



M Ű E G Y E T E M 1 7 8 2

Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Control Engineering and Information Technology

3D Shape Recognition Methods for Tangible User Interfaces

PHD THESIS

MÁRTON SZEMENYEI

Supervisor

Dr. Ferenc Vajda

September 9, 2020

Contents

1 Preliminaries	2
2 Aims and Methodology	3
2.1 Objectives	4
2.2 Methodology	4
3 Novel Scientific Results	6
3.1 Shape Classification	7
3.2 Feature Compression	9
3.3 Arrangement in Scenes	12
4 Summary of Results	15

1 Preliminaries

User interfaces (UI) are our primary way of interaction with computers and other machines, therefore their efficiency and usability are primary concerns in the field of Information and Communication Technology (ICT). In fact, one may observe that two of the three IT revolutions are linked to the development and spread of a novel interaction technology: The application of the mouse and Graphical User Interfaces (GUI) arguably played a major role in the spread of personal computers in the 1980s, while high-quality touchscreens were a major component in the popularity of smartphones, tablets and other similar devices.

At the time of this writing, the mouse and keyboard are the primary interaction devices used for personal computers, while smart handheld devices rely on touchscreens primarily. Arguably, one of the main reasons for the popularity of both mouse and touch-based interfaces is that both require the user to learn interface metaphors that come naturally to humans, since they rely on instinctive understanding of space-time and object manipulation. The result of relying on natural metaphors is an easy-to-learn and use interface, that - as long as it functions robustly - provides a powerful tool for users.

Notably, both of these interaction techniques are inherently two dimensional, which makes the manipulation of higher-dimensional objects or environments cumbersome. This is a significant shortcoming, since in the real world, users frequently have to work with three or four dimensional structures. This problem makes the research and development of novel, high-dimensional interfaces a worthwhile endeavor.

Tangible User Interfaces (TUI) have been an area of intense research in the last decade. The goal of these systems is to allow users to interact with virtual objects by manipulating real objects, which results in a natural interface. One particular subfield of TUI is Tangible Augmented Reality (TAR), which aims to combine Tangible User Interfaces with augmented reality by introducing real-world objects into the user interface design.

2 Aims and Methodology

Objects of various forms require different manipulation techniques, and provide different sensations when touched. The inconsistency of visual and haptic senses might cause neural conflict, which greatly diminishes user experience. Nonetheless, by matching real and virtual objects similar in shape, this conflict can be mitigated. Moreover, a shape-based matching process enables us to create Adaptive Mixed Reality systems (AMR): Environments that can adapt to the specific structure of the real scene they are projected into by arranging virtual objects intelligently, so that they would fit into the real scene.

The subject of this thesis is to propose a novel solution for determining a logical pairing of virtual and real objects. The main criterion of the pairing method is that the real and virtual objects must have similar physical properties, so that the users could effortlessly manipulate the real object as they would do with the virtual one. However, many important physical properties, like mass, roughness of surface cannot be measured visually, therefore the proposed algorithm uses shape and size only.

Notably, AMR systems may have other requirements: For instance the presence of certain virtual objects might be required, in which case even poor matches have to be accepted. Moreover, some environments may benefit from having multiple instances of certain virtual objects placed in the scene, while some others may not. Finally, there might be virtual objects that benefit from the proximity of another, in which case it is recommended to reward the algorithm for placing them close to each other.

This way AMR systems are able to reduce the setup and preparation work required by the user: The *developer* of the AMR environment is responsible for creating the virtual object categories, training the system to recognize suitable placeholder shapes, and setting properties and requirements for these virtual object categories. Using these, the system prompts the user to do a walk-around of the scene to perform 3D reconstruction, and then it proceeds to determine object placements automatically.

2.1 Objectives

Arguably, the object pairing method needs to be based on shape matching or shape similarity. While there are numerous shape matching methods, the aim of this work is to propose an algorithm that requires no prior information on the virtual objects, but can learn from instances of labeled real objects. Also, in order to make no assumptions on whether large-scale or small-scale shape attributes are important for pairing, the proposed methods attempt to encode all information on the shape of an object, and thus allow the learning algorithm to make that decision. Since the shape of a scene or certain objects is best represented as a structured object, the main focus of this thesis is to develop a learning algorithm for structured objects used to describe complex shapes.

However, shape recognition alone cannot solve this problem, since the additional requirements posed by the designer of the environment result in co-dependent assignments. To resolve this issue, some form of global optimization scheme has to be introduced that optimizes all assignments simultaneously, arriving at an optimal "compromise" solution.

2.2 Methodology

The methods proposed in this thesis are largely based on machine learning and heuristics, attempting to satisfy criteria that cannot be precisely defined, such as similarity of shape. Moreover, they are applied to a problem that is not guaranteed have a feasible solution (for instance if there are less real objects than required virtual ones). Consequently, the correct working or feasibility of the proposed algorithm for all possible inputs cannot be verified through traditional means.

For the above reason, we choose to demonstrate the performance of the proposed methods empirically by evaluating them a numerous different datasets. This, however presents another problem: since the problem stated earlier is not a particularly common one (the proposed application is novel to our best knowledge), there are no publicly available datasets to test our methods with.

To resolve this, we created our own datasets containing real object types commonly found on office desks, one containing 3D reconstructions from synthetic images created in Blender, while the other is using real images instead. Nonetheless, we argue that two datasets are not enough to evaluate our methods on, therefore we created numerous other datasets synthetically. These synthetic datasets are generated automatically, each dataset using a different, random set of hyperparameters. We argue

that having a higher number of structurally different dataset is a better choice for demonstrating the universality and robustness of our method, than having a single dataset with more training examples.

Once the methods have been evaluated on all datasets, we perform a statistical test to infer the difference between the performance of the proposed and the previous state-of-the-art methods. Since the methods are evaluated on the same datasets, the paired-samples t-test is used to evaluate the difference of the compared methods' efficiency. Having a higher number of different datasets also makes the results of these test more robust.

It is essential to point out, that the standard, frequentist versions of the statistical testing (such as the Student's t-test) infer the probability of the data under the null hypothesis. It is a grave (and all too common) error to infer the probability of the null (or any other) hypothesis based on these tests. Unfortunately, the probability of certain hypotheses is precisely what we aim to prove in this work.

For the above reasons, we employ the Bayesian version of the paired-samples t-test, which allows us to infer the probability of the effect size (the improvement caused by our proposed methods) being positive, as well as the 95% Credible Interval (CI). The cost of using such a test is having to explicitly choose a prior distribution of the effect size. In this work we use a t-distribution, since it is zero-centered, symmetric, meaning that the tests are unbiased.

3 Novel Scientific Results

The *first step* of the algorithm proposed in this work is segmentation and shape description. Object detection algorithms capable of detecting multiple instances usually employ a segmentation procedure, in order to produce object candidates for a subsequent classification method. This is a viable method of shape recognition, especially when applied to scenes where segmentation is relatively straightforward.

Nonetheless, segmentation is difficult in indoors scenes, since objects much more likely to be cluttered in this context. For this reason, a different approach is used: the input 3D scene is first segmented into primitive shapes, which may be interpreted as the “building blocks” of the scene. Then, primitive shapes are classified individually, and objects are determined based on the segment labels.

In order to avoid ignoring the geometric relations between objects, a graph is constructed from the primitive shapes, and a graph node embedding procedure is applied to produce a feature vector for each node that encodes the local context of the primitive as well. The initial classification is performed using these descriptors.

The *next step* of the algorithm is a discriminant analysis technique that aims to reduce the number of features used in the embedding to reduce the amount of computation and the complexity of the learning algorithm the method uses for classification. The aim of this discriminant analysis technique is not only to preserve the features that are useful for separating the nodes belonging to different classes, but also to keep the nodes *within a single instance* unique, since these might be useful for determining the pose of the real object.

The *final step* is to use the initial individual classification to find a globally optimal arrangement for the entire scene that is able to take all additional criteria into account. For this step a Genetic Algorithm (GA) is used, with novel, problem-specific genetic operators. This step also introduces a way to optimize the contextual requirements of virtual objects, allowing designers to set the preferred proximity of certain categories.

Our aim in the thesis is to show that the proposed solutions for these three steps are viable for a shape recognition and arrangement task, while they also provide improvement over previously existing methods.

3.1 Shape Classification

As mentioned above, the first part of the object pairing method is to perform simple shape recognition on the parts of the 3D scene independently to provide an initial pairing, which will be refined by later steps. In the current approach, this is formulated as a classification problem: The whole scene is broken down into parts, and these individual parts are classified according to which virtual object category they might belong to. The scores for the different classes will be used by a later global optimization step as initialization.

To increase the classification accuracy, a novel graph node embedding framework is proposed. These methods are able to handle both directed and undirected graphs of any structure, while at the same time they are applicable to graphs with vectorial edge and node features. The proposed algorithms also improve classification accuracy by including information about the local neighborhood of the individual nodes.

The first proposed algorithm is an explicit graph node embedding framework based on the spectral decomposition of graphs. Since the spectral embedding is only partially invariant to the node ordering, an initial ordering step ensures that the resulting feature vectors vary for different nodes. Then, a descriptor matrix is constructed for the node, called the node feature matrix \mathbf{F} , which contains information on the neighboring nodes as well. It is computed according to the equation below.

$$\mathbf{F} = \begin{bmatrix} T_{1,1} & T_{1,2} & \cdots & T_{1,N} \\ T_{2,1} & T_{2,2} & \cdots & T_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ T_{N,1} & T_{N,2} & \cdots & T_{N,N} \end{bmatrix} \quad (3.1)$$

$$T_{ij} = T(n_i, n_j, e_{1i}, e_{1j}, e_{ij}), \quad (3.2)$$

where T is a feature transform function, n_i is the i_{th} node of the graph, while e_{ij} is the edge pointing from the i_{th} to the j_{th} node. N is the maximum number of neighboring nodes considered in the embedding. The feature transform function may be chosen freely: In this work, we used distance-weighted kernel functions between the nodes and edges of the graph. Spectral decomposition can then be performed on the \mathbf{F} matrix, resulting in the feature descriptor of the given node.

The second proposed method is a variant of the random walk graph kernel. To compare graph nodes instead of entire graphs, the starting probability vector is modified so that walks would always start from the nodes being compared. To limit

the exploration to the local context, the maximal length of the walks is limited. Here, kernel functions between the nodes and edges are also used to compute the direct product of the graphs.

Both methods were tested extensively on shape graph datasets, and compared to classification accuracies achieved without any embedding. The Bayesian t-test performed on the results clearly shows that the proposed methods achieve a clear improvement in cross-validation accuracy, with the random walk node kernel *credibly* outperforming the explicit embedding. On the other hand, the explicit method is considerably faster, while *credibly* less prone overfitting. Also, the explicit method was compared with a neural network baseline in terms of generalization to scene graphs. It was shown that the explicit method was able to generalize robustly, while the neural network's accuracy was comparable to random guessing.

Furthermore, the proposed algorithmic frameworks are generic enough to be applicable to problems other than classification of shape graph nodes. In fact, with well chosen kernel functions the methods could be applied to 2D vision tasks, such as Bag of Visual Words-based classification or Deformable Parts-based detection methods. In these cases, an embedding method considering the presence of other nearby parts or visual words and the geometry of this neighborhood could improve the accuracy of these methods.

Lastly, the methods are not limited to graphs with vectorial edge and node features. As long as a sensible feature transform or kernel function can be defined between nodes and edges, both methods remain applicable.

Thesis 1

I created novel methods to improve the classification of vector-graph nodes, which make use of both the features of the given node, and its local context (other nearby nodes and the geometric relations between them). I showed via statistical testing that the embedding methods increase the classification accuracy on 3D shape-graphs. This contribution includes the following:

(a) I created a novel solution for embedding vector-graph nodes in a vector space, which also encodes the local context of the embedded node. I demonstrated via statistical testing that this method is able to robustly generalize to context, unlike previous state-of-the-art solutions.

(b) I extended the random walk graph kernel to function as a kernel on vector-graphs.

[1, 2, 3, 4, 5]

3.2 Feature Compression

The previous chapter introduced a novel method for embedding neighborhoods of graph nodes with vector features on both the nodes and edges. The suitability of this embedding for classification was also demonstrated. Still, the proposed method might result in a high-dimensional feature space, which is computationally expensive and prone to overfitting. For this reason, it is appropriate to apply a dimension reduction method to compress the representation.

In this section the first method for performing discriminant analysis for Structured Composite Classes (SCC) is presented. With SCCs, instances of objects are represented as unordered sets or graphs of feature vectors. The individual feature vectors of an instance are referred to as *nodes* of the object or class. The method assumes that the nodes are drawn from normal distributions, however, all of the nodes within a single object are drawn from different distributions.

Also, *all* the nodes in every object of every class are in the same vector space. Lastly, the proposed method makes no assumptions regarding the number of nodes per object, they may vary both between and within classes. However, there is no prior labeling available for these types of similarities, nodes are only labeled according to which class and instance they belong to.

An extra requirement of the SSC Discriminant Analysis problem is to separate nodes within a single instance, since this may be helpful for pose estimation at later steps. This can be addressed in a relatively straightforward way: by adding a second discriminant criterion that encourages selecting dimensions that separate nodes within instances. This addition is called the within instance scatter matrix, which is computed as follows:

$$\mathbf{S}_{wi} = \sum_{i=1}^C \sum_{j=1}^{N_i} \sum_{k=1}^{n_{i,j}} (\boldsymbol{\mu}_{i,j} - \mathbf{x}_{i,j,k})(\boldsymbol{\mu}_{i,j} - \mathbf{x}_{i,j,k})^T, \quad (3.3)$$

where C is the number of classes, N_i is the number of instances in the i -th class, $n_{i,j}$ is the number of nodes in the j -th instance, $\mathbf{x}_{i,j,k}$ is the k -th node of the j -th instance of the i -th class, and $\boldsymbol{\mu}_{i,j}^{inst}$ is the mean of nodes in the j -th instance of the i -th class. With this, the between class node scatter matrix can be defined similarly to classic LDA:

$$\mathbf{S}_{bcn} = \sum_{i=1}^C (\boldsymbol{\mu} - \boldsymbol{\mu}_i)(\boldsymbol{\mu} - \boldsymbol{\mu}_i)^T, \quad (3.4)$$

where $\boldsymbol{\mu}$ is the mean of all nodes and $\boldsymbol{\mu}_i$ is the mean of all nodes in the i_{th} class. Then, the optimization criterion of Structured Composite Discriminant Analysis (SCDA) can be written as:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_{bci} \mathbf{w}}{\mathbf{w}^T \mathbf{S}_i \mathbf{w}}, \quad (3.5)$$

$$\mathbf{S}_{bci} = \mathbf{S}_{bcn} + \mathbf{S}_{wi}. \quad (3.6)$$

Despite the addition of the within instance scatter, SCDA still suffers from the inaccuracy of the between class scatter matrix due to the invalid normal assumption. This problem was, however, solved by Subclass Discriminant Analysis (SDA) by assuming a Mixture of Gaussians model and computing the between subclass scatter accordingly. However, this method relies on expensive (and in the case of SCCs, inaccurate) clustering.

Thankfully, it is possible to address this issue by improving the clustering procedure (SDA-IC). SDA-IC estimates of the number of subclasses in each class separately by setting it to the number of nodes in the largest instance of the given class. The subclass means are initialized with the nodes of the largest instance. Then, the remaining nodes can be assigned to the different subclasses using a nearest mean approach. This is equivalent with using a single iteration of the k-means clustering algorithm. During the assignment step it is possible to penalize the algorithm for putting two nodes of the same object into the same subclass cluster.

Since the clusters are initialized by considering the requirement of separating the nodes of a single instance, the resulting clusters are much more likely to be close to optimal. The improved clustering method makes use of the additional information in the dataset to achieve significantly more accurate subclass clusters at reduced computational cost.

A minor issue with the SDA-IC algorithm is that it compresses the two separability criteria into the same scatter matrix. This takes away the ability to weight the relative importance of the two matrices which may be important. An additional reason for doing this is that certain applications demand significantly less tolerance for one type of error than the other. This may justify adding artificial bias to the discriminant analysis algorithm. The solution to this problem may be achieved by combining the SDA-IC and the SCDA methods by defining the \mathbf{S}_{bci} matrix of Subclass Structured Composite Discriminant Analysis (SSCDA) by:

$$\mathbf{S}_{bci} = \mathbf{S}_{bsb} + \mathbf{S}_{wi}, \quad (3.7)$$

where \mathbf{S}_{bsb} is the between subclass scatter of SDA, and the actual subclasses are determined by the improved clustering method presented earlier.

The best-performing method proved to be the Subclass Structured Composite Discriminant Analysis (SSCDA), which clearly outperformed all other methods on versatile datasets, with the exception of SCDA, where the improvement is less certain (though still likely positive). The SCDA and SSCDA methods are also the highest performing on the 3D shape-graph databases, meaning these methods are viable choices for the shape recognition system.

Notably, the proposed structured composite discriminant analysis methods are not limited in their application. As mentioned before, structured composite classes occur in numerous perception problems, such as object classification and detection. The SSCDA method is arguably viable as an image feature descriptor, although - unlike SURF - it is not necessarily invariant to standard image transformations, such as rotation or scaling.

In general, the SSCDA method can be used to perform discriminant analysis for component-based data structures, as long as the individual components can be described using the same set of variables. Furthermore, all proposed discriminant analysis methods assumed that the classes have a Mixture of Gaussians distribution, which is an important limitation of these algorithms. If this assumption is not true, the results are likely to be suboptimal.

Thesis 2

I created a novel method to perform dimension reduction for classes composed of sets or graphs of vectors (structured composite classes). I showed on multiple datasets that the proposed method provides a viable solution not only for separating different classes, but also for simultaneously separating individual components within a single instance. I demonstrated via statistical testing that the new method provides descriptors that allow for higher classification accuracy compared to previous methods for structured composite classes. [6, 7, 8]

3.3 Arrangement in Scenes

In this chapter a methodology for performing globally optimal scene arrangement based on individual part-by-part classification results was introduced. This is needed for several reasons: First, the optimal labels for the individual parts are interdependent: since most real-world objects are spatially constrained, it can be argued that nearby parts are more likely to belong to the same object, and therefore have the same label. Consequently, the choosing the final label of a given part of the scene influences the optimal label for other nearby parts.

Moreover, the application might pose other constraints and requirements that the final pairing has to satisfy. In the Tangible Mixed Reality setting the presence of some objects might be required, or the total number of objects placed in the scene might be limited.

The method is based on constructing an objective function that - while attempting to stay close to the individual classification - also considers other constraints and requirements, such as the compactness of the objects. Moreover, the cost function needs to be rather flexible, allowing the designer to introduce further constraints, such as requiring certain categories to be present, or encouraging the placement of multiple objects of the same class. Lastly, the function needs to present the opportunity to influence the spatial arrangement of category pairs by introducing context reward into the objective function. To summarize, the cost function needs to encompass the following elements:

- The labeling has to be valid: every node should have exactly one label.
- The final labeling has to have acceptable classification scores.
- Nearby object parts should have the same label.
- All required classes must have at least one object part labeled as them.
- The algorithm should be encouraged to place multiple instances of objects that have been selected for multiple placements by the designer.
- The algorithm should be encouraged to place 'compatible' objects nearby and 'incompatible' ones afar.

The constructed objective function is made of four distinct parts: The first is the classification scores provided by the node-only classification, while the second is a compactness term, encouraging nearby objects to have the same label. The third

term rewards the correct context (closeness of certain object categories), while the fourth rewards the placements of certain categories. Required objects are introduced as constraints into the optimization problem.

Following the construction of the cost function, several optimization strategies were evaluated and compared on this problem, including simulated annealing and genetic algorithms. To increase the performance of the latter method, three novel operators were proposed for initialization, mutation and crossover.

The proposed Class Score Optimal Initialization (CSOI) method begins by adding a single individual to the population: the individual that maximizes the classification scores. Then, the mutation operator is used to generate the rest of the initial population. To ensure the diversity of the initial population, identical individuals are checked for and replaced with newly mutated objects. Moreover, to further increase the diversity, a certain percentage of the individual population is generated by applying multiple mutations in succession.

The second proposed operator is the Random Drag and Shuffle Mutation (RDSM). The standard mutation operator for binary/nominal integer genomes is the random flip operator, which randomly changes the label of a single node. In this case, however, the random flip operator is very likely to create a significantly worse candidate solution, because of the compactness element in the cost function. To solve this, we introduce the concept of random drag: the node whose label is changed may drag other nearby nodes with it with a certain probability, meaning these other nodes will be assigned the same new label. The drag probability influences the trade-off between node mutation and cluster mutation. Note, that the former is still vital to allow the genetic algorithm to correct mistakes in the classification or to separate close objects. By allowing both kinds of mutation to occur, the parameter space can be explored more efficiently.

The last proposed operator is the Clustered N-Point Crossover (CNPC). The standard crossover operator for binary/nominal integer genomes is the N-point intersection. This operator randomly places N points in the genome, effectively dividing it into $N + 1$ parts. Then, the offspring inherits these sections from the two parents in an alternating way. The problem with this operator is similar to the case of the random flip mutation, namely, that defining the crossover on the level of nodes may lead to the creation of a high number of inferior offspring.

This problem can be solved by applying an idea similar to the random drag: the intersection operator should be defined on node clusters instead of the nodes themselves. This means that all nodes are assigned to a cluster based on their proximity,

using an adaptive distance threshold to divide clusters. Then, the clusters are ordered randomly, and divided into N intervals. The labels are then inherited from the parents alternatingly.

After evaluating the methods on the datasets, the results clearly show the global optimization scheme outperforming the individual classification scheme. Moreover, the genetic algorithm with custom operators was shown to be far superior to other methods, with each proposed operators contributing credibly positively to the final accuracy. Moreover, it was shown that the embedding method credibly increases the accuracy of the final arrangement. Finally, comparison against a neural baseline demonstrated that the global optimization can outperform graph neural networks trained on the scene dataset.

Exploiting spatial compactness of physical objects was a key element of both the objective function and the genetic operators, which means that these methods are likely to result in inferior performance if this assumption does not hold. While on paper this is undeniably a limitation, in reality, however, the spatial compactness of logical entities is a relatively universal principle. This is not to say that exceptions do not exist, rather that in our view they are rare.

Thesis 3

I developed a comprehensive, efficient search method for globally optimal arrangement of classes in scenes composed of objects by exploiting geometric properties of the problem. I formulated of the problem so that the optimization method could consider design concerns relevant to Tangible Mixed Reality Systems. I verified via statistical testing that the new algorithm surpasses previous graph-based object detection methods on numerous scene graph datasets.

(a) I developed novel genetic operators that provide a statistically significant increase in the probability of finding the global optimum in optimization problems where compactness of object can be assumed.

[1, 2, 5, 9]

4 Summary of Results

In this thesis our stated goal was to propose a scheme composed of novel methods that can efficiently search for a globally optimal arrangement of virtual objects in a complex 3D scene. The algorithm's object matching is based on 3D shape similarity, while it is augmented by several further optional criteria that may be of use to the designer of the Adaptive Mixed Reality environment. These criteria were designed to aid the automatic placement of virtual objects so that the final arrangement would be valid and intuitive.

The thesis introduced three novel solutions for embedding the local context of graph nodes into a vector space, performing discriminant analysis on structured composite classes, and finding globally optimal arrangement of labels using part-based classification results as initialization. These algorithms were shown to be efficient for the problem described above, furthermore, they were shown not to be limited to a single application.

We have evaluated the results of the proposed algorithmic set on numerous different datasets, and our results show low error rates achieved at reasonable speed. Since the initial setup of the scene is performed only once at the beginning, a few seconds of processing time is well within the acceptable range. Our experiments also showed that the proposed methods provided an increase in accuracy compared to previous or standard solutions.

In conclusion, our methods provide a feasible scheme for automatically pairing virtual objects to real placeholders in an Adaptive Mixed Reality environment. It is worth noting, that while the error rate of our method is relatively low (0.62% on the real image-based dataset), it still provides erroneous assignments at times. These, however only present a mild inconvenience to the user, as the occasional error could be easily corrected manually. Without automatic pairing, manual assignment would have to be performed for all virtual objects. For reproducibility, all code and datasets are available online at <https://www.github.com/szemenyeim/PhD-Research>.

Thesis Publications

- [1] M. Szemenyei, “Neural graph node classification via self-attention,” in *Workshop on the Advances of Information Technology*, B. Kiss and L. Szirmay-Kalos, Eds., BME-IIT, 2020, pp. 33–38.
- [2] M. Szemenyei and F. Vajda, “3d object detection and scene optimization for tangible augmented reality,” *Periodica Polytechnica Electrical Engineering and Computer Science*, vol. 62, no. 2, pp. 25–37, 2018. DOI: 10.3311/ppee.10482.
- [3] M. Szemenyei and F. Vajda, “Learning 3d object recognition using graphs based on primitive shapes,” in *Workshop on the Advances of Information Technology*, B. Kiss and L. Szirmay-Kalos, Eds., BME-IIT, 2015, pp. 67–71.
- [4] M. Szemenyei and F. Vajda, “3d object detection using vectorial graph node embedding,” in *Workshop on the Advances of Information Technology*, B. Kiss and L. Szirmay-Kalos, Eds., BME-IIT, 2017, pp. 45–53.
- [5] M. Szemenyei and P. Reizinger, “Attention-based curiosity in multi-agent reinforcement learning environments,” in *2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)*, IEEE, May 2019. DOI: 10.1109/iccairo47923.2019.00035.
- [6] M. Szemenyei and F. Vajda, “Dimension reduction for structured composite classes in multi-object environments,” in *Proceedings of the 15th International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED)*, V. Mladenov, Ed., ser. Recent Advances in Electrical Engineering, WSEAS, vol. 58, WSEAS Press, 2016, pp. 134–141.
- [7] M. Szemenyei and F. Vajda, “Dimension reduction for objects composed of vector sets,” *International Journal of Applied Mathematics and Computer Science*, vol. 27, no. 1, pp. 169–180, 2017. DOI: 10.1515/amcs-2017-0012.
- [8] M. Szemenyei and F. Vajda, “Optimal feature selection for objects composed of vector sets,” in *Hungarian Conference on Computer Graphics and Geometry*, L. Szirmay-Kalos and G. Renner, Eds., NJSZT, 2016, pp. 7–14.
- [9] M. Szemenyei, “Evolutionary scene arrangement for adaptive augmented reality systems,” in *Workshop on the Advances of Information Technology*, B. Kiss and L. Szirmay-Kalos, Eds., BME-IIT, 2018, pp. 10–17.