Sub-project 4: „Experimental development of augmented reality tools for online social networks and the study of their impact on users" (AR Media)

REPORT ON PHASES 4-3, 4-4 și 4.5

PHASE 3 2020

Phase 4-3: Development and implementation of options and services (subsystems) for online social network using augmented reality AR Media

Activity 4-3-3: *Developing the object recognition service using cameras and displaying various information about them, such as translating the objects' names into foreign languages according to the geographical location of AR Media user*

STUDY 1: Objects identification using cameras (object recognition) and displaying various information about them, such as translating object names into foreign languages depending on the geographical location of AR Media user

During this activity, the research team studied the current state of knowledge in the field of Computer Vision and tried to find the best methods for creating the proposed application. Computer Vision is an interdisciplinary field that deals with how a computational platform can be built and used to obtain an analysis of information extracted from digital images or videos. From an engineering perspective, Computer Vision automates the tasks that the human visual system can perform. (1)

The tasks included in this category are:

- procurement methods,
- processing methods,
- methods of analysis and understanding of real-world digital images for producing information

The classic problem is image processing and the ability of algorithms to determine whether or not the image data contains a specific object, feature or activity.

Different types of recognition problem are described in the literature: object recognition - one or more preferred or learned objects or object classes can be recognized, usually together with their 2D positions in the image or 3D position in the scene. Blippar, Google Goggles and LikeThat offer stand-alone programs that illustrate this functionality; identification - an individual instance of an object is recognized (examples: identification of a person's face or fingerprint, identification of handwritten numbers or identification of a specific vehicle); detection - image data is scanned for a specific condition (examples include detecting possible abnormal cells or tissues in medical images or detecting a vehicle in an automated road taxing system; detection based on relatively simple and fast calculations is sometimes used to find regions of lesser interest in image data, which can be further analyzed by more demanding computerized techniques to produce a correct interpretation). (1)

In addition to classic Computer Vision problems, such as differences between classes or different angles from which the object can be photographed, the location of objects must consider the rotation and scaling of the object (small objects), precise location of the object, detection speed etc.

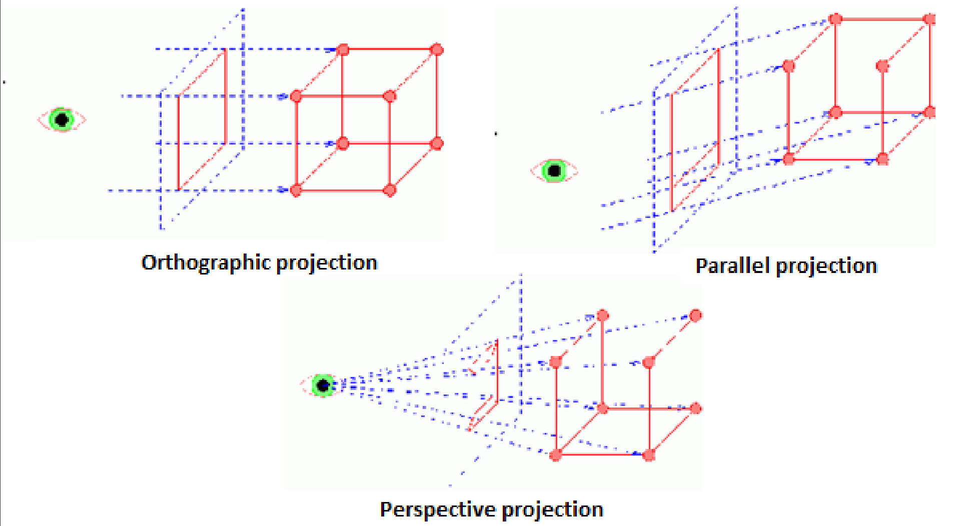
The most commonly used model for image formation geometry is that of the darkroom, where the image is formed by light rays passing through an infinitesimal orifice and projecting them onto a plane. In a traditional device, this plane is a photosensitive material; Digital cameras have the plan composed of electronic elements sensitive to radiation in the spectrum of interest (visible, infrared etc.). (2)

The geometric transformation in the darkroom model is intuitive and easy to express mathematically - the technology is *perspective projection*. The big problem with this transformation, however, is that the relationship between the points coordinates inside the image and the tridimensional coordinates of the photographed world is inherently nonlinear.

Because of this, another type of projection is often used, the *orthographic projection*, where an object in 3D space is first projected perpendicular to the image's plane, ignoring *the depth* (meaning that a point (x, y, z) in 3 dimensions becomes a point (x, y) in the image plane). Then this projection is reduced in scale, depending on the average distance of the object from the camera. (2)

This model, which uses an orthographic projection followed by a scaling down, is an acceptable approximation of the model in perspective, only in cases where the object is close to the optical axis of the camera and relatively small, compared to the distance between the object and the optical device. (2)

Another linear transformation is the *parallel projection*. In this case, too, the object is projected on the image's plane and then reduced to scale, but, this time, the projection is made by parallel rays oriented in the "average" direction of the rays that would make up the projection through perspective. This method gives good results even if the object is not close to the optical axis. If the object is right on the optical axis, the parallel and orthographic projections are equivalent. The three projections are illustrated in the following figure (2):



Three types of projection in modelling the cameras which capture the image

Main steps in creating an object recognition application (2):

- Image capturing
- Images comparison
- Creating the multimedia database
- Object recognition
- Segmentation
- Stereoscopy
- Learning

Image capturing - An image is a function, defined by two dimensions whose value is the color (or light intensity for black and white images). Because each photosensitive element has finite dimensions, in the process of image capturing the domain is discretized in a grid; each cell in the grid corresponds to a photosensitive element (pixel). This transformation from continuous to discrete space is called sampling. The value of color or light intensity is also in a continuous space; the devices will discern a finite number of different values. (2)

Images comparison - A simple operation is the comparison of 2 images: we search for images on the Internet or inside digital libraries with object recognition (by comparing with a prototype image). Algebra provides us with very simple tools for solving very effectively at least part of the comparison problem. For each image we "measure" in a certain way a series of parameters that seem important to us for classification. For each such parameter we obtain a value. (2)

Here are some examples of parameters:

- image size in pixels;
- the number of different colors or shades of gray;
- the size of the largest matte "spot" in the image;
- the number of points of a certain color;
- the average number of consecutive points of a certain color in a given direction;
- the average of the colors inside the image;
- maximum and minimum intensity of points inside the image.

We assemble for each image this collection of values in a vector. If we have n different values for an image, we get a vector in an n-dimensional space. In the extreme case, we can see the whole image as a vector in an x * y-dimensional space, where x and y are the dimensions in pixels, and the value at each point inside the image is the color intensity at that point.

Generally, this extreme solution is not practical, for two reasons (2):

1. We want all the images we operate with to be represented by vectors with the same number of dimensions, because this is crucial to compare them.

2. The number of pixels inside an image is usually much too large, compared to the number of features we are interested in. We often want to compress the image description into a relatively small number of features (from a few to a few hundred).

Once we have a description of an image through a vector, we can use two simple metrics to compare the vectors (2):

- We can calculate the distance between two vectors;
- We can calculate the plane angle between two vectors.

The second method is recommended, because it is insensitive to vector scaling (for example, for the parameters exemplified above, the angle between two vectors does not change if we double the size of one of them). Both metrics can be calculated very easily: the first requires n subtractions, and the second n multiplications and additions plus a radical.

Multimedia databases - The distinction between the process of searching for an image by content and the image recognition process is quite fine. Many recognition systems can be used to search for images inside a database. The distinction is that in the case of databases, images are available before starting the search so they can be pre-processed. We can thus create offline data structures with algorithms that are too expensive for interactive object recognition. (2)

Object Recognition - In objects recognition, all start with a set of object models; when an image is presented to the system for analysis, the available models are used to determine the objects from the image. But the problems to be solved are very difficult: to recognize an object in an image, we should know what it looks like seen from any position, and with any light source. In practice this requirement is impossible to meet. If we have a complete model of the object, the recognition algorithms try to determine which is the most plausible position and illumination to `` explain '' the image. (2)

Segmentation - Another fundamental problem is to separate the different objects that appear inside an image. The difficulty of this problem varies depending on the image; the more uniform the illumination, the more contrasting the objects are and less noise the image contains, the simpler the segmentation problem is. (2)

By some method, we mark the “centers” of the main objects from the image. We can do this either with information from the user, or choosing the maximum contrast points, or other information, depending on the application.

We then calculate the “distance” between each of two neighboring points in the image (we can use either four or eight neighbors) as follows: the more different the colors, the greater the distance. We then run an iterative computational process in which each point is assigned in the same region as the nearest neighbor. The algorithm ends when each point is assigned to one of the initial regions. (2)

Stereoscopy - Stereoscopy is the process used for obtaining tridimensional information from two or more images. Human perception is based on this principle. (2)

Learning - Practically any algorithm that determines operating parameters can be assigned to this category. An example is the "training" of an image segmentation system. (2)

Recognition service for object detection - FUTUREWEB project

The recognition service for objects detection using cameras is made in Python using Pytorch and the Facebook library, Detectron for detecting objects. PyTorch is an optimized library for deep learning using GPUs and CPUs. Detectron is a performant code base for object detection, covering the identification and segmentation possibilities for the identified objects. Detectron was built by Facebook AI Research (FAIR) to support the rapid implementation and evaluation of new computer vision research. It includes implementations for the following object detection algorithms: Mask R-CNN, RetinaNet, Faster R-CNN, RPN, Fast R-CNN, R-FCN.

Detectron can be used immediately to detect general objects or modified to train and execute inference on its own data sets. It is written in Python and will be powered by the PyTorch 1.0 deep learning framework.

A very important aspect is the database that contains the "known" application objects. Each object is decomposed into its essential features, and can be recognized from any 3D angle.

The application has an obvious advantage by using Deep Learning, which allows "learning" new objects. Basically, the number of objects in the database can be unlimited and thus the objects recognition degree in any image increases significantly.

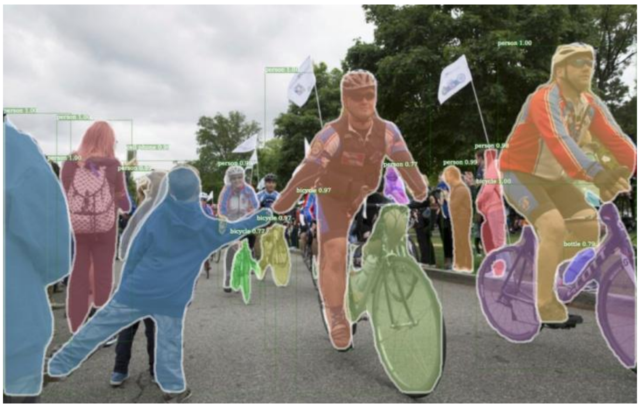


Fig.01 – Example of detecting objects inside a photo using Detectron code

Activity 4-3-4: ***Development of the object tagging service through the augmented reality of real objects from the environment and label dating so this information is available for being viewed by other registered users of the AR Media network***

STUDY2.

Object tagging service through the augmented reality of real objects from the environment and label dating so this information is available for being viewed by other registered users of the AR Media network

The object tagging service through the augmented reality of real objects from the environment is performed for each object that is recognized as existing in the database.

Basically, the following steps are followed:

P1 - Loading the image chosen to be analyzed

P2 - Analyzing the image, identifying the objects inside the image

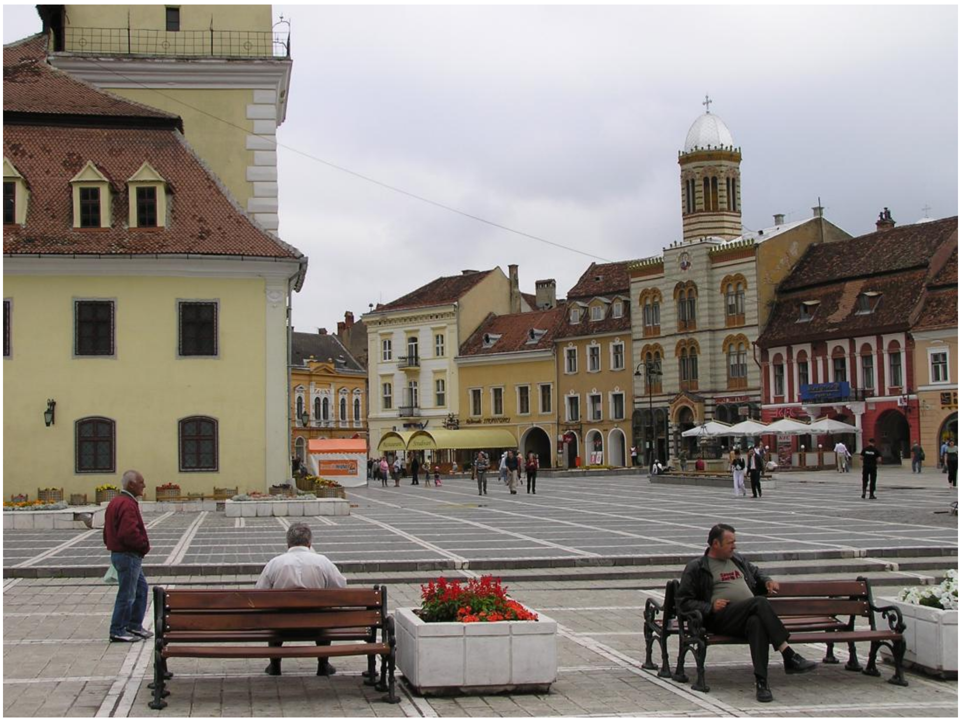
P3 - For each identified object in the image, a similarity comparison is done for the objects in the object database

P4 – Identifying inside the database the object with the maximum similarity index with that of the object in the analyzed image

P5 - An augmented reality label is applied to the identified object, containing information that is considered relevant as a description of the object. (This information is stored in the database next to objects).

F1-1 – Initial image that will be analyzed

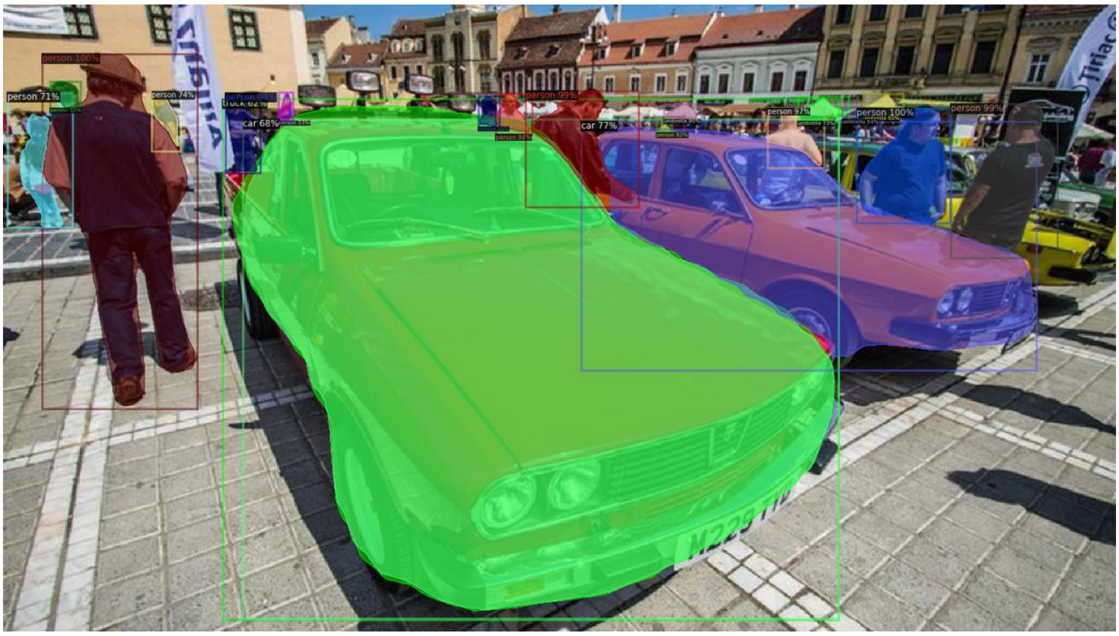
F1-2 – The image containing labels for identified objects (person 98%, person 99%, person100%)



F2-1 - Initial image that will be analyzed

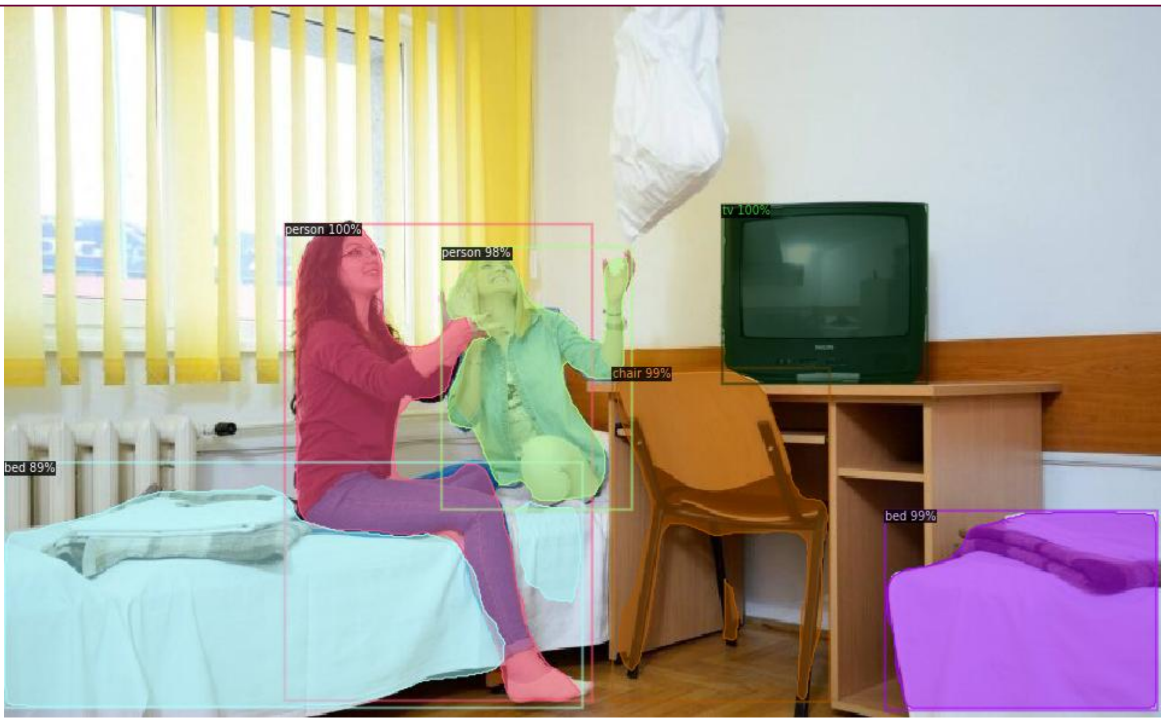
F1-2 - The image containing labels for identified objects (person 100%, bench 100%, person 85%, bench 84%, person 85%...)

F3-1 - Initial image that will be analyzed



F1-2 - The image containing labels for identified objects (person 100%, person 99%, car 77%, car 68%...)

F4-1 - Initial image that will be analyzed



F1-2 - The image containing labels for identified objects

(person 100%, person 98%, chair 99%, TV 100%, bed 99%)

F5-1 - Initial image that will be analyzed

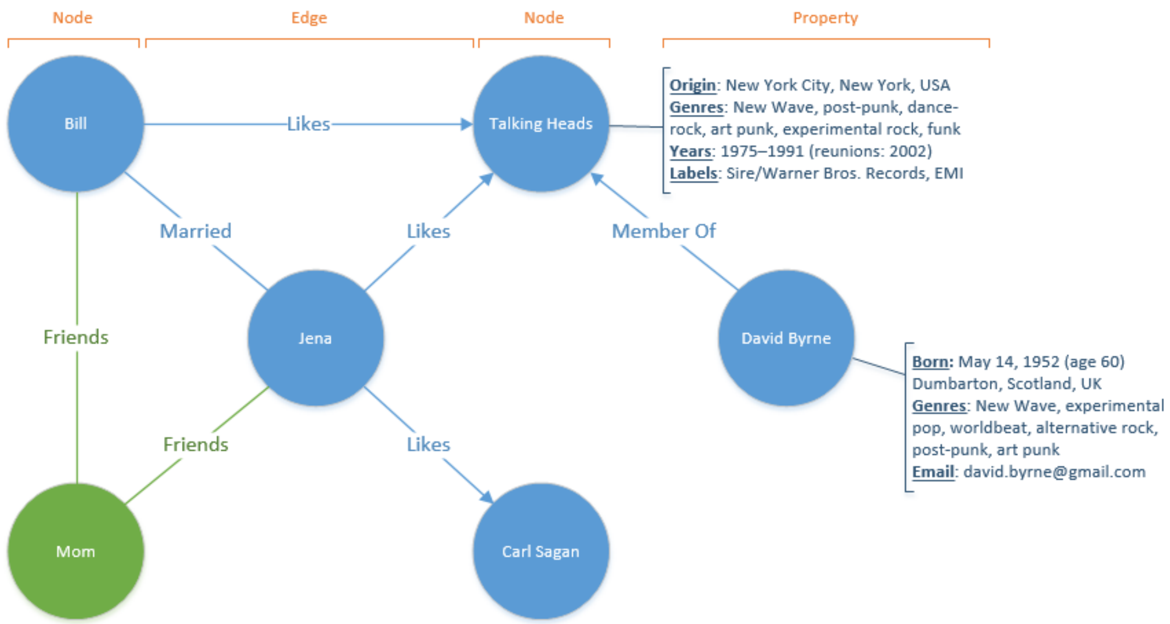
F1-2 - The image containing labels for identified objects (person 100%, person 99%, bike 99%, bike 100% ...)

Activity 4-3-5: ***Developing the service for connecting, retrieving and analyzing the users' activity from other specific social networks such as LinkedIn, Twitter, Flickr, Youtube, Spotify or other networks within AR Media to analyze their recent activity, images and favorite videos.***

STUDY 3. Connecting, retrieving and analyzing the users' activity from other specific social networks such as LinkedIn, Twitter, Flickr, Youtube, Spotify or other networks within AR Media to analyze their recent activity, images and favorite videos.

The service for connecting, retrieving and analyzing the users activity on other social networks is based on APIs specific to each interconnected network. The Graph API is the main way to extract data from the Facebook platform and also to post data on it. It is an HTTP-based API that applications can use to query data, post new articles, manage ads, upload photos, and perform a wide range of other tasks through programs / scripts.

The structure of a Graph API uses nodes to obtain data about individual objects. Marginal elements are used to obtain the collection of objects associated with a node or to publish objects in those collections. Fields are then used to specify that all desired data are included in the responses.



LinkedIn provides a service that allows people to bring their LinkedIn profiles and networks to your site or application through their OAuth-based API.

However, automatic data retrieval must be done taking into account data portability in the GDPR - “the new right to data portability aims to hold data subjects accountable for their own personal data as it facilitates their ability to move, copy or transmit data with personal character easily from one IT environment to another".

Phase 4-4: Content development and promotion of the use of the online social network with augmented reality AR Media and its integration on FutureWeb platform

Activity 4-4-1: ***Improving and optimizing the AR Media online social network through the direct contribution of members of the research consortium and application's users and its integration on FutureWeb platform***

STUDY 4.

Improving and optimizing the AR Media online social network through the direct contribution of members of the research consortium and application's users and its integration on FutureWeb platform

An important aspect of the application developed within the project was its improvement and optimization. For this, an important role was played by the research project team, which tested both the practical functionalities of the social network and the stability and reliability over time.

Specifically, the dependencies on other online applications or databases, the Internet connection, the database load degree etc. were followed.

At the server level, the log lists were analyzed, following the possible problems as well as the causal elements, trying to obtain a platform as stable as possible for the future.

Phase 4-5: Wide dissemination of project results

Activity 4-5-1: Wide dissemination of project results

In 2020, the research team members published 4 articles with FutureWeb project acknowledgment and a brochure presenting the project on the site.

Published articles / to be published articles:

PERSONALIZED ONLINE MARKETING USING FACIAL AND EMOTION RECOGNITION/2020/Journal of Smart Economic Growth/Radu Lixăndroiu, Cătălin Maican, Gheorghe Epuran, Gabriel Brătucu, Lavinia Dovleac (published)

https://jseg.ro/index.php/jseg/article/view/109

AUGMENTED REALITY TECHNOLOGIES IN EDUCATION - A LITERATURE REVIEW / 2020/ Bulletin of the Transilvania University of Brasov/ Daniela Roxana Vuță (to be published)

http://webbut.unitbv.ro/bulletin/Series%20V/Contents_V_1_2020.html

AUGMENTED REALITY AND FACIAL RECOGNITION TECHNOLOGIES. BUILDING BRIDGES BETWEEN THE HOSPITALITY INDUSTRY AND TOURISTS DURING PANDEMIC/ 2020/ Bulletin of the Transilvania University of Brasov/ Ioana-Simona Ivasciuc (to be published)

http://webbut.unitbv.ro/bulletin/Series%20V/Contents_V_1_2020.html

OPPORTUNITIES OF USING NEW TECHNOLOGIES (VR/AR) IN ORDER TO FACILITATE THE ACCESS OF PERSONS WITH DISABILITIES TO TOURISM PRODUCTS/ 2020/ Bulletin of the Transilvania University of Braşov/ Gheorghe Epuran, Alina Simona Tecău, Cristinel Petrișor Constantin, Bianca Tescașiu, Ioana Bianca Chițu (to be published)

http://webbut.unitbv.ro/bulletin/Series%20V/Contents_V_1_2020.html

Promoting brochure

https://futureweb.unitbv.ro/images/brosura_aplicatie.pdf

References:

(1) Samuel Sandu - Modelarea și simularea algoritmului de detecție a obiectelor în Python pentru Raspbery Pi (Modeling and simulation of the object detection algorithm in Python for Raspbery Pi) (https://www.researchgate.net/publication/346003836_Modelarea_si_simularea_algoritmului_de_detectie_a_obiectelor_in_Python_pentru_Raspbery_Pi)

(2) https://www.cs.cmu.edu/~mihaib/articole/vedere/vedere-html.html

(3) https://analyticsindiamag.com/top-8-algorithms-for-object-detection/

(4) https://www.analyticsvidhya.com/blog/2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1/

(5) https://heartbeat.fritz.ai/introduction-to-basic-object-detection-algorithms-b77295a95a63

(6) https://machinelearningknowledge.ai/different-types-of-object-detection-algorithms/

(7) https://cv-tricks.com/object-detection/faster-r-cnn-yolo-ssd/

(8) https://blog.netcetera.com/object-detection-and-tracking-in-2020-f10fb6ff9af3