULTS

Considering the economical input data required by the two-level fuzzy clustering algorithm presented above – an aiming optimal clustering determination (Garth, 1989) -, all the coefficients and membership functions are computed and visually presented using the graphic platform.

The simulations on the fuzzy clustering algorithm points out a relatively low level of the data dispersion, the Dunn coefficient for this case is around 0,55 and the silhouette is higher than before 0,56.

Given the statistical data mentioned above and provided by the main EU statistics body – Eurostat, for an economically stable envisaged period, results generated by the graphical platform are presented below. Due to recent economic evolution, and to the most recent crisis generated issue of the excessive deficit procedure and the high public debt acquired in order to sustain these deficits, we shall be using these two as main indicators for a three year stable period before the debut of the crisis in November 2007.

In order to represent the relative positioning of the countries inside the clusters, positions need to be determined based on initial criteria figured on the two axis. The reference position here should be the basis for comparison – with a fixed range – 60% of GDP for the debt and 3% of the GDP for the budget deficit.

Second positioning envisages also the budget deficit indicator – considered of major significance and the exchange rate volatility of the national currency against the Euro. It is important to specify that the represented countries were not part of the Euro Area for the analyzed time interval.

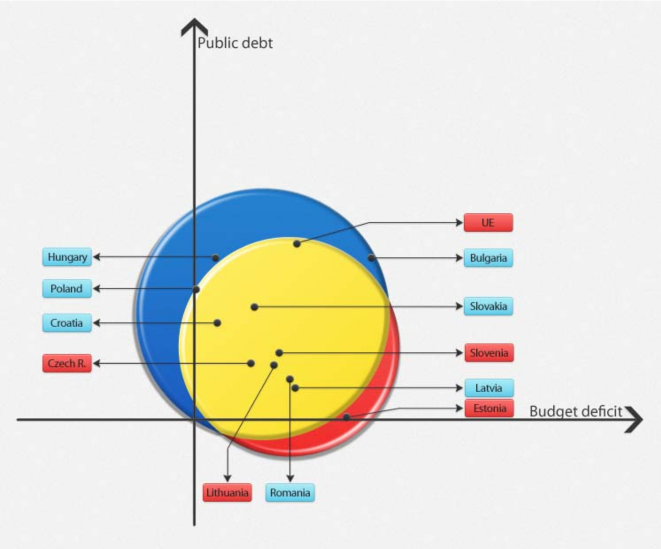

Fig. 4. Budget deficit–public debt clusters representation.

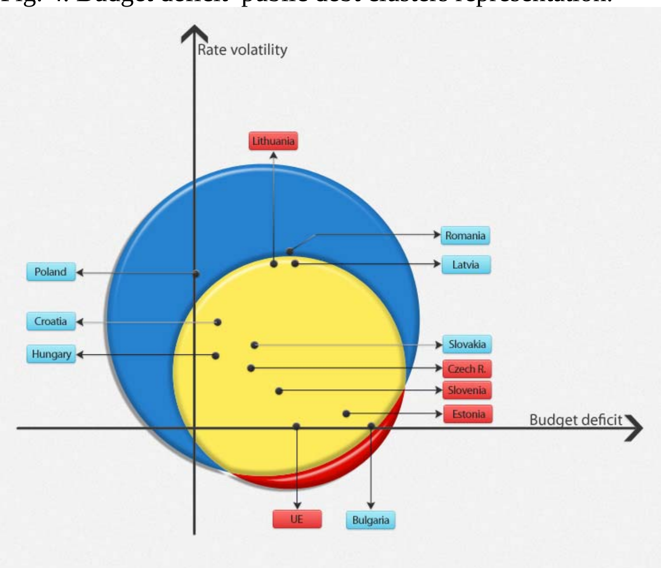

Fig. 5. Budget deficit-exchange rate volatility clusters representation.

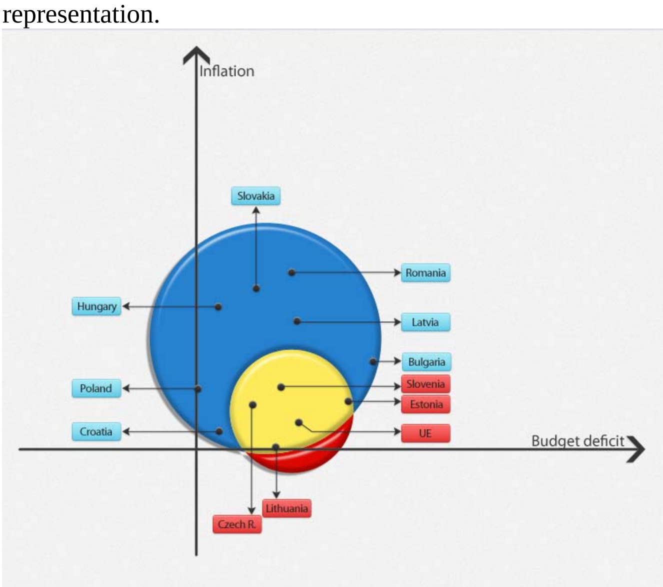

Fig. 6. Budget deficit-inflation clusters representation.

The EU an Bulgaria have a special position inside the cluster, as the Euro against Euro variation is null and the monetary board in Bulgaria has fixed the Bulgarian Leva against the DM, and then against the Euro. Polarisation in this case is rather high.

Third positioning also includes inflation as significant criterion during economic growth time intervals. As inflation averages and then normalized values are computed, dispersion tends to be higher than in the case of the exchange rate volatility. Even if inflation is targeted in most these countries, the exchange rate is easier to control and thus keep in a predefined variation interval. This explains most of the dispersion degree.

For the case of the partitioning based on the data for the real convergence indicator, simulations encountered the following results: - the optimum data partitioning is for two clusters – the best performance for the Czech Republic, Latvia, Lithuania - as best performing group. The second performing group is formed of Bulgaria and Romania with lowest levels of GDP per capita. The Polish case is closest to the boundary between the two clusters and has a middle ranking position.

The aim here is to locate, using the fuzzy clustering algorithm and the graphic platform, the more or less performing countries but also the optimal number of clusters. Distance to the reference points described above is also relevant in terms of the catching-up and convergence process. Together with the average growth rate of each indicator – nominal or real – the optimal Euro adoption moment may be determined for each candidate country.

6. CONCLUSIONS AND FURTHER DEVELOPMENT

A realistic graphic multimedia platform was developed in order to be able to simulate in a virtual environment both engineering systems and other applications such as economical applications. Also, a two-level fuzzy clustering algorithm was developed and tested on economical data to determine European countries clustering with regard to different statistical values provided by Eurostat (Krishnapuram, 1992).

The first level of the fuzzy system receives as inputs economical and financial statistical data provided by Eurostat and computes fuzzy membership values, using the Euclidian distance between clusters to determine the need for a new iteration. The second level computes the optimal number of clusters. The entire system is a loop, which ends when a set of preselected economical coefficients are within an accepted range.

The platform allows a user who is un-familiarized with the advanced graphic techniques to build, at any point, his own engineering applications. He can set their constraints and visualize their graphic simulations in a very pleasant and realistic manner. Also, this platform permits a graphic programmer to properly test his algorithms, without being concerned with the environment development.

The fuzzy clustering algorithm and the graphic platform simulation on macroeconomic data in the field of monetary integration confirm a two stage Euro Area enlargement, based on nominal and real convergence indicators analysis.

Further developments in this field would be the determination of the Euro adoption timing, sequencing and optimal moment, leaving from present achievements and computing distances to reference points.

Also, a further development would be a way to visualize the resulted clusters and the *k*-dimensional input vectors data simultaneously, in a pseudo-*k*-dimensional space.

ACKNOWLEDGMENT

This work was supported by the strategic grant POSDRU/89/1.5/S/61968, Project ID61968 (2009), co-financed by the European Social Fund within the Sectorial Operational Program Human Resources Development 2007-2013

REFERENCES

Adams, A., (2003), *Advanced animation with DirectX*, The Premier Press Game Development Publ.

Barr, D., Breedon F., Miles, D. (2003*), Life on the Outside: Conditions and Prospects Outside Euroland*, Economic Policy

Boreiko, D. (2002), *EMU and Accession Countries: Fuzzy Cluster Analysis of Membership*, Central Bank of Chile, Working Papers no.189

Gath, I., Geva, A.B. (1989), *Unsupervised Optimal Fuzzy Clustering*, TPAMI, 11(7):773-781.

Jungnickel, D., (2004), *Graphs, Networks and Algorithms*, Springer, 2nd edition.

Kaufmann, A. (1986), *Fuzzy subsets applications in O.R. and management*, in Fuzzy Set Theory and Applications, eds. A. Jones, A. Kaufmann and H.-J. Zimmermann (Reidel, Dordretch).

Kaufman, L., Rousseuw, P. J. (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: John Wiley & Sons, Inc

Krishnapuram, R., Nasraoui, O., Frigui, H. (1992), *The Fuzzy C Spherical Shells Algorithm: A New Approach*, TNN, 3(5):663-670.

Luna, F., (2003), *Introduction to 3D Game Programming with DirectX 9.0*, Wordware Publishing Inc.

Mandani, E.H. (1974), *Application of Fuzzy Algorithms for Control of Simple Dynamic Plant*, Proc. IEE, v. 121, no. 12.

Sanchez, J., Canton, M., (2000), *DirectX 3D Graphics Programming Bible*, IDG Books Wordwide.

Tanasie A., Tanasie R.T., Fratotiteanu C., (2008), *A fuzzy Clustering Algorith Application for Monetary Unions*, International Journal of Economic Persectives, Vol.2, Issue 4, ISSN 1307-1637, p.246

Wang, L. X., (1997), *A Course in Fuzzy Systems and Control*, Prentice Hall PTR.

Yen J., Langari R., (2000), *Fuzzy logic, intelligence, control and information*, Prentice Hall, New York.