Evolving visual artefacts based on consumption patterns

Catarina Maçãs*, Nuno Lourenço and Penousal Machado

CISUC – Department of Informatics Engineering,
University of Coimbra,
Coimbra, Portugal
Email: cmacas@dei.uc.pt
Email: naml@dei.uc.pt
Email: machado@dei.uc.pt
*Corresponding author

Abstract: In information visualisation, the visual artefacts should have a functional dimension, allowing the analysis of information, and an aesthetic dimension, to seize the users' attention to the information being displayed. However, in the data aesthetics field, the main concern is to produce aesthetically appealing artefacts. With this project, our goal is to try to join these two fields by exploring the aesthetic dimension of a functional visualisation model characterised by a series of parameters that can make the visualisation more functional and/or more aesthetically appealing. In concrete, we propose a framework based on interactive evolutionary computation (IEC) to evolve the parameters of the visualisation model, enabling the user to explore possibilities and to create different aesthetics over the same data. Our case study will be based on a dataset containing the consumption patterns within a Portuguese retail company. Through different validation methods – automatic fitness function, usage scenarios, and a user study – we show that our system is able to create a wide and diverse set of emergent visual artefacts that can be intriguing and aesthetically appealing for the user.

Keywords: information visualisation; data aesthetics; visualisation model; visual artefacts; swarm systems; emergent artefacts; aesthetically appealing; consumption patterns; genetic algorithms; interactive evolutionary computation; automatic fitness function; user study.

Reference to this paper should be made as follows: Maçãs, C., Lourenço, N. and Machado, P. (2020) 'Evolving visual artefacts based on consumption patterns', *Int. J. Arts and Technology*, Vol. 12, No. 1, pp.60–83.

Biographical notes: Catarina Maçãs has an Undergraduate and Master's in Design and Multimedia from the University of Coimbra. She is currently a computational designer and researcher at the Computational Design and Visualization Lab (CDV), which is part of the Centre for Informatics and Systems of University of Coimbra (CISUC), is enrolled in the Doctoral Program of Information Science and Technology of the University of Coimbra, and is also a teaching assistant at the Department of Informatics Engineering of the University of Coimbra. Her research is based in the area of information visualisation, more specifically in the visualisation of temporal events for the understanding of patterns and trends.

Nuno Lourenço is an Assistant Professor (non-Tenure) at the University of Coimbra. He has been with the Centre for Informatics and Systems of University of Coimbra (CISUC) since 2009, a member of the Evolutionary and Complex Systems (ECOS) group. His main research interests are in the areas bio-inspired algorithms, optimisation and machine learning. He is the creator of structured grammatical evolution, a genotypic representation for grammatical evolution, and participated in the proposal of a novel approach to automatically design artificial neural networks (ANNs) using evolutionary computation called DENSER. He is member of the executive board member of SPECIES. He has authored or co-authored several articles in journals and top conferences from the evolutionary computation and artificial intelligence areas.

Penousal Machado is an Associate Professor at the University of Coimbra, the Deputy Director of the Centre for Informatics and Systems (CISUC), coordinator of the Cognitive and Media Systems group and the Scientific Director of the Computational Design and Visualization Lab. His research interests include evolutionary computation, computational creativity, artificial intelligence and information visualisation. He is an author of more than 200 refereed papers, chair of several scientific events, and member of the program committee of numerous conferences. He is the recipient of several scientific awards, including the EvoStar Award for Outstanding Contribution to Evolutionary Computation in Europe. His work was featured in the *Leonardo Journal*, Wired, and included in the 'Talk to me' exhibition of the Museum of Modern Art, NY (MoMA).

This paper is a revised and expanded version of a paper entitled 'Interactive evolution of swarms for the visualisation of consumptions' presented at ArtsIT/DLI 2018, Braga, Portugal, 24–26 October 2018.

1 Introduction

Information visualisation has its roots in scientific reasoning and is usually seen as an analytical tool. However, with the democratisation of tools which enable the creation of visualisations, the conceptual boundaries of information visualisation were expanded to a more exploratory and user-oriented field (Friendly, 2008). Furthermore, with the increasing necessity to collect data within big enterprises, the interest in information visualisation tools has also increased and companies from all lines of work began to take interest in exploring their data to improve the knowledge over their own business sectors (Keim et al., 2013).

In a partnership with a Portuguese retail company, we had access to a high volume of data about the Portuguese consumption. With this dataset, they asked us to research, propose, and develop a set of visualisations models and to explore two different dimensions – analytical and aesthetics. In what concerns the analytical aspect, we developed a series of visualisation tools to optimise their operations at different levels, ranging from the understanding of how the consumption values are distributed along their product hierarchy, to the finding of the optimal location to open new supermarkets. Regarding the aesthetic dimension, the company's goal is to communicate with a wider public, allowing them to visualise their consumption data in an appealing manner and providing them with an aesthetic experience. With this, the company intends to give more visibility to their projects and their investment in scientific research, and at the same time, straighten the relation with their clients. This is aimed to be a step towards captivating the big public, through aesthetic experiences.

It is in the context of an aesthetic exploration that the present work arises. With the data of the Portuguese's consumption routines, we aim to create emergent visual artefacts driven by those routines, positioning our work in the field of data aesthetics (Manovich, 2000; Moere, 2007). The dataset is rich in daily, weekly and monthly repetitions of consumption patterns, offering us the opportunity to transform the Portuguese consumption values into visual artefacts, while exploring, highlighting, and visualising their periodic nature.

Taking into account a previous project (Maçãs et al., 2015a), we use a swarm system to create emergent visual artefacts (Jones, 2007). In the referenced work, the user had to define the parameterisations of the system to create a balance between a more functional or more aesthetically intriguing visualisation. In this project, this parameterisation is automatically defined through the use of an interactive evolutionary framework, opening the possibilities to create a wider range of visual solutions. Whence, in this project the main concern is to create artefacts that are aesthetically appealing to the user, and not artefacts placed in the functional spectrum.

The result of this exploration is an interactive framework that is able to create a wide and diverse set of solutions. We test the validity of the system through different usage scenarios in which the system relies on the users' preferences as an input to guide the solution towards what they find attractive. We defined three objectives for the user guidance: explore specific parameterisation attributes; guide the evolution to a functional artefact; explore randomly the system. Additionally, we used an automatic function to qualify each solution and test the capacity of the algorithm to traverse the search space. We also tested the resulting visual artefacts of the new system through an user study, to understand what is more appealing to the general public, and to have a first validation on the hypothesis of using this system as an avatar generator, that creates unique avatars for each customer, representing its consumption patterns without being too intrusive.

We were able to achieve the goals of the company, since the developed framework can be used by an experienced user that understands the parameters and the possibilities of the visualisation model, as well as by a user that has no knowledge of the system and only selects artefacts visually relevant to his/her own tastes.

1.1 Background

This article is part of a larger project developed in association with a Portuguese multinational company – Sonae – that manages a diversified portfolio of businesses in retail, financial services, technology, shopping centres, and telecommunications. Through the years, they have gathered numerous distinctions in the areas of sustainability, efficiency and innovation, making them one of the most important and well-known companies in Portugal.

For the present project, Sonae gave us the opportunity to explore the data about the Portuguese's consumption in their retail chains. The dataset consists of 278 GB of information about customer purchases in 729 Portuguese supermarkets and hypermarkets of the company's chains, in a time span of 24 months (from May 2012 to April 2014). This dataset also includes the geolocation of 682 supermarkets, the regions of the country to which they belong, and information about the transitions of products between warehouses and supermarkets over a time span of six months.

While shopping in the retail stores, costumers tend to use their client cards to accumulate discounts and other benefits. This enables the company to create personalised discounts and to aggregate data by specific geolocations. Currently, the number of active cards is above 6 million, which is an impressive number, especially if we take into consideration that the Portuguese population is below 11 million and that the cards are issued by 'household' and shared by the entire family. The dataset comprises approximately 2.86 billions transactions where each transaction has the following attributes: customer card id, the amount spent, product designation, the quantity of the purchased products, and date and time of the transaction. Each product is placed within the product hierarchy of the company, which has six levels – department, BizUnit, category, sub-category, unit base and product.

This dataset triggered our interest by its richness, size, quality and nature. We believe it offers us the opportunity to explore the Portuguese's consumption patterns analytically, highlighting and visualising their periodic nature to ease their understanding. Notwithstanding, we also believe that it is a good opportunity to transform the Portuguese's consumption into aesthetic artefacts that intrigue the users and entice them to further explore the data, promoting the company.

With this high number of collected data, the need to analyse and make sense of it, is of the utmost importance. Data visualisation is an invaluable asset in this process, and therefore, it is the main focus of our research. We believe that, through the visual representation and inherent simplification of complex datasets, we can improve the analysis and enhance the understanding of the data, improving, as a consequence, the company's business.

As a guideline for the project, the company was interested in two main dimensions, one related to the analytical analysis of their data and another related to aesthetics, giving us the opportunity to freely explore the last. In the analytical dimension, and with the available dataset, the company's goals were to:

1. understand how consumption evolves through time in specific regions of Portugal, allowing the understanding of seasonal variations (Maçãs et al., 2016)
2. visualise the transition of customers among supermarkets over time (Polisciuc et al., 2015)
3. identify potential sites to open new supermarkets (Polisciuc et al., 2016a)
4. visualise the transition of products between warehouse and supermarkets (Polisciuc et al., 2016b)
5. represent the evolution of consumption within the company's product hierarchy, with emphasis on periodic behaviours and their disruptions (Maçãs et al., 2015b).

A set of works were already made in the analytical dimension. These works can be divided into two sub-areas of Information visualisation: geo-visualisation and small-multiples. In the first, the main goal is to represent in a map the evolution of a certain data variable. Next, we will present a short resume of the work done in those sub-areas.

In Maçãs et al. (2016), the goal was to represent geographically the evolution of consumption, enabling a quick understanding of how the consumption distribute throughout the entire country, as well as within each specific county district. To do so, we applied an Isoline technique to distinguish the areas with different consumption values. By adding a timeline, in which the user can change the period of time of the visualisation, we emphasise the changes in the patterns of consumption at specific times of the year. Additionally, this enabled the detection of areas which are not covered by the company's hypermarkets and the highlight of areas of low consumption and high population density. This was a first step in the development of a visualisation tool to assist decision-making process regarding the opening or renewal of facilities.

In Polisciuc et al. (2015), the main goal was to reveal the flow patterns of customers' transitions among supermarkets over time. In this project, a transition is characterised by the change of a transaction's location in the customer's shopping history. The visual exploration of these transitions enable the understanding of the amount of transition flows among geographical locations on a global perspective, as well as on specific geographical areas and time spans. In this project, we explored two techniques to visualise the transitions' flow. The first, relied on an arc representation, which proved to have a high density of clutter and the second technique was based on a swarm-based system, improving the former, reducing visual clutter, and enhancing clarity.

In Polisciuc et al. (2016a), the goal was to improve the identification of potential sites to open new supermarkets. We implemented a new approach to hexagonal gridded maps to enhance the used space and facilitate the accurate positioning of information on the map. We also defined a set of different information layers. Those layers were based on dot maps and were coloured depending on the type of information being presented. This kind of visualisation allows the company's analysts to explore and analyse in the same visualisation tool:

1 the frequency of consumption in certain areas

2 the impact of the supermarkets' location on customer preferences

3 the areas of low coverage by supermarkets.

In Polisciuc et al. (2016b), the aim was to create flow maps representing the products being transported from warehouses to supermarkets. For this project, we used the dataset concerning the warehouse-to-supermarket transitions. To represent the flow between supermarkets and warehouses, we used a swarm-based algorithm to represent the large amount of transitions from one location to another. This method uses a customised swarming system to trace edges in an intuitive and organic fashion. To further improve clarity, we also applied a technique known as Dorling Cartograms (Dorling, 1996). With this technique, the overlapping points were separated but retained some degree of spatial relationship.

In Maçãs et al. (2015b), the main goal was to represent the evolution of consumption values in a single chart. In addition, we wanted to provide the mechanisms to:

1 visually explore the consumption evolution over time within the company's product hierarchy

2 detect periodic behaviours

3 enable the comparison between different days.

Moreover, we aimed at using the display space efficiently, maximising data density and minimising the use of 'ink'. To do so, we applied the small-multiples technique (Tufte, 1991) to enhance the comparison among consumption days while providing a general overview of the annual behaviours. This calendar view, represents the daily consumption deviations from a weekly baseline, thus, highlighting these deviations along time, eliminating their intrinsic periodic behaviour, and emphasising moments of disruption.

With regard to the aesthetic dimension, the company's aim was to lure the user to explore their data and to be involved. They wanted to create a closer relationship with their customers through the creation of appealing and entertaining visual artefacts. This relation is also emphasised by the fact that it is the customer itself that, in a certain way, creates the artefacts, or at least, made them possible. Therefore, we intend to create artefacts that, with the simplification of the complex dataset, can be easily captured by the more general public.

The developed works in the aesthetic dimension focus on the representation of the consumption values over the time span of two years. Although they are intended to explore the aesthetics of visualisation, they are also concerned with functionality, trying not to overpass the barrier of legibility. In Maçãs and Machado (2016), we explored how the visualisation can morph depending on data, creating movement and highlighting the rhythm and disruptions of the normal consumption patterns. We represent changes using two layers of visual variables: colour and size. With this, our goal is to ease the understanding of the evolution of the general shape, and consequently of the datum. For this project, to overview the customers shopping habits and how their priorities change over time, we characterised the transactions in the company's chain into three different types of consumption – essential, non-essential, and unknown. We also explored two different approaches, through a small-multiples technique and through a video. These two representations have different objectives: while the first is mainly analytical, the second focuses on captivating and entertaining the users, highlighting differences among days through behavioural changes. In the static representation, we can see how the consumption values are affected by special events and vacations, and easily compare different days that are distant from each other. In the video, we can, above all else, perceive the rhythmic patterns of consumption, marked by the weekends and the disruptive effects caused by special events.

In Maçãs et al. (2015a), we explored the qualitative representations of data, providing an overview of how data behaves through time. We apply a swarm-based system as a method to create emergent visualisations of the consumption values with the intent to convey meaningful information and, at the same time, explore the boundaries between data visualisation and information aesthetics. The application of swarms systems to visualise data can also be seen in Moere and Lau (2007), Ogawa and Ma (2009) and Kamvar and Harris (2011). Additionally, in the field of generative art, swarm forces were also used in a variety of projects (Bornhofen et al., 2012; Greenfield and Machado, 2014). We focused on the ability of this emergent system to communicate information while engaging the viewer with organic visuals (Jones, 2007). Additionally, with different parameterisations, we were able to create a set of renderings with different levels of legibility and attractiveness. The approach presented in this article builds on this work, by developing an interactive framework that is able to evolve the configuration of the visualisation model through the use of an evolutionary algorithm (EA) (Eiben and Smith, 2015). Our goal is to create new, diverse, and surprising visual artefacts, which, in spite of not being completely functional, are engaging and entertaining for the user.

1.2 Our contributions

In the present article, we present our research which intersects the fields of data aesthetics and evolutionary algorithms. Our main goal is to develop a computational framework that is able to automatically generate visuals artefacts that represent analytical data in a more aesthetic manner. To this end, we describe the following:

- The interactive evolutionary framework used to evolve visual artefacts that can be placed in the data aesthetics field (Section 2).
- The validation of the interactive framework through the definition of an automatic fitness function (Section 3).
- The use of the framework with three different goals to show the interaction with the system and to analyse the generated results (Section 4).
- The validation of the generated artefacts through an user study with the main aim of perceiving the how the general public reacts to the diversity of the artefacts (Section 5).

This paper is built upon previous work (Maçãs et al., 2015a), which proposed and discussed the development of the interactive evolutionary approach. In concrete, we extend the previous work, by validating the evolutionary algorithm with an automatic fitness function. Then, and to validate the ability of the system to generate artefacts that are aesthetically appealing to the general public, we also conduct an user study.

2 Interactive evolutionary approach

To further explore the aesthetic dimension, and to improve the process of parameterisation of the swarm system detailed in Maçãs et al. (2015a), we couple the swarm system with an EA. By doing so, we increase the degree of freedom during the creation of the visual models, enabling a wide and diverse range of solutions. Note that, instead of being interested in creating a balanced artefact between aesthetics and functionality, we are only concerned with the former. Additionally, we add the advantages of IEC to provide the user with the ability of guiding the system in the creation of visual artefacts according to his/her taste. These new artefacts are intended to amuse the viewers and lure them into decode the visualisation, or, at least, to entice them to further explore the framework. Our intention with this new approach is to develop visual artefacts that are continuously readjusting to the intentions of the user, evolving to a state of complete interest to him/her. Additionally, we aim to enable the user to explore artefacts not imagined before by him/her. Thus, the company can deploy the system to different audiences, and the system will evolve and adapt to their aesthetic preferences.

2.1 Visualisation model

The visualisation model used for this project consists of a swarm system that, through different forces of separation, attraction, and cohesion, creates emergent visualisations (Jones, 2007) about the Portuguese's consumption routines. This system is constituted by several boids –artificial objects that simulate the flocking behaviour of birds – Reynolds (1987) in an environment (i.e., the canvas), that react to the changes in consumption over time.

To simplify the dataset, we aggregated all transactions (products bought in a super or hypermarket) by the highest level in the product hierarchy, i.e., the department. There are a total of seven different departments in the dataset: grocery (biscuits, cereals, frozen foods, hygiene, and cleaning products), fresh food (meat, fish, vegetables, and fruits), food and bakery (bread, cakes, and coffee), home (household essentials), leisure (books, office supplies, and DIY), textile (clothing), and health (products from nutrition to beauty). Each transaction has the hour, minute, and second of purchase. However, the representation with this degree of detail would be too subtle to the human eye. Therefore, we aggregate the data in intervals of two hours, resulting in 12 intervals per day. Although the transactions occur only between 8 am and 10 pm, we keep the representation of the 24 hours aggregated every two hours to enhance the gap between days, facilitating the distinction between different days.

2.1.1 Swarming forces

The swarm system simulates the behaviour of multiple boids (Reynolds, 1987). In our visual artefacts, each boid represents the consumption in a specific Department, is represented through a circle and is described by properties such as velocity, position, size, and colour. Only the position and velocity of the boids are affected by the swarming forces. Based on the work of Reynolds (1987), each boid follows three basic rules:

1. cohesion
2. separation
3. alignment.

To explore the system, we applied different values to each force, so it was possible to create different outputs. For example, if two neighbouring boids are from the same department, we apply higher forces of attraction and lower forces of separation, but, if they are from different departments, the attraction force is lower then the previously defined, and the separation force higher.

To prevent the boids from randomly moving on the canvas and to enhance their periodic behaviour, we defined a target that all boids should look for. Hence, in addition to the previously described forces, all boids are under the influence of an attraction force towards this moving target. The target, although not represented visually, starts from the centre of the canvas and swirls around, creating a spiral with equal distances between each lap.

To represent the passage of time, we consider that each lap of the swirling boids represent one month of the dataset. Then, depending on the angle of the boid with the centre of the canvas, we obtain the day and corresponding hour of the month. To do so, each lap is divided by the 31 days multiplied by 12 hours ($2\Pi \div (31 \times 12)$), since, as previously stated, the data is aggregated in intervals of two hours. Note that all laps have 31 days. By doing so, all months start with the same angle (at the top of the circle), and, if they have less than 31 days, the consumption values are null – and the circles are almost invisible – during those non-existent days.

2.1.2 Rendering

As stated before, each boid is represented through a circle that has a specific size and colour. While colour identifies the department to which the boid belongs (Figure 1), the size represents the consumption at a certain time: the bigger the circle, the bigger the consumption value. As the boids wander through the canvas, they leave a trail of their shape, enabling the user to see its path, and consequently, interpret the consumption values. The boids' size is mapped to a predefined minimum and maximum radii, that can represent the minimum and maximum consumption value of each individual department (local normalisation), or the minimum and maximum consumption value of all departments (global normalisation). Additionally, we defined three different styles to represent the boids: through a filled circle, through the outline of the circle, and through a line that connects all boids belonging to the same department. Additionally, we implemented a mechanism to sort the circles in depth according to their radii. With this, the smaller circles are drawn over the larger ones and are never hidden by them.

Figure 1 Colours used to distinguish the seven departments of the company's product hierarchy (see online version for colours)

2.2 Evolutionary algorithm

The visualisation model described above is easy to understand and use, yet it requires the definition of several parameters in order to create visual artefacts that can be appealing for the user. Therefore, the understanding of these parameters is an important asset to maintain the balance between functionality and aesthetics. If they are not selected with care, we might end with a model that is unreadable and/or visually uninteresting.

To aid in the task of creating appealing and/or functional visualisation models, we propose a framework based on EAs (Eiben and Smith, 2015). This framework will relieve the burden of searching for parameters that balance functionality and aesthetics. To evolve the swarm system, we will be searching for the best combination of the following system's parameters:

1. the separation, alignment, and cohesion forces
2. the minimum and maximum radius

3 the use of a global or local normalisation

4 the representation modes (lines, filled circles, transparency, sorted circles).

Note that the boids' size is always mapped according to the consumption value on the data. In the following subsections, we present the parameters used to evolve the visual artefacts and the used genetic operators.

2.2.1 Representation

Each EA solution is encoded as a set of values that correspond to the number of parameters needed by the swarm system. In concrete, we have ten different parameters that are required by the visual model:

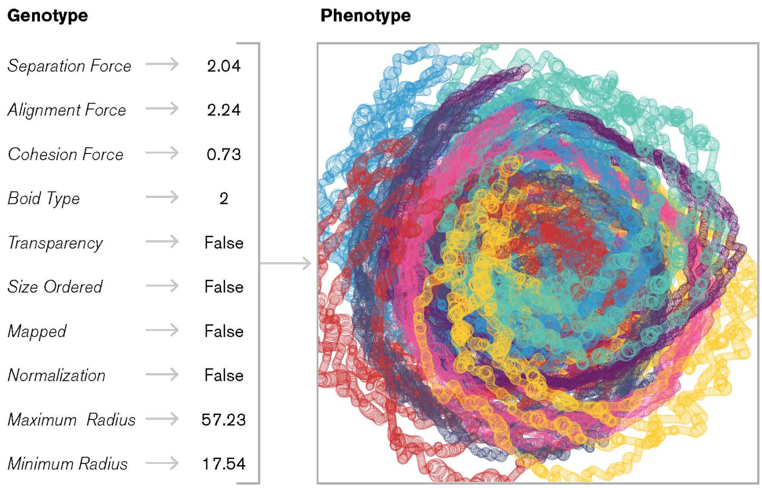
- *Separation force:* A real value between 0 and 3.
- *Alignment force:* A real value between 0 and 3.
- *Cohesion force:* A real value between 0 and 3.
- *Boid render:* One of the following options: lines, circles and filled circles.
- *Transparency:* A boolean value that enables transparency of the boids.
- *Size ordered:* A boolean value. If this value is true, the visualisation model sorts the boids by radius, i.e., boids with smaller radius will be on top of boids with larger radius.
- *Mapped:* A boolean value. If it is true, it indicates that the separation force is mapped depending on the radius of the boids.
- *Normalisation:* A boolean value that enables normalisation based on the maximum sales values.
- *Maximum radius:* A real value between 30 and 80 that corresponds to the maximum radius of the boids.
- *Minimum radius:* A real value between 0.1 and 15 that corresponds to the minimum radius of the boids.

An example of a possible solution alongside with its phenotype representation is depicted in Figure 2.

2.2.2 Genetic operators

To promote the evolution and the proper exploration of the search space we rely on two operators: recombination and mutation. The recombination operator is the uniform crossover and combines two solutions by creating a random mask of the same size of the genotype, and then swap the genetic material according to the previously generated mask. Regarding the mutation operator, we apply a per gene mutation to the candidate solutions. This allows the algorithm to change, from generation to generation, a significant percentage of the genes to other valid ones.

Figure 2 On the left, the genotype of a possible solution; on the right, the resulting phenotype, i.e., the visualisation model (see online version for colours)



2.2.3 *Fitness evaluation*

The main goal of the EA is to promote the creation of solutions based on the users' preferences. To accomplish that, practitioners often resort to IEC systems, which ask for the user's input to rank the solutions that are being evolved depending on their preferences (Takagi, 2001). However, we also defined a simple metric to evaluate the capacity of the algorithm to traverse the search space, and to find if the EA is working properly. In concrete, we defined a fitness function in which the quality of the individual is proportional to the PNG file size, after the rendering of a solution. The larger the file size, the better the individual.

3 Experimental results

We will use the algorithm detailed above to address the ability of the automatic fitness function to promote evolution towards visualisation models that result in images with large sizes in terms of storage. To give some coherence between the explorations, we defined a fixed period of time of four months, from October 2012 to January 2013. We chose this period of time since it contains moments of high consumption, such as during Christmas and New Years periods, and more calm periods of time, such as in January (after the holidays).

3.1 *Experimental setup*

Table 1 details the EA parameters used in the experiment conducted in the following sections. All the values, except the number of generations, are kept fixed throughout the experiments. We use high mutation and crossover rates so that the individuals in each generation have noticeable differences, promoting a diverse set of visualisation models. This enables a faster convergence towards feasible solutions, i.e., the emergence of individuals that are considered of high quality by the user or by the automatic function. Additionally, parent selection is performed using tournament selection.

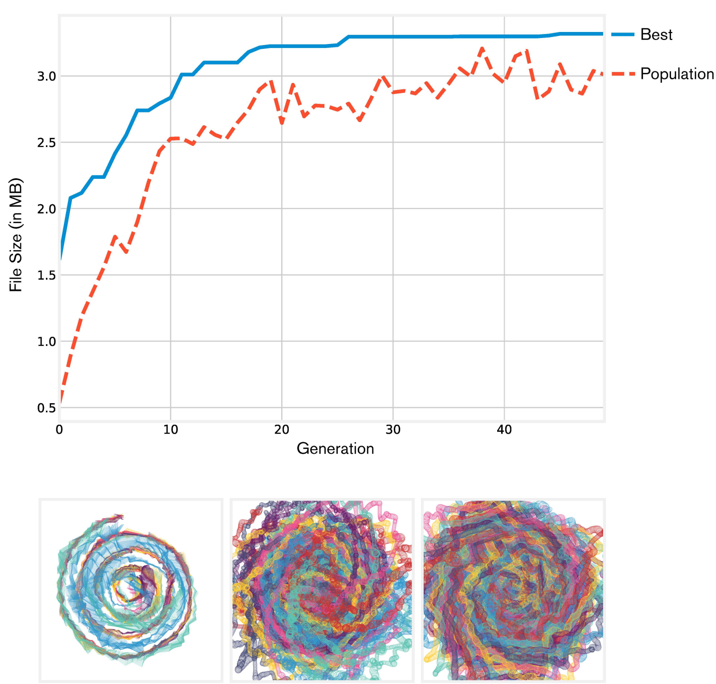
Table 1 Experimental parameters

Parameter	*Value*
Number of runs	30
Population size	20
Number of generations	50
Crossover rate	90%
Mutation rate	50%
Elite size	3 individual
Tournament size	3 individuals

3.2 Automatic fitness results

The first set of experiments focus only on the use of the automatic fitness component. As such, we will analyse the capacity of the fitness function to guide the EA towards solutions with visualisation models that take a large amount of space when stored on the hard drive. A video showing the best individuals for this experiment can be seen in: https://vimeo.com/289093672.

Figure 3 Automatic fitness results (see online version for colours)



Generation 0
File size: 1.4MB

Generation 24
File size: 3 MB

Generation 40
File size: 3.2MB

Figure 3 shows the evolution of the quality of the solutions, i.e., the size of the files in megabytes (MB), across 50 generations. The results are averages of 30 runs. An inspection of the results shows that the system is able to converge towards regions of the search space with visualisation models that have large file sizes. Looking at the fitness of the best solutions (solid line), it is possible to see that they have the smallest size in the first generation. Between the 1st and 10th generation, we see a rapid increase in size, which stagnates around the 20th. After this point, evolution is still occurring, but at a slower pace. The fitness of the population follows the same trend, i.e., it increases rapidly in the first generations, and roughly after the middle of the evolutionary process, it stagnates between files with a size of 2.6 and 3.1 MB.

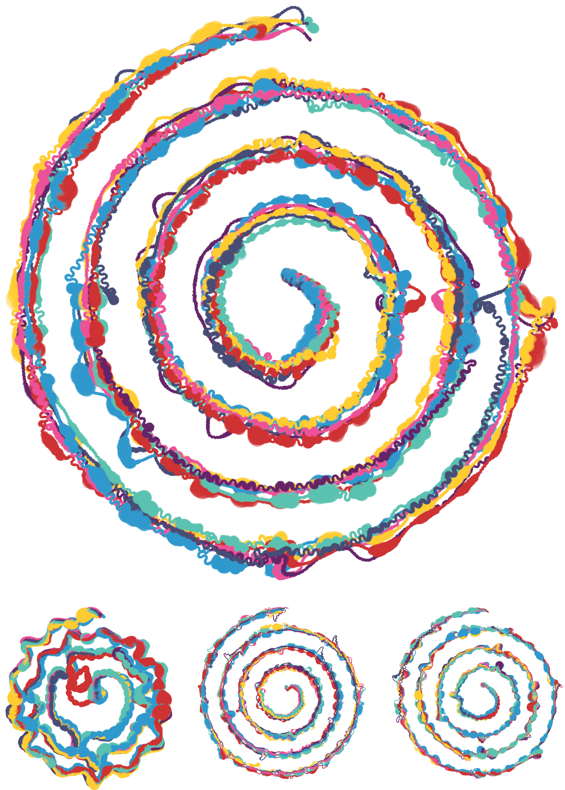
Looking at the phenotype of the solutions (Figure 3, bottom panel) it is possible to see that they start with something that has redundant information since there are not many differences in the colours of the pixels (e.g., many parts of the image are white). Since the PNG uses the Deflate algorithm to compress the image, in order to increase the file size we need to reduce the compression rate which is done by increasing the number of different pixel colours. This is precisely what is happening during the evolutionary process. In the first generations, the system explores a wide variety of solutions, based on a small set of colours, and with a cleaner canvas. As the evolutionary process progresses, the individuals start to be more visually cluttered (augmenting the density of the coloured pixels) and to use a wider palette of colours. This increases the file size of each individual, due to the fact that the Deflate algorithm is not able to compress the information (i.e., pixels colours) if they are not redundant. Note that we are not changing the palette colours, to represent the different departments. These differences emerge from the combination of the different parameters in the genotype such as normalisation and transparency which influence the used colours.

4 Usage scenarios

In addition to the previous results, we explored the evolutionary system based on the user's preferences through IEC. We implemented a simple interface to enable the interaction with the system. A video showing the interaction with the system can be seen in: https://vimeo.com/289093672. These explorations were based on three different objectives and different types of users. In the first exploration, the goal is to evolve solutions with specific parameterisation attributes: the boids must be represented with filled circles, use the local normalisation, and have a zigzagging pattern. In the second, the goal is to intersect the functional dimension, enabling the readability of the artefacts. Finally, for the third exploration, there is no predefined objective, so the solutions must be diverse according to the user's taste. For the first two explorations, the user must have some experience on how the system works, its parameters and the data. In the last, the user needs no experience with the data nor with the parameterisations, the user only explores the possibilities and creates artefacts suitable for his/her own taste. As the generations evolve, the user chooses only the visual artefacts that visually intrigue, amaze, and/or correspond with his/her preferences.

To avoid user fatigue, we used a reduced number of individuals per generation. For each exploration, the number of individuals is set to 20 and the maximum generations to 10 (this value can be increased). With this, we aim to enable the creation of various solutions per generation, giving the user multiple possibilities, without causing a sensation of weariness. Additionally, the user can stop the evolutionary process at any given time if the visual artefacts are in accordance with his/her expectations.

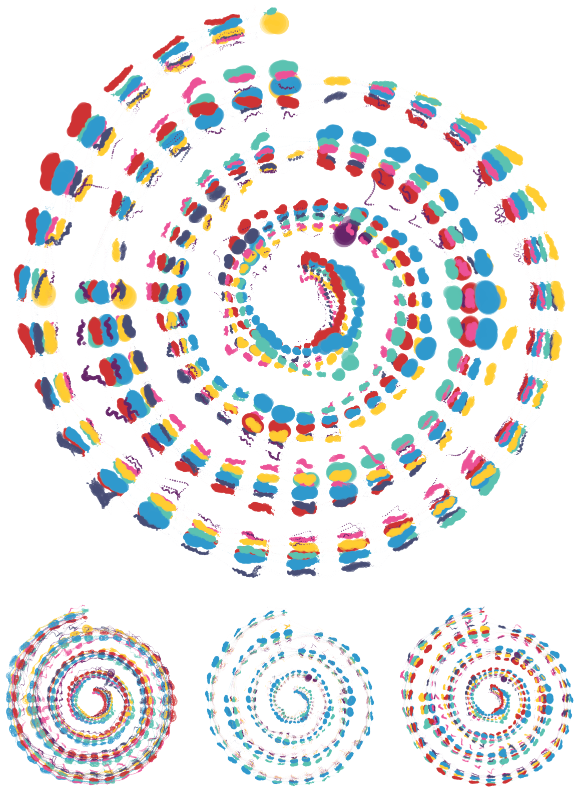
Figure 4 Chosen individuals by the user during the interactive evolution (see online version for colours)



Notes: In this example, the user aimed to create artefacts with a balanced colour palette. The shown individuals were selected in the 10th, 1st, 3rd, and 6th generations, respectively from up to bottom, left to right.

In the first exploration, and as the user had already a target solution in mind, it was possible to perceive that the evolutionary system was evolving correctly towards the user predefined goal (Figure 4). As the individuals were being created, the user selected only the ones with filled circles. The user also chose only the ones where all colours appeared balanced, meaning the user chose only artefacts which uses the local normalisation. When using the global normalisation, the artefacts colours consist mainly of blues and greens – the departments represented by those colours are the ones with higher consumption values, and thus, the ones with more visual presence. With the local normalisation, each consumption value is mapped between zero and the maximum consumption value of the represented department, enabling a more balanced distribution of colours.

Figure 5 Chosen individuals by the user during the interactive evolution (see online version for colours)



Notes: In this example, the user aimed to create artefacts that can be placed in the functional dimension. The shown individuals were selected in the 10th, 1st, 3rd, and 6th generations, respectively from up to bottom, left to right.

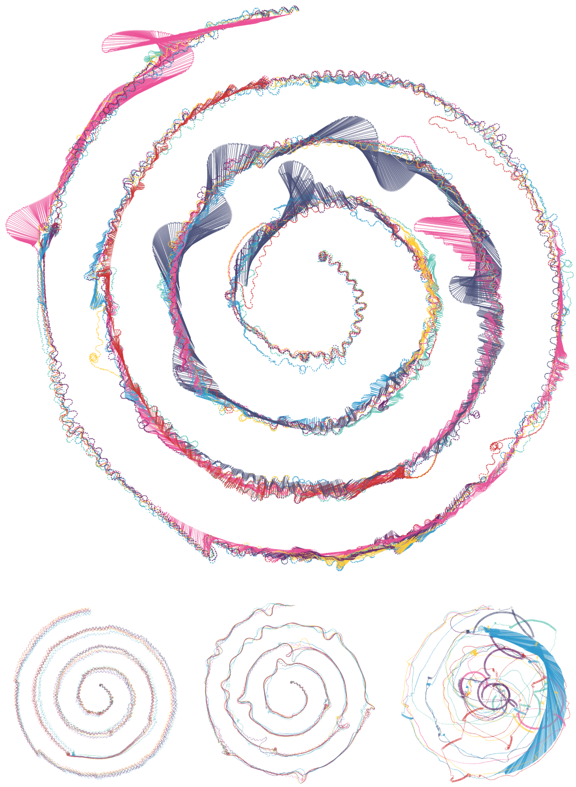
For the second exploration, and with the intention to guide the evolutionary system to generate artefacts which can be aesthetic and at the same time functional, the system also proved to be capable of evolving good solutions (Figure 5). In the last generations, the majority of the individuals were readable. This means that they had less clutter and lower separations forces, which enables them to have a more strict behaviour as they swirl, not deviating from the spiral path. Additionally, the majority of the generated artefacts are similar to the explorations made in the previous project (Maçãs et al., 2015a). For the majority of the generations, there would still appear cluttered visual artefacts, reducing its functionality, but introducing some degree of novelty. Furthermore, if the user continued to further explore the system, these non-functional artefacts could contribute to new findings over the data. In this exploration, it was also possible to see an artefact which was never created by the previous system, but could be inserted in the functional dimension (Figure 6). In this artefact, the user can visualise the moments in which each department has its higher consumption values. This is accomplished by the predominant colour throughout the spiral. For example, in the second lap, we can see a wide purple circle, which corresponds to the consumption in the Leisure Department during a promotional event on 3 and 4 of November 2012 in which all toys had a 50% discount. It is also possible to perceive the highest days of consumption in the Food and Bakery Department on the 24 and 31 of December. These growing consumption are explained by the festive seasons of Christmas and New Year. This outcome is a remarkable example of how the evolutionary system can work to aid in the design of functional artefacts.

Figure 6 An individual that intersects the functional dimension (see online version for colours)

Note: This solution was retrieved from the second exploration, in which the user aims to create functional visualisations.

For the last exploration, the user had no predefined objective. As the artefacts were being created and presented, the user chose freely the preferred ones. In the beginning, there was no guidance as the user had no previous knowledge of the system, not knowing the possible outcomes and how to achieve them. In this way, the user chose only the ones which she found visually intriguing. In the end, the user managed to guide the system through a specific style, the use of lines to represent the departments (Figure 7). As the solutions appeared, the user constantly opted for the ones with lines and with swarming forces which make the boids to deviate from the spiral path, as can be seen especially in the bottom right image of Figure 7.

Figure 7 Chosen individuals by the user during the interactive evolution (see online version for colours)



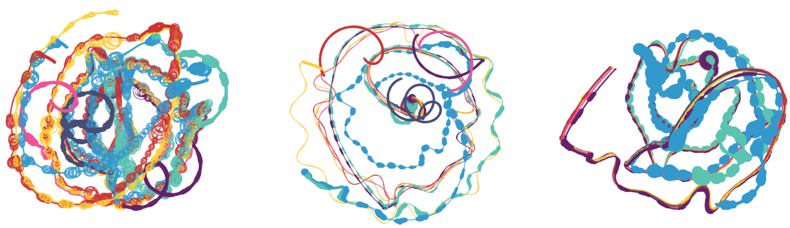
Notes: In this example, the user has no predefined target solution, choosing only the solutions which fit his/her tastes. The shown individuals were selected in the 10th, 1st, 3rd, and 6th generations, respectively from up to bottom, left to right.

As a summary, we found that the system was capable of discovering artefacts similar to the ones created by human designers, but also to find new ones. Additionally, it was capable to create a new functional artefact, that was not achieved previously by human designers. In conclusion, our explorations showed that the system can be guided by the user, showing similar artefacts to the ones the user chose previously, and, at the same time, giving to the user others to enhance his/her choices.

4.1 Results' diversity

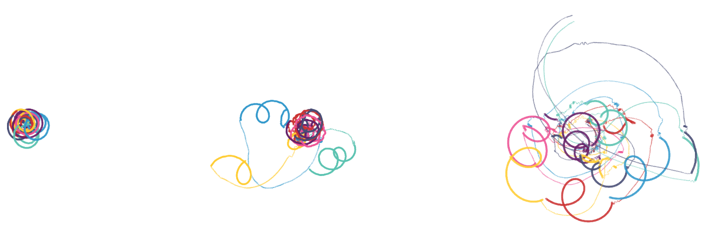
One of the important aspects that we looked for in the previous experiments was the capacity of the system to generate and evolve a diversified range of visual artefacts. The random initialisation of the visualisation models creates a wide variety of behaviours. For example, if the forces are too strong, the boids will deviate from the spiral path, creating random zigzagging patterns (Figure 8). In a similar way, these forces can also cause the boids to stagnate their position in the centre. In Figure 9, it is possible to see how the system generates a set of individuals, based on a model that was selected by the user. In the beginning of the evolutionary process, the boids are trapped, swirling around without distancing themselves from the centre of the canvas. This behaviour emerges due to the forces applied to the boids. As the generations go by, the parameters change, which will result in the modification of the boids' behaviour. This is caused by the fact that as the user selects the individuals, they have higher chances of being chosen to produce descendants for the next generation. These descendants will be visually close to the previous chosen individual once the crossover and mutation operators inflict subtle modifications in the genotype.

Figure 8 Individuals that deviate from the spiral path (see online version for colours)



Note: This is caused by the strong swarming.

Figure 9 Different individuals with similar parameterisations (see online version for colours)



As the system evolves the parameters, without the direct intervention of the user, it can generate solutions which the user was not expecting and has not seen before. Even an experienced user, that knows the parameters and the system itself, can be surprised by the solutions found. As it can be seen in Figure 10, the developed system is able to create a diverse set of emergent visual artefacts. For example, if the system chooses the lines to represent the departments, the visual artefact can even lose its swirl effect, as the boids of the same department can be distant from each other, generating long lines throughout the canvas, and eliminating the spiral path (artefacts in the bottom and upper right corners of Figure 10). This last figure presents only a small set of the possible variety of visual solutions that our system can create. As the users' objectives change, the system evolves other visual solutions, creating an undetermined number of different artefacts.

Figure 10 A small set of the variety of visual artefacts that the evolutionary system can create (see online version for colours)

Note: Depending on the user guidance this set can be extended to other visual solutions not presented in this article.

5 User study

To validate our system with real users we performed an user study. The objective of this study is to validate the ability of the EA to create visual artefacts that are aesthetically appealing, i.e., that are visually pleasing, and to create a diverse set of artefacts. The study was conducted using an online survey, where the participants were asked to perform three main tasks. In the first, they were asked to select between two images: one that was generated using the evolutionary framework and another that was taken from the previous system, which generates visual artefacts using the same system, but in which the parameterisation is defined manually and is intended to explore the functional domain of the artefacts (Maçãs et al., 2015a). The used images were randomly paired. To select the previous system's artefacts we defined that they should intersect the functional domain, which results in images with fewer deviations from the spiral path.

In the second task, we asked the users to grade, from 1 to 5 the diversity of a set of images (Figure 10), where 1 corresponded to low diversity and 5 to high diversity. Additionally, we asked the users to rate the set of images depending on their aesthetics (Q13) and their ability to captivate the user (Q14), both in a scale from 1 to 5, being 1 not aesthetically pleasing or not captivating, and 5 very aesthetically pleasing or very captivating.

Finally we asked the users if they would use the images as an avatar. Our main purpose with this last question was to understand if the images were interesting enough to the participants, so they would use them as a virtual public display of their preferences. This is a first step in perceiving the applicability of the system in a real world situation.

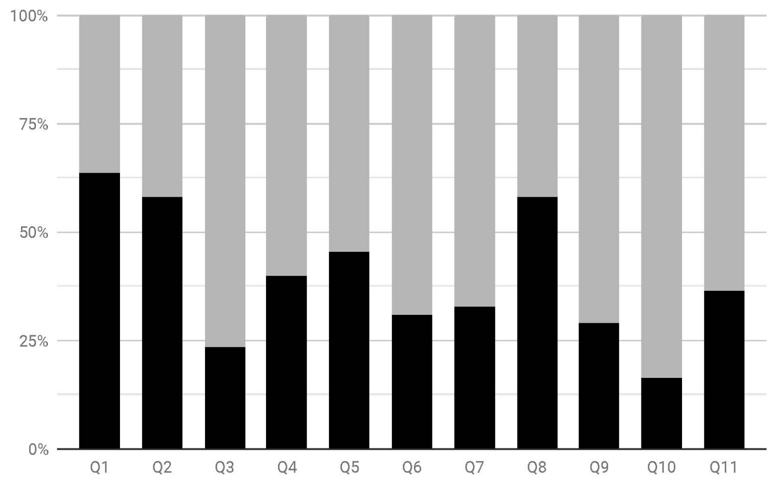
The study sample was composed of 56 participants with ages between 21 and 66 years old, with a median of 27. Their background ranged from computer science, design, psychology, and economy. The questionnaire can be found here https://goo.gl/forms/XlhFL66GhIN7iRkI2.

5.1 Results

After gathering the questionnaire results, we analysed the answers using the statistical software SPSS version 24. To analyse the questions we used the χ^2 test since all of our data is categorical. We considered a significance level of $\alpha = 0.05$. The distribution of the answers per questions is depicted in Figure 11. The black portion of the bars corresponds to the percentage of participants that selected the image generated by the evolutionary algorithm. As we can see by the results there are six questions (Q3, Q6, Q7, Q9, Q10) where the users clearly select the image that was not generated by the EA. On the contrary in Q1, Q2, and Q8 the users clearly selected the image that was generated by the proposed approach. In both situations the results have statistical meaningful differences. In all other cases, there are no clear preferences. Based on these outcomes, we analysed the differences between the images presented in the questions, and tried to understand why such differences existed. In the first task, the users tend to select images that are more packed and coherent. This is evident by the answers to Q6, Q7, an Q10. However, in Q8, both images had patterns that deviated from the normal spiral path, and the users had a preference for the EA artefact.

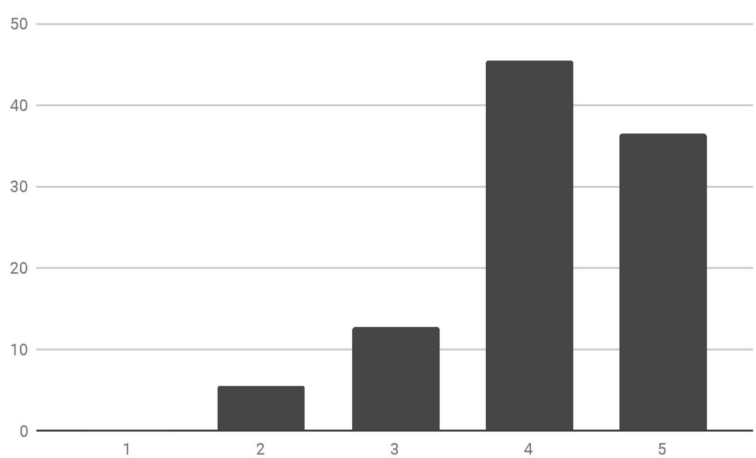
Concerning the second task, where the user was asked to evaluate how diverse a set of images was (Figure 13), the results show that, in general, the users found the images diverse, which gives an indication of the capacity of the EA to generate different images. Then, and in order to understand if the images were appealing, we showed the same set of images and asked the users to grade, from 1 to 5, the pleasantness of the images. Results are depicted in Figure 12. The results obtained in this question, together with the ones from Q12, indicate that the EA is capable of generating images that are diverse and at the same time are visually interesting.

Figure 11 Distribution of answers in the first task of the questionnaire



Note: The black portion of the bar corresponds to the percentage of users that selected the artefact created with the EA.

Figure 12 Distribution (in percentage) of the answers given to the question about the aesthetics in the images in Q13, grouped from 1 to 5



Moreover, Figure 14 shows the results for the question about how captivating the set of images was. Looking at the results, it is possible to see that the majority of the participants think that the images are captivating. This leads us to hypothesise that these visual artefacts would also have value if used in an outdoor, publicising the retail company, as they would captivate the passersby to look and analyse the outdoor.

Finally, in Q15, we asked the users if they would use the images as their avatar, for example in a social network or website. The majority of the users (61.9%) said that they would use the images as an avatar, whilst (38.1%) say that they would not. Note that the users were not given the information that the images were generated using data about consumption's in retail. If they were allowed to use their data, and generate an image that reflected their own consumption habits, we are convinced that the number of no answers would decrease substantially, as the artefacts would gain more personal value.

Figure 13 Distribution (in percentage) of the answers given to the question about the diversity of the images in Q12, grouped from 1 to 5

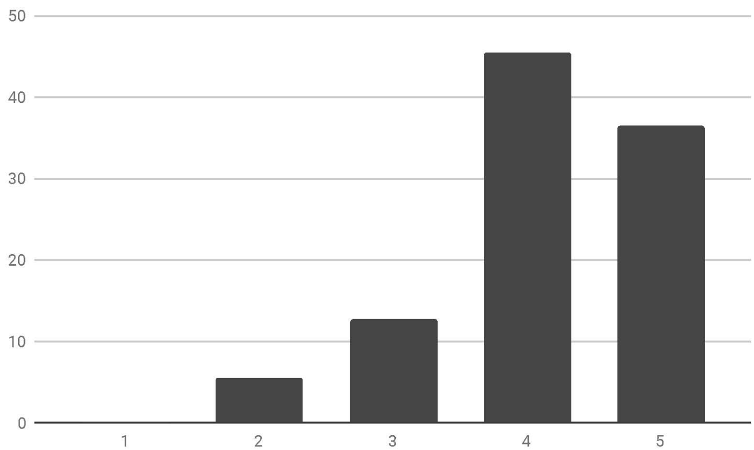
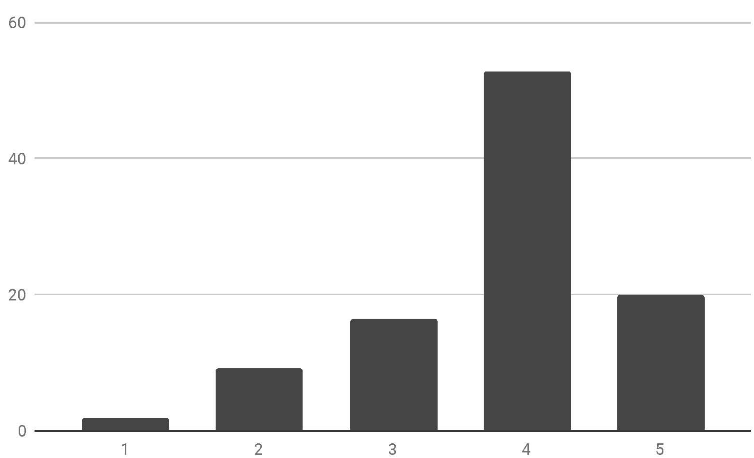


Figure 14 Distribution (in percentage) of the answers given to the question about how captivating the images in Q14 are, grouped from 1 to 5



6 Conclusions

Over the last few years, we have seen an exponential increase in data in all sectors of business. Companies have seen this as an opportunity to improve their operations and increase the satisfaction of their costumers. In previous works, we have proposed, developed, and implemented visualisation models for one of the biggest retail companies in Portugal. We have developed a set of artefacts that refined the way the company looked at their data in order to improve their business strategy. In this article, we extend a previous work (Maçãs et al., 2015a) and explore an automatic manipulation of a swarm system through the application of an evolutionary algorithm. Our main goal was to explore the aesthetic dimension of the consumption data provided by the Portuguese retail company. We validated the system with three different methods:

1. the use of an automatic fitness function, to perceive if the evolutionary algorithm was evolving the artefacts depending on the fitness function
2. the use of three different usage scenarios, to understand how to interact with the system
3. a user study, to validate the system and its capacity to evolve diverse and visually appealing artefacts.

As future work, we intend to improve the interactive evolutionary system by enabling it to learn the specific parameters which are more relevant to the user tastes and guide the evolution based on those parameters. With this, we can also augment the number of individuals per generation and show only the fittest (based on the previous choices) for the user to choose, diminishing user fatigue.

Acknowledgements

The work is supported by Portuguese National Funds through the FCT-Foundation for Science and Technology, I.P., under the grant SFRH/BD/129481/2017.

References

Bornhofen, S., Gardeux, V. and Machizaud, A. (2012) 'From swarm art toward ecosystem art', *International Journal of Swarm Intelligence Research*, Vol. 3, No. 3, pp.1–18.

Dorling, D. (1996) 'Area cartograms: their use and creation', *Concepts and Techniques in Modern Geography series*, Environmental Publications, University of East Anglia.

Eiben, A.E. and Smith, J.E. (2015) *Introduction to Evolutionary Computing*, Natural Computing Series, Springer, Berlin, Heidelberg.

Friendly, M. (2008) 'A brief history of data visualization', in *Handbook of Data Visualization*, pp.15–56, Springer, Berlin, Heidelberg.

Greenfield, G. and Machado, P. (2014) 'Swarm art', *Leonardo*, Vol. 47, No. 1, pp.5–7.

Jones, D. (2007) *Swarm Aesthetics: A Critical Appraisal of Swarming Structures in Art Practice*, Master's thesis, MA Sonic Arts.

Kamvar, S.D. and Harris, J. (2011) 'We feel fine and searching the emotional web', in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, ACM, pp.117–126.

Keim, D., Qu, H. and Ma, K-L. (2013) 'Big-data visualization', *IEEE Computer Graphics and Applications*, Vol. 33, No. 4, pp.20–21.

Maçãs, C. and Machado, P. (2016) 'The rhythm of consumptions', in *Expressive 2016 – Proceedings of the Joint Symposium Expressive 2016, Co-located with the Eurographics 2016*, ACM.

Maçãs, C., Cruz, P., Martins, P. and Machado, P. (2015a) 'Swarm systems in the visualization of consumption patterns', in Yang, Q. and Wooldridge, M. (Eds.): *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015*, AAAI Press, Buenos Aires, Argentina, 25–31 July, pp.2466–2472.

Maçãs, C., Cruz, P., Amaro, H., Polisciuc, E., Carvalho, T., Santos, F. and Machado, P. (2015b) 'Time-series application on big data visualization of consumption in supermarkets', in Braz, J., Kerren, A. and Linsen, L. (Eds.): *IVAPP 2015 Proceedings of the 6th International Conference on Information Visualization Theory and Applications*, SciTePress, Berlin, Germany, 11–14 March, pp.239–246.

Maçãs, C., Cruz, P., Polisciuc, E., Amaro, H. and Machado, P. (2016) 'ISO-edges for the geovisualization of consumptions', in Braz, J., Kerren, A. and Linsen, L. (Eds.): *IVAPP 2016 – Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, SciTePress, Vol. 2, pp.220–227.

Manovich, L. (2000) *Information and Form* [online] http://manovich.net/index.php/projects/information-and-form (accessed 12 September 2018).

Moere, A.V. and Lau, A. (2007) 'In-formation flocking: an approach to data visualization using multi-agent formation behavior', in *Australian Conference on Artificial Life*, Springer, pp.292–304.

Moere, A.V. (2007) 'Aesthetic data visualization as a resource for educating creative design', in *Computer-Aided Architectural Design Futures (CAADFutures) 2007*, Springer, pp.71–84.

Ogawa, M. and Ma, K-L. (2009) 'code_swarm: a design study in organic software visualization', *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, No. 6, pp.1097–1104.

Polisciuc, E., Cruz, P., Amaro, H., Maçãs, C., Carvalho, T., Santos, F. and Machado, P. (2015) 'Arc and swarm-based representations of customer's flows among supermarkets', in Braz, J., Kerren, A. and Linsen, L. (Eds.): *IVAPP 2015 – Proceedings of the 6th International Conference on Information Visualization Theory and Applications*, SciTePress, Berlin, Germany, 11–14 March, pp.300–306.

Polisciuc, E., Maçãs, C., Assunção, F. and Machado, P. (2016a) 'Hexagonal gridded maps and information layers: a novel approach for the exploration and analysis of retail data', in *SIGGRAPH ASIA 2016 Symposium on Visualization*, ACM, New York, NY, USA, pp.6:1–6:8.

Polisciuc, E., Cruz, P., Maçãs, C., Amaro, H. and Machado, P. (2016b) 'Flow map of products transported among warehouses and supermarkets', in Braz, J., Kerren, A. and Linsen, L. (Eds.): *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, SciTePress, Vol. 2, pp.177–186.

Reynolds, C.W. (1987) 'Flocks, herds and schools: a distributed behavioral model', in *ACM SIGGRAPH Computer Graphics*, ACM, Vol. 21, pp.25–34.

Takagi, H. (2001) 'Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation', *Proceedings of the IEEE*, Vol. 89, No. 9, pp.1275–1296.

Tufte, E.R. (1991) 'Envisioning information', *Optometry & Vision Science*, Vol. 68, No. 4, pp.322–324.