A Concept of Visual Knowledge Representation

Tatiana Jaworska

Abstract. The image semantic representation is a very challenging task. This article presents a concept of using visual analysis to represent knowledge based on large amounts of massive, dynamic, ambiguous multimedia. This concept is based on the semantic representation of these visual resources. We argue that the most important factor in building a semantic representation is defining the ordered and hierarchical structure, as well as the relationships among entities. This concept has stemmed from the content-based image retrieval analysis.

Keywords: Image semantics · knowledge representation · order · similarity

1 Introduction

For many years researchers have been intensively striving to describe image semantics. It is an element of a widely understood knowledge representation for further knowledge retrieval.

So far, all knowledge has been represented in language form, in the beginning artificial, and now more or less natural which, at the same time, is the biggest obstacle in the proliferation of the knowledge repository. The best example is Wikipedia, without detracting from its merit, where articles differ depending on the national versions.

Fig. 1 represents the location of visual knowledge representation in the whole pyramid of decision making support. So far, we have developed the data and information retrieval systems. It means that the decision maker has received raw, or slightly processed, mainly aggregate data. Recently, content-based image retrieval systems have caused a great breakthrough in information analysis, becoming the front-end element in the domain of knowledge retrieval systems [1].

With a deluge of images and photos, and the development of graphical interfaces in computers, mobiles, etc., the new generation is more and more dependent on

T. Jaworska(✉)
Systems Research Institute, Polish Academy of Sciences,
01-447, 6 Newelska Street, Warsaw, Poland
e-mail: Tatiana.Jaworska@ibspan.waw.pl

visual information rather than textual. Producers and programmers steadily multiply icons, emoticons and other graphical symbols of all kinds. It concerns not only human-machine interaction systems but, first of all, pattern recognition and machine learning, as well as artificial intelligence. All this suggests that we should construct a visual knowledge representation system, rather than textual ones, e.g. domain ontologies. Our objective is the creation of a visual knowledge representation as the first step to a visual knowledge retrieval system because effective retrieval is possible only when a proper representation has been prepared.

We are aware of the fact that we cannot totally avoid description in knowledge representation, but in this paper we present a concept of knowledge representation, focused on images as much as possible. We will demonstrate that images and, broadly understood, multimedia have such a large information potential that we can reduce the use of a natural language to nearly zero and, thanks to this, make our system much more universal.

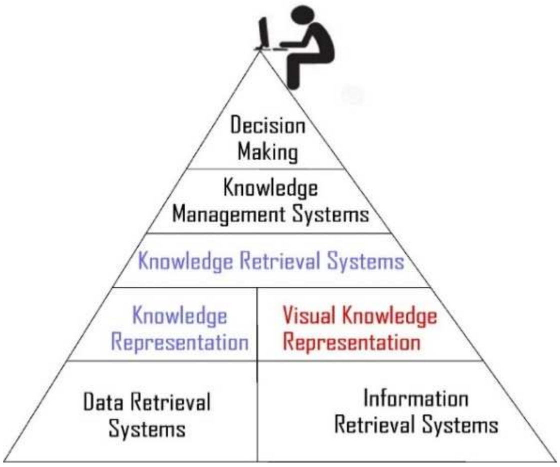
Decision Making
Knowledge Management Systems
Knowledge Retrieval Systems
Knowledge Representation
Visual Knowledge Representation
Data Retrieval Systems
Information Retrieval Systems

Fig. 1 The decision making support pyramid.

1.1 Categorical Perception

We claim that the most important observation is the way in which the human brain perceives images and what conclusions it draws from visual information. Let us think for a moment about the task of identifying an animal – a potentially dangerous one. First of all, we recognize the general posture. If we know this kind of animal because of our earlier experience, we can recognize it based on the fragment of its silhouette, for example a head. Then we decide whether to escape because of the threat the animal poses, or to catch it because it might promise a tasty meal. This knowledge is deduced without the name of the animal [2].

In this example two aspects are important: (i) assignation to category: danger / safe and edible / inedible; (ii) the kind of categories which are orthogonal to each other.

Thus categorization is fundamental in prediction, inference, classification, and further, in knowledge representation and decision making. Categorical perception is the phenomenon of perception of distinct categories when there is a gradual change in a variable along a continuum. Our perception is based on the different aspects of reality, whereas the understanding of images is inextricably connected with the human experience and the reality we live in. That is why we started preparing the framework of our system from defining the most universal and mutually separable areas of abstraction in order to cover the widest possible range of semantics. Moreover, to organize the knowledge in each area of abstraction, we had to define the way of ordering according to which we would arrange images to enable the user to navigate the system.

The mathematical framework for human perception problems, involving ambiguous image perception, was provided by Tim Poston in 1978 in [3]. These problem shall be ignored here, as they reach beyond the scope of our interests.

2 The Concept of Image Wiki

Visual knowledge representation and, later, visual knowledge retrieval systems will offer quite new capabilities to the decision makers. A unified measure reflecting visual semantic similarity has not been developed yet, although we claim that semantic information can be described in context. In each context each image can be treated as an information granule, classified by a vector of categories. We can go even further: each object in an image can also be an information granule [4]. For each seed of information which constitutes a visual object, we define a coverage and specificity for which the trade-off has to maintain. Information granules of type-2 give most abstractive notion, e.g. a set of many people and buildings defines a city. Information granules characterise with/by:

- The semantic similarity which means that feature vectors for them should be more or less similar;
- The functional similarity understand as similarity of classes assigned to similar visual objects;
- The semantic unambiguity - communication impact is defined by semantic classes;
- Representation of objects from the real world.

Information granularity connects strictly/inseparably with the notion of scale.

Thinking globally about all the sets of existing images, we can tentatively select the following most common areas of abstraction which we understand as orthogonal dimensions, by which we can characterize each image:

- Scale[2] – an image can present information about objects of different size: from galactic clusters to atoms. The scale should change linearly or at intervals, depending on a structure of data.
- Time – changes linearly in our system, according to our common intuition. However, in the system there is information about a photo acquisition date, but this same photo appears in a domain chronology, which means that if a photo presents a geological mesozoic formation, it is presented in a geological (time?) chronology.
- Hierarchy – our algorithm organizes images in a general-to-specific relationship of content (depth of abstraction) which is deliberately less strict in comparison with formal linguistic relations.
- Content domain – taxonomy – covers different areas of abstraction. It is the most voluminous because it has to address as many domains as humans are interested in. It shall connect to existing ontologies [5].
- Geographical location – connects the location where a photo has been taken with GIS or Google maps.
- Image author – information about the image author, can link to the author's website or images of the author's masterpieces.
- Action – presents some actions and movies in videos.
- Information granularity – images in their nature contains information in the form of visual information granules. Semantic objects/segments are the best candidates for such granules, with bounding box as coverage and centroid as location. Objects have assigned a vector of classes or prototypes;
- Exemplification – at the level of abstraction where the detailed exemplification is possible. At the lowest level images will be organized according to some similarity measures to make browsing among them easier.

As we can see from the above list, each of these dimensions represents an important aspect of each image and has to be organized by means of a different, immanent order (as mentioned in the introduction). All the above-listed parameters will be elaborated on below, beginning from the basic notions

The assumptions in our project are very wide, however, the framework of the system is still under construction. In order to create the system, obviously gradually, we use the existing CBIR systems [6], [7] in different domains to obtain a sufficient number of images and order them based on their similarity in different metrics. The help of experts is needed in the proper organization of domain knowledge. The latest achievements, namely CNNs [8], [9], allow a quick selection of objects in an image in order to add to each selected object some more detailed images at the lower levels. The object images which will be attached on a lower level to the more general image shall be accepted by experts if the images are relevant in terms of the taxonomy of the particular branch of domain knowledge. At each level of browsing, the user will

[2] There are two notions called *scale*. A s*cale* here means the size of the object represented by an image. In subsec. 2.2 we use the notion *scale,* in fact, as a **scale of measure** which is different from the one described here.

be able to jump to another aspect. For example, the user finds a satellite image of a fragment of their town and wants to see the changes in urban development in the last 30 years. Then they switch to the time search and can see the photos and archive plans of this selected area.

Obviously, we assume the availability of these data. The system shall offer free access to any user willing to edit the content similarly to the rules in the textual Wikipedia, thanks to which the amount of images will increase very fast and will be revised by experts to maintain semantic correction.

Each new added image contains all the information required to be localized in a proper place in the whole structure. At each level, many exemplifications will be accessible and each object in them will be connected to the lower level images in a domain hierarchy.

The system also has to contain drawings, schemes, slides and other graphical materials in order to help the user to understand the presented knowledge. In many cases we understand an image not as a photo, but as a specially prepared photographic or animated illustration to visualize how a scheme or process works. We would like to emphasize that the system in its assumption is much wider that GIS, CBIR systems and ontologies, but connects them together and, because of this fact, it partially uses the mechanism and algorithms implemented in these systems.

2.1 Preliminaries

In order to unify our concept, we have applied the notion of a total order and a partial order, according to which we build relations over a set of visual entities X. By an entity of the system we understand any visual information granules, such as: images, videos, 3D graphics, depth maps, etc.

A (non-strict) partial order $P = (X, \leq)$ is a binary relation $\leq$ over a set X, satisfying particular axioms. The axioms for a non-strict partial order state that the relation $\leq$ is reflexive, antisymmetric, and transitive [10]. That is, for all a, b, and c in X, it must satisfy:

- $a \leq a$ (reflexivity: every element is related to itself).
- if $a \leq b$ and $b \leq a$, then $a = b$ (antisymmetry: two distinct elements cannot be related in both directions).
- if $a \leq b$ and $b \leq c$, then $a \leq c$ (transitivity: if a first element is related to a second element, and, in turn, that element is related to a third element, then the first element is related to the third element).

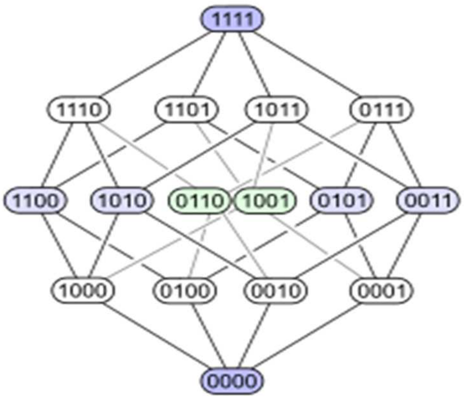

Fig. 2 An example of a Hasse diagram for a 4-element set ordered by inclusion $\subseteq$. (Follow Wikipedia: https://en.wikipedia.org/wiki/Hasse_diagram).

In other words, a partial order is an antisymmetric pre-order. A set with a partial order is called a partially ordered set (also called a *poset*). A well-known example of a partial order is a linear order, particularly one based on real numbers.

Graphically, a poset can be represented in the form of a Hasse diagram (see Fig. 2) [11] where each element of *X* is a vertex and the upward line symbolizes the binary relation which holds between comparable elements that are immediate neighbours. Such a diagram, with labelled vertices, uniquely determines its partial order.

In the most common sense of the term, a graph is an ordered pair *G* = (*X*, *E*) comprising a set *X* of vertices, nodes or points together with a set *E* of edges, arcs or lines, which are 2-element subsets of *X* (i.e. an edge is associated with two vertices, and that association takes the form of an unordered pair comprising those two vertices) [12].

We have introduced all these notions in order to be able to organize images semantically. Some aspects, such as time or scale, are linear, so the linear order is used naturally, but others, such as hierarchy or taxonomy, can be defined only in a different way [13].

Then, we have to use a different scale of measure which describes the multidimensional nature of information. Here, we follow S.S. Stevens [14], being aware that there are some other approaches to the problem of scale and scaling. In our case the division of scales into four types: nominal, ordinal, intervallic, and linear is appropriate for the dimensions we have proposed.

The nominal scale is apparently the simplest one, but it is formally described by the category theory, from the mathematical point of view, and it is basic to all kinds of classification, which is one of the most important parameters when we discuss semantic image analysis, especially when the value of the quality of the information is difficult to assign. As W.W. Rozeboom *[15]* claims, a semantic scale is equivalent to a formal scale. In such a situation, classes or categories and properties are coded by the values of a natural variable, which are equivalent to a formal scale. Strictly speaking, a scale-name A and set $\mathbf{a}$ of symbols compose a semantic scale $\langle A,\mathbf{a}\rangle$ where $\mathbf{a}\in\{x_1,\ldots, x_n\}$ for natural variables $\mathbf{\Delta}\in\{d_1,\ldots, d_n\}$ in a given coding system. Then, we can always find a formal scale ϕ of $\mathbf{\Delta}$ $\langle\phi,\mathbf{\Delta}\rangle$, such that A and ϕ have the same argument domain $\mathbf{D} = 1,\ldots,n$ and there exists a function f from $\mathbf{a}$ into $\mathbf{\Delta}$, such that: $f:\mathbf{a}\rightarrow\mathbf{\Delta}$, $d = f(x)$. Then, f is called a scaling transformation of $\mathbf{a}$ into $\mathbf{\Delta}$ and it must be one-one.

This scheme has to be extended in the image case because image information is much more complex. Semantic variables are represented by a vector of attributes for each of the above-mentioned dimensions.

The linear and ordinal scales are based on an order described above.

The interval scale will be used to present information for which it is impossible to assign an exact value of a particular dimension, such as geological eras, for instance, for images of geological layers.

2.2 How to Build such a System?

Data structure should be similar to a Geographic Information System (GIS) data structure for raster or grid data, with the difference in the attribute representation. As we mentioned in the preliminaries, a Hasse diagram for a poset $P = (X, \leq)$ represents an acyclic graph $G = (X, \prec)$, such that maximal elements are situated at the top of the diagram, and for two vertices $a \prec b$, where a vertex a is over b. As we know, an acyclic graph represents tree as a data structure. Hence, as a basic data structure, we use a tree and because of the fact that images have rectangular structure and objects selected from them are polygons we apply the R-tree structure. There are no pure R-tree structures because there is an option of changing aspects which forces the correlations between R-tree structures [16]. The layer structure, being a characteristic feature of a GIS, in our case, is equivalent to levels in trees. We use Oracle DBMS for a set of all images, data, attributes, etc.

It is also important that the user has an access to an image in most possible semantic correlations. Hence, once more we relate to the image of a geological mesozoic formation. Not only will this image appear in domain geologic knowledge/context, but also the user will see it analyzing the works of the geologist who took this photo, as well as looking through mineralogical domain.

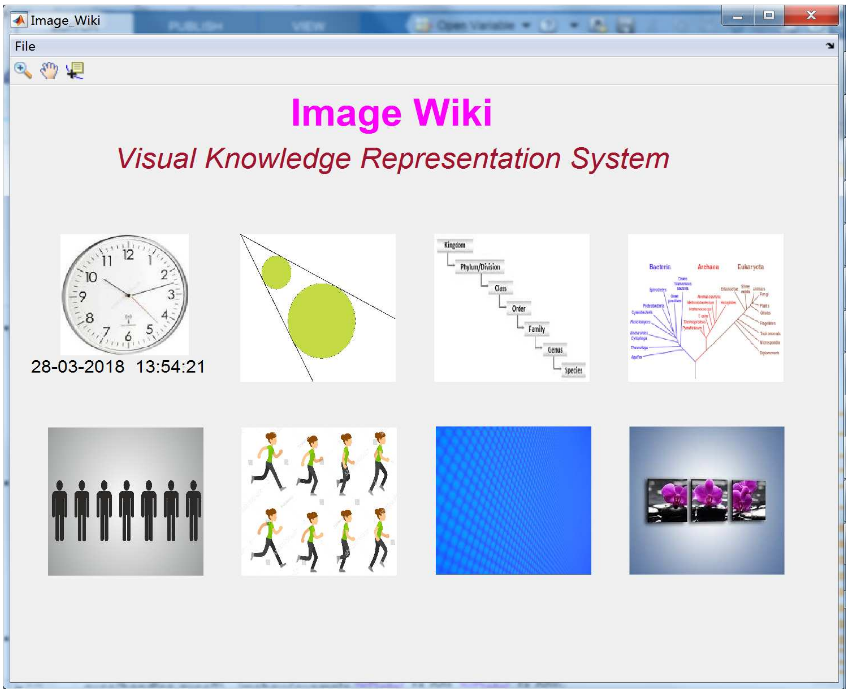

Fig. 3 The main screen of the Image Wiki system. Each square is an active icon.

3 The System Navigation and the User Interface

Above, in Fig. 3, we present the main screen of our system where icons symbolizing the main search parameters/aspects (from the top left time down to the bottom right information granularity and sequence of the example of the same kind of entity). Each of them opens further windows, enabling the user to surf down a particular dimension to find the image he/she wants to learn something. It is not an attempt to present similar images, it is a system which organizes images in a semantic order which depends on the aspect the user is browsing at a particular moment.

For example, the user is in an aspect *hierarchy* and then they go through the anatomy of an animal (see Fig. 4). When the user wants to look at different photos of a fruit fly (drosophila), they can switch into an enumeration, then look at other photographs of the same species (a fruit fly in this particular example). Browsing examples of objects at the same level is available after selecting and by clicking on the left/right grey arrow situated on both sides of the screen (see Fig. 4). But when the user wants to continue analysing the fruit fly anatomy details she\he clicks on the eyes, legs, or wings and moves lower to images from a microscope, presenting in detail particular organs.

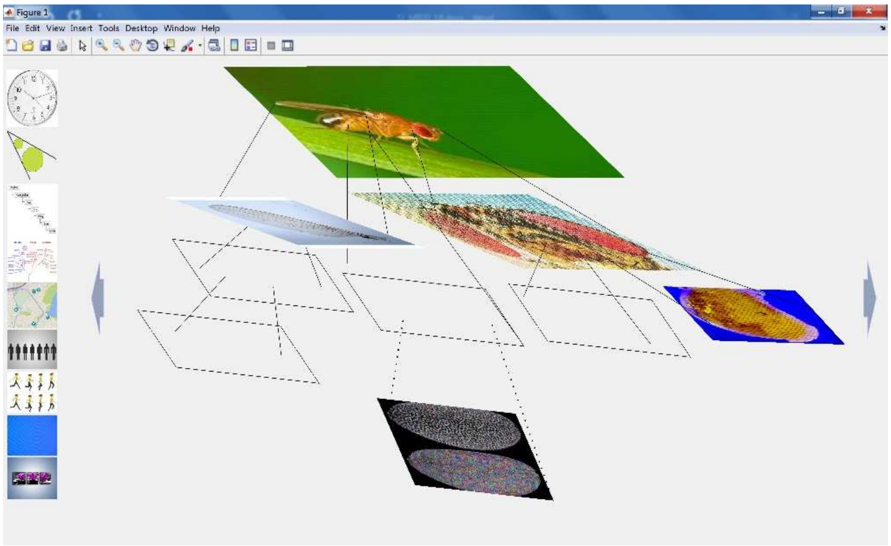

Fig. 4 The manner of understanding the mechanism of visual organization and presentation of taxonomy. All component images CC.

A fast search for interesting information by the user will be available through a system of icons which will facilitate quick navigation across the most general levels. The system will start from the previous browsing stage to save the time of going down to a particular level and it will remember the search history to aid returning to the previous level of search.

In order to change a domain, the user clicks the icon symbolizing taxonomy and selects the proper domain of interest. The system is based on different kinds of images, which means not only photographs, but also drawings, schemes, sketches, videos, etc. The domains can overlap. At each level, there is a list of tools to switch into another aspect.

One of the domains contains the zoom-in maps, similar to the Google map. The user can move to street view and later to a particular building. At present, there are photographs of buildings, and through satellite stereo images their 3D models are calculated more often. When 3D models of different building interiors are available, we will be able to incorporate them into our system in order to enable the user to visit these buildings inside.

Navigation at a particular level will be intuitive or rather similar to the navigation of present graphical programs, where the scroll wheel zooms in and out the image, a click on an object with the left button presents this object in more detail on a more precise level, a click with the right button anywhere in the image moves the user to the upper level.

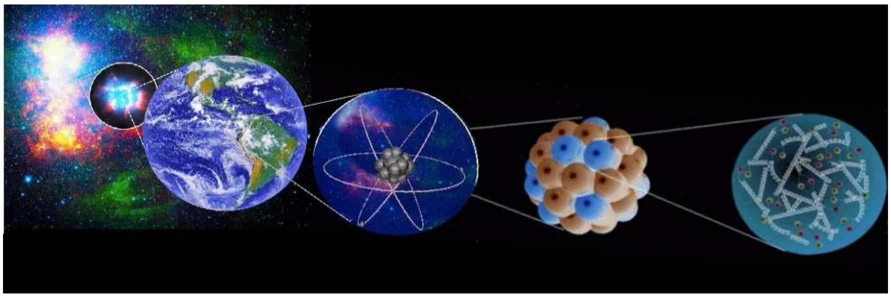

Fig. 5 An example of space distances of image content, enabling the user to select the scale of interest, from light years for astronomical images, through kilometers and meters related to the Earth, down to atom diameters and even further, to the subatomic scale.

All visual objects in an image are active. For example, Fig. 5 presents a quick selection of the scale in which the user wants to operate. It means that a click on a left nebula moves the user to astronomical knowledge, a click on the Earth moves them to the Earth map and in the next step at once to the level of the Google map, a click on the atom moves to chemistry and the atomic scale, etc.

4 Conclusions and Further Works

In this paper, we present the framework which is being created in order to elaborate the visual knowledge representation for further visual knowledge retrieval. Taking into consideration the state-of-the-art semantic analysis methods, we can safely assert that the construction of the above-described system is fully technically feasible, though time consuming. However, the fact that the semantic analysis of image by machines is still far from human abilities remains a considerable challenge. Additionally, effective management of such huge sets of visual information will require intensive research in terms of both hardware resources and new algorithms for sharing and managing visual information.

References

[1] S. Belongie and P. Perona, "Visipedia circa 2015," *Pattern Recognition Letters,* no. 72 , pp. 15-24, 1 Mar. 2016. doi: 10.1016/j.patrec.2015.11.023

[2] J. Hawking and S. Blakeslee, On intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines, NY,: Henry Holt and Company, 2004, p. 262.

[3] T. Poston and I. Stewart, "Nonlinear Modeling of Multistable Perception," *Systems Research and Behavioral Science,* vol. 4, no. 23, pp. 318-334, 1978. doi: 10.1002/bs.3830230403

[4] F. Yu and W. Pedrycz, "The design of fuzzy information granules: Tradeoffs between specificity and experimental evidence," *Applied Soft Computing,* vol. 9, p. 264–273, 2009. doi: 0.1016/j.asoc.2007.10.026

[5] P. Chmiel, M. Ganzha, T. Jaworska and M. Paprzycki, “Combining semantic technologies with a content-based image retrieval system – Preliminary considerations,” in *Application of Mathematics in Technical and Natural Sciences, Conference Proceedings*, Albena, Bulgaria, Oct. 2017. doi: 10.1063/1.5007405

[6] T. Jaworska, “The Concept of a Multi-Step Search-Engine for the Content-Based Image Retrieval Systems,” in *Information Systems Architecture and Technology. Web Information Systems Engineering, Knowledge Discovery and Hybrid Computing*, Wrocław, 2011.

[7] T. Jaworska, “A Search-Engine Concept Based on Multi-Feature Vectors and Spatial Relationship,” in *Flexible Query Answering Systems*, vol. 7022, H. Christiansen, G. De Tré, A. Yazici, S. Zadrożny and H. L. Larsen, Eds., Ghent, Springer, 2011, pp. 137-148. doi: 10.1007/978-3-642-24764-4_13

[8] A. Krizhevsky, I. Sutskever and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, Toronto, 2012, pp. 1-9.

[9] N. Zhang, J. Donahue, R. Girshick and T. Darrell, “Part-based R-CNNs for Fine-grained Category Detection,” in *13th European Conference on Computer Vision - ECCV. Proceedings, Part I*, Zurich, Switzerland, 6-12 Sep, 2014. doi: 10.1007/978-3-319-10590-1_54

[10] D. A. Simovici i C. Djeraba, „Partially Ordered Sets,” w *Mathematical Tools for Data Mining: Set Theory, Partial Orders, Combinatorics*, London, Springer Science & Business Media, 2008, p. 615. doi: 10.1007/978-1-4471-6407-4

[11] R. Freese, “Automated lattice drawing,” in *Second International Conference on Formal Concept Analysis*, Sydney, Australia, Feb. 23-26, 2004. doi: 10.1007/978-3-540-24651-0_12

[12] J. Bang-Jensen and G. Z. Gutin, “Basic Terminology, Notation and Results,” in *Digraphs Theory, Algorithms and Applications*, London, Springer, 2009, pp. 1-11. doi: 10.1007/978-1-84800-998-1

[13] E. Bart, I. Porteous, P. Perona and M. Welling, “Unsupervised Learning of Visual Taxonomies,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, USA, 23-28 June, 2008.

[14] S. S. Stevens, “On the Theory of Scales of Measurement,” *Science*, vol. 103, no. 2684, pp. 677-680, 7 june 1946. doi: 10.1126/science.103.2684.677

[15] W. W. Rozeboom, “Scaling Theory and the Nature of Measurement,” *Synthese*, vol. 16, no. 2, pp. 170-233, Nov. 1966. doi: 10.1007/BF00485356

[16] Y. Manolopoulos, A. Nanopoulos, A. N. Papadopoulos and Y. Theodoridis, R-Trees: Theory and Applications, III ed., London: Springer Science & Business Media, 2010, p. 194. doi: 10.1007/978-1-84628-293-5