

The Emotion Detection Tools in the Museum Education EmoDeM Project

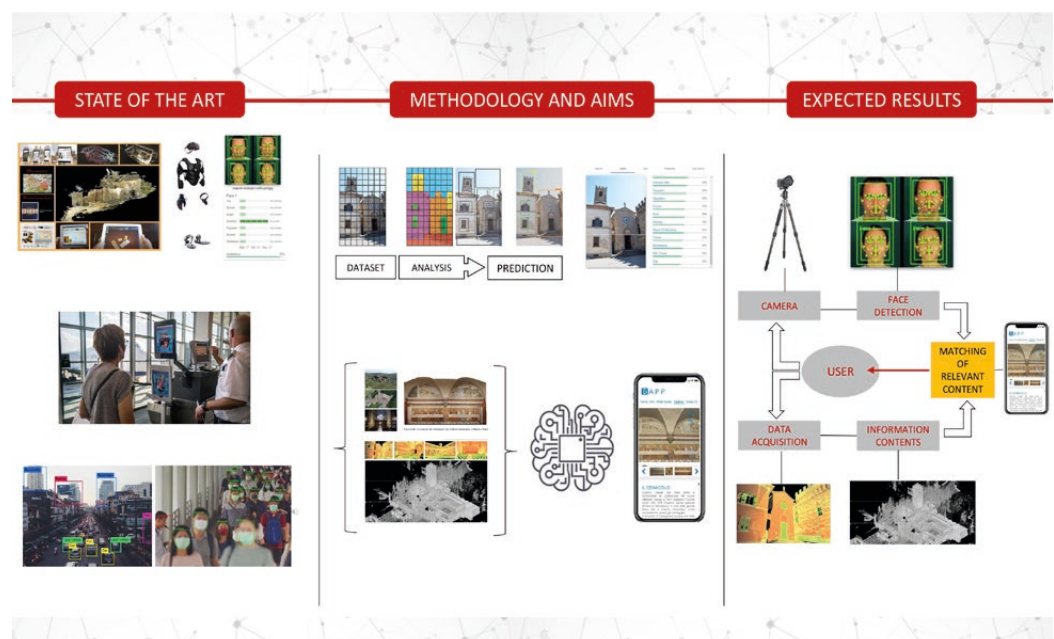
Paola Puma
Giuseppe Nicastro

Abstract

Facial recognition technologies, already used nowadays in many applications, i.e. to support security systems in sensitive buildings, could in a short time achieve widespread use also in others sectors including culture institutions like museums or art galleries. The state of the art in the field of facial recognition allows discriminating factors not only related to the essential somatic characteristics of a person to recognize, with an ever-greater degree of precision, the emotional reactions that may occur on person's face. The article intends to describe the research's EmoDeM experimentation in the museum environment in order to provide a tool capable of interpreting the reaction of a user in front of an artwork and propose a responsive information content coherent what is manifested through facial expressions.

Keywords

digital heritage, museum education, machine learning, image classification, emotion detection.



The most important national and international cultural strategies clarify the importance of actions that include the use of digital languages and virtual environments allowing visitors to better understand the informative features about the observed objects and contexts. The hybridization of traditional spreading methodologies with the tools offered by contemporary digital languages in the field of cultural heritage education therefore becomes an essential asset for those who work in the field of spreading of tangible and intangible values of the cultural heritage [COE 2005; New European Agenda for Culture, European Commission 2018]. The most innovative aspect is not solely in the technological advancement of the supports used to collect, record and transmit information: the kind of digital contents, their being an informative material whose language has a communicative shape with potential and effectiveness only partially investigated, makes their experimentation absolutely central in the contexts connected to education and enhancement like museums and cultural heritage institutions. In fact, digital data has taken on a role not new but certainly enhanced in the current meaning of cross-media contents and effective integration with traditional museum educational tools [COE 2018; Work Plan for Culture 2019-2022; Council 2018]. The experimental use of new media and A.I. in the EmoDeM project represents, from this point of view, an attempt to further advance in the dissemination of cultural heritage as an integral *pivot* of society and human knowledge.

Field of the Research

The EMOTION DETECTION Museum project – EmoDeM concerns a study using some AI applications in the museum environment to provide valuable indicators for museum education. Among the multiple thematic axes of the project, also relevant in negative terms, some can be identified as priorities:

- the cultural axis relating to the EmoDeM's human-centered characterization of 'AI for society' as explicitly pursued by the EU "Between 'AI for profit' and 'AI for control', Europe could embrace 'AI for society', a human-centred, ethical, and secure approach that is true to our core values" [Annoni et al. 2018];
- the scientific axis of the multidisciplinary concept focused on the Survey discipline and its various articulations: Architectural Survey (as a tool for acquiring and representing data for the knowledge of historical contexts), visual communication and digital storytelling;
- the axis of the regulatory policy concerning privacy issues, the governance of the data sharing ecosystem and the management of information.

Aims of the Research

The article illustrates the early stages of the EmoDeM study consisting of the application of facial recognition technologies to museum education [Li, Li 2020, pp. 1-25; Baraldi et al. 2015, pp. 2705-2714]. These AI functions are used to detect and interpret the facial expressions of the visitor in front of the exhibited artwork to propose a content corresponding to what is manifested [Artstein et al. 2014]. The project foresees that, by comparing the images acquired by an 'intelligent' shooting system in the immediate vicinity of the visitor with a dataset of pre-selected and classified images, it is possible to identify and classify with a good reliability the visitor's reaction and consequently propose an information content identified in the connected and predetermined repository as 'associable' to the expressed facial expressions.

State of the Art

The state of the art in the field of facial recognition allows discriminating factors not only related to the somatic characteristics of an individual: there are several experiments in which a software is instructed to identify the emotional states of the framed subject arriving at recognizing, with a constantly increasing degree of precision, the emotive reactions that can be

manifested on the face of an individual (amazement, curiosity, attention, etc.) [Sharma 2019, pp. 834-837]. Facial recognition solutions that use Machine Learning programming are built on a sw models based on the analysis of large information datasets, called Training Datasets, and on the identification of recurring patterns inside it [1] the machine can therefore employ what it has learned from the analysis of the datasets to making autonomous predictions and choices without explicit programming instructions [2] [Alpaydin 2020]. Research and experimentation in the field of machine learning have been demonstrating that the simulation performances of the human processes are currently very unstable as the detection is not always correct and reliable and yet the phases referring the identification of the components of a specific problem, the comparison, the definition of a result, the refinement of the degree of reliability are in very rapid evolution [Alpaydin 2020]. This suggests that software of facial recognition currently used to support the security systems could soon also be used in different scenarios like museum education [Duguleană et al. 2020, pp. 1-17].

Methodology

The methodology used in EmoDeM integrates a workflow of operations divided into three macro-phases: 1) acquisition of facial expressions; 2) image analysis and emotion detection; 3) response of the software through the matching of coherent contents. The first phase of the process has been the data acquisition using a video camera placed near an artwork, used to acquire the visitor's facial expressions [Fyffe et al. 2016; Legendre et al. 2018, pp. 1-2]. The captured image has been subsequently processed therefore starting the emotion detection sequences. This phase certainly represents the most relevant and innovative part of the application as uses specific Machine Learning solutions responsible to evaluate the facial expressions detected and the visible visitor's reactions. The purpose of a machine learning program is to provide a computer with an algorithm consisting of a series of elements that can be compared in order to establish general criteria to learn from and to rely on in subsequent queries. For this purpose, it's essential to develop datasets consisting of many specific data including the significant characteristics to be correctly interpreted from the software during the preliminary training phases. Therefore to enhance the sw detection, in addition to the images directly acquired from the authors during the test, EmoDeM will employ specific datasets expressly designed for emotion detection applications, i.e. CK+ [3] [Lee, Kang 2020, pp. 15-27]. In its complete configuration the last output of EmoDeM is the proposal of relevant contents coherent to facial expression.

First Results

The technical solutions used in the EmoDeM study have been inspired from the results obtained in a previous research on the documentation and spreading project about the abbey of Badia a Passignano [Nicastro 2020]. At the end of that experience, an enhancement project was drawn up whose AI images detection functions have been later integrated into the EmoDeM study: in the first research an application was developed to provide information for the visitors of Badia a Passignano [4]. For intrinsic reasons, the architectural components are, however, characterized by visual patterns that are much more easily typified than those characterizing a totally inhomogeneous and variable set such as people's facial expressions; this constitutes the major limitation of the test, illustrated below also in the failure aspects. At present, the experimentation phase relating to detection in a simulated environment was held, the purpose of which was to verify the degree of confidence of the software in the correct classification of the apparent reactions shown by the faces of people photographed in front of a pictorial work. For this purpose, a set was set up in which 6 participants were invited to observe three different reproductions of pictorial works. A smart workstation was used to capture the images of facial expressions, catalog and analyze them with the Google Cloud Vision API software. Once the acquisition phase was completed, the subjects involved were asked to fill in a questionnaire in order to explicitly indicate the reactions felt during the test, then comparing the answers with the answers provided by the software. The analysis of the acquired data reported that the matching on 18 answers (6 participants examined on three works) between classification operated by sw and by question-

naires resulted in: 9/18 occurrences of 'too low match', 5/18 occurrences of correct outcome, 4/18 occurrences of incorrect recognition. The phases relating to the training of the platform are therefore already partially underway, and the first application test of EmoDem is expected to be carried out in a short time in a museum environment. The on-site installation will allow the feasibility check to evaluate the efficiency of the concept, and to expand the dataset of images collected in different contexts than laboratory experimentation, and test the activation of the response by proposing coherent contents.

Conclusions

Although the research is in its initial stages, the tests carried out in the laboratory have highlighted the first technical critical issues, especially about the images to be used in the sampling phases of the faces. In fact, the most significant advantage of pre-compiled dataset consists of using images suitably prepared to be used during the training phases of the software; on the other side, the images acquired in a controlled research environment present a certain degree of theatricality in the emotions expressed and a lesser naturalness. In the museum workfield phase, in order to improve the responses of the application, a second training dataset will have to be compiled with an expanded number of images with greater expressive naturalness. The major general problems we have encountered so far using AI tools are of two types. The first one is clearly represented by the skills, especially in computer programming languages, necessary to profitably conduct research and experimentation using AI applications; however, it should be noted that the issues relating to AI and machine learning systems are nowadays at the centre of public debate, and explains why in recent years the available tools and documentation have multiplied and its possible operate with these tools with an appropriate level of awareness in a multidisciplinary team. The second threat is represented by the regulatory policy and visitors privacy's protection issues, which must be defined with law criteria to allow the activation of information support only for consenting visitors. In the feasibility analysis of the project, the use of AR solutions could appear as a more effective choice than the efforts required to use an image recognition algorithm. The advantage in this last case is related to the fact that unlike the AR targets solutions, in which the references to the information contents are defined in the design phase and remain so once defined, in this mode the information is provided contextually to the reactions of the observer without there being a preordained sequence. EmoDeM tries to shift attention from the ever-increasing production of information artifacts to the adoption of technological solutions that can assist museum curators in proposing as much as possible inclusive and customized informations. The technological advancement in the field of AI therefore gives us the possibility of operating in the space between the physical object and the related digital contents, with respect to which the concept of EmoDem has been set up, designed precisely with the function of interface between the user and the digital information contents.

The editorial and scientific responsibility of the chapters is recognized to: Paola Puma for Introduction, Design of the research, Aims of the research, Conclusions; Giuseppe Nicastro for State of the art, Methodology, Expected results; both the authors Paola Puma and Giuseppe Nicastro for the figure and References.

Notes

[1] These images must be suitably prepared to be used in a Machine Learning workflow: in a first phase, the content of each single image will be divided into cells, according to a 5x5 grid, containing a defined number of pixels. Subsequently, the contents of any cell will be examined to identify classes of pixels with similar characteristics in areas defined by appropriate tags. The datasets thus prepared will therefore be used in the software training phases, or when further images, not belonging to the datasets used, will be analyzed to identify and classify their content.

[2] Using Machine learning solutions and computer networks in the analysis of large datasets represent practices becoming increasingly widespread. Companies like Google or Amazon offer today increasingly affordable solutions for those who intend to approach these technologies: if until a few years ago, developing an experiment in the field of machine learning implied to access to computer networks available only to specialists and researchers, today Cloud Computing has simplified the access to these resources.

[3] Among these, one of the most used is CK + [Lee, Kang 2020, pp. 15-27], an archive of images acquired in a laboratory environment and catalogued in frame sequences with which it's possible to investigate different types of faces during the gradual transition from a neutral emotional state to one of the seven emotions available (anger, disgust, sadness, happiness, amazement, contempt and fear).

[4] The Passignano abbey, located in Italy between Florence and Siena, has been documented by digital survey that used the integration of data acquisition methodologies to document the consistencies and architectural features of the complex for the purposes of knowledge and cultural enhancement. In this case, the operational scenario included the LS survey, the topographic survey, and terrestrial and aerial SFM. At the end of the research, an enhancement project was drawn up and application was developed to provide information support for visitors. BAPP application uses the Badia a Passignano complex and its works of art figures acquired during the survey campaign to train the image recognition software. By this procedure Bapp makes it possible to frame with the camera of your device any portion of the complex located along a defined visit path to access contents consistent with the object framed by the camera attaining the output deriving from the machine learning processing.

References

Alpaydin Ethem (2020). *Introduction to Machine Learning*. Cambridge: MIT Press.

Annoni Alessandro et al. (2018). *Artificial Intelligence: A European perspective*. Publications Office of the European Union. <https://ec.europa.eu/jrc/en/facts4efuture/artificial-intelligence-european-perspective>

Artstein Ron, Traum David, Alexander Oleg, Leuski Anton, Jones Andrew, Georgila Kallirroi, Debevec Paul, Swartout William, Maio Heather, Smith Stephen (2014). Time–offset interaction with a Holocaust survivor. In *Proceedings of the 19th international conference on Intelligent User Interfaces (IUI '14)*. New York: Association for Computing Machinery, pp. 163-168.

Baraldi Lorenzo, Paci Francesco, Serra Giuseppe, Benini Luca, Cucchiara Rita (2015). Gesture Recognition Using Wearable Vision Sensors to Enhance Visitors' Museum Experiences. In *IEEE Sensors Journal*, vol. 15 (5), pp. 2705-2714.

Council conclusions on the Work Plan for Culture 2019-2022 (2018). In *Official Journal of the European Union* 460, 21.12.2018, pp. 12-25. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018XG1221%2801%29>

Council of Europe Framework (2005). *Convention on the Value of Cultural Heritage for Society, Faro 2005*. <https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680083746>

Duguleană Mihai, Briciu Victor–Alexandru, Duduman Ionuț–Alexandru, Machidon Octavian Mihai (2020). A Virtual Assistant for Natural Interactions in Museums. In *Sustainability*, Vol. 12/17, pp. 1-17.

European Commission (2018). *Commission staff working document A New European Agenda for Culture*. <https://ec.europa.eu/culture/document/new-european-agenda-culture-swd2018-267-final>

Fyfe Graham, Graham Paul, Tunwattanapong Borom, Ghosh Abhijeet, Debevec Paul (2016). Near Instant Capture of High Resolution Facial Geometry and Reflectance. In *ACM SIGGRAPH 2015 Talks (SIGGRAPH '15)*, 32 (1). New York: Association for Computing Machinery.

Lee Hyun–Soon., Kang Bo–Yeong (2020). Continuous motion estimation of facial expression on Jaffe and CK+ dataset for human–robot interaction. In *Intelligent Service Robotics*, vol. 13, pp. 15-27.

Legendre Chloe, Bladin Kalle, Kishore Bipin, Ren Xinglei, Yu Xueming, Debevec, Paul (2018). Efficient Multispectral Facial Capture With Monochrome Cameras. In *ACM SIGGRAPH 2018 Posters (SIGGRAPH '18)*. New York: Association for Computing Machinery, pp. 1-2.

Li Hongfei, Li Quing (2020). End–to–End Training for Compound Expression Recognition. In *Sensors* Vol. 20, pp. 1-25.

Melinte Daniel Octavian, Vladareanu Luige (2020). Facial Expressions Recognition for Human–Robot Interaction Using Deep Convolutional Neural Networks with Rectified Adam Optimizer. In *Sensors*, Vol. 20 (8), pp. 1-21.

Nicastro Giuseppe (2020). *Badia a Passignano, un monastero fortificato in Val di Pesa. Rilievo digitale integrato per la valorizzazione delle strutture architettoniche e del patrimonio artistico*. PhD Thesis, University of Florence. Tutor: Prof. Paola Puma.

Puma Paola (2018). Multidisciplinary experiences of virtual heritage for the documentation of architecture and archaeology within the DigitCH Group – Digital Cultural Heritage Group. In Ioannides, Fink, Brumana, Patias, Doulamis, Martins, Wallace (eds.), *Digital Heritage Progress in Cultural Heritage: Documentation, Preservation, and Protection, Proceedings*. Cham: Springer International Publishing, pp. 242-252.

Puma Paola (2019). The digital image as complex environmental interface: a scenario additional reading. In *IMG Journal*, 1, pp. 268-277. Bologna: Alma DL Journals, pp. 1-10.

Sharma Akshdeep (2019). Emotion Detection and Feedback Generation: Survey. In *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 7 (1), pp. 834-837.

Suguna Ramadass, Devi Shyamala M., Kushwaha Akash, Gupta Puja (2019). An Efficient Real time Product Recommendation using Facial Sentiment Analysis. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1-6.

Authors

Paola Puma, Dept. of Architecture, University of Florence, paola.puma@unifi.it

Giuseppe Nicastro, Dept. of Architecture, University of Florence, giuseppe.nicastro@unifi.it

