



HAL
open science

Face Value: Analyzing and Visualizing Facial Data

Everardo Reyes-García

► **To cite this version:**

Everardo Reyes-García. Face Value: Analyzing and Visualizing Facial Data. *Lexia - Rivista di semiotica*, 2021, <10.4399/978882553853323>. <hal-03181816>

HAL Id: hal-03181816

<https://hal.science/hal-03181816v1>

Submitted on 31 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Face Value: Analyzing and Visualizing Facial Data

EVERARDO REYES*

TITOLO IN ITALIANO: *Valori facciali: analisi e visualizzazione di dati facciali*

ABSTRACT: The increasing popularity of face recognition software introduces new forms to interact with digital face representations. This article explores some of the forms that have emerged in the field of automatic face recognition including face detection, face identification, or face classification. The visual representations generated in technical and scientific literature are taken into account in order to study their relationship with user interface elements. At this point, the article adopts a cultural perspective based on semiotics, media studies and cultural analytics. Human-computer interaction is conceived as an enunciation process and some modalities are regarded according to the practice of face recognition software. In the end, the article discusses an original project in cultural analytics, created from images of album covers and analyzed from their occlusion and distortion values, which is an interpretative reading that technical faces also evoke.

KEYWORDS: Facial Recognition Software; Facial Data; Face Occlusion; Face Representation; Interface Logics

1. Introduction

Face recognition technology has become a popular functionality in software applications that we use on an everyday basis, from automatic tag-

* University of Paris 8.

ging in web-based social networks to face authentication to log into smartphones and computers. The way in which such functionalities are deployed is similar to a black box where the components and mechanisms are hidden to the user, provoking that applications are taken for granted. In this article we are interested in taking a glance at the field of automatic face recognition from the perspective of semiotics, media studies and cultural analytics.

First, we overview the principal areas of interest that constitute the domain of face recognition. Then, we focus particularly on the different manners in which the face is described and represented, both in graphical form and for computing purposes. At this stage, we evoke the characteristics and variety of facial data, as well as the challenges it rises in cultural-oriented fields of study. In a final endeavor, we present an experimental approach to face visualization using album covers as corpus of analysis.

2. Overview of Face Recognition Technology

Since the 1970s, the development of automatic face recognition has achieved substantial progress thanks to its close relation to visual computing, including a combination of techniques for image processing, computer vision, and image synthesis. In fact, automatic face recognition encompasses a set of different areas according to the overall problems it tries to solve: face detection, face representation, face identification, facial expression recognition, and face classification.

Today, there is a vast number of scientific literature in each one of those areas. To have an idea, repositories like arXiv² have registered only in 2020 more than 550 scientific papers on face recognition. The ACM Digital Library³ registers almost 4,000 papers since the 1970s, while the IEEE Xplore⁴ almost 3,500. In these texts, the diversity of topics ranges from methods and algorithms to applications in different fields: law enforcement, video tracking, gaming, computer animation, robotics, to mention a few.

2. <https://arxiv.org/> (last accessed 18 September 2020).

3. <https://dl.acm.org/> (last accessed 18 September 2020).

4. <https://ieeexplore.ieee.org/> (last accessed 18 September 2020).

A broad overview of face recognition technology can be seized by following each of the problems trying to be solved in a sequential order (in an analogous way as Samal and Iyengar 1992):

- Face detection is mainly focused on locating faces within the space of a digital image. Typically, a successful face detection routine outputs at least the vertex coordinates of a bounding box surrounding the face space. Other forms of showing a detected face rely on how face representation and description are implemented in the system.
- Face representation deals with the manner in which data is stored and handled. A regular practice is to constitute a face dataset (a set of image files respecting same format, size, color depth, or face alignment) and its description in form of a database (an text-based file in standard format such as CSV or JSON) that organizes each image file according to measures and values.
- Face identification is in charge of assessing the similarity of a single face (the probe) among a group of images (the gallery). It could also exist the case of face validation, comparing a face image with a gallery of images of a single person. Furthermore, with deep learning techniques we talk about face clustering as another type of identification, using unsupervised learning (i.e., no gallery is used) to automatically classify the different images of a data set (Kroon, Hanjalic & Boughorbel 2007).
- Facial expression recognition evaluates a language-based emotion potentially present in a face. Perhaps the most influential research in this area is the Facial Action Coding System (FACS) (Ekman and Frieder 1977). The idea was to organize a combination of muscle movements, called Action Units, that together defined six universal emotions: happy, sad, surprise, fear, anger, and disgust. Other developments have been more ambitious and defined 135 emotions based on the prototypical approach introduced by Rosch (Shaver et al. 1987).
- Face classification assigns high-level (semantic) categories that might refer to gender, age, style, type, or even identity (like in celebrity identification). Broadly speaking, these methods and algorithms inherit background knowledge from the computing branch of pattern classification. One the most efficient methods is, for example, the Support Vector Machine (Vapnik 2000), which has been one of the principal building-blocks for many contemporary systems and models.

3. Computational representation of faces

Let us take a look at different manners in which faces are represented in computing research and development. Although this task implies to dig into specialized technical literature, the scope is useful if we want to find relations between the visual models and their experiences in other disciplines and contexts. Moreover, as automated scripts like crawlers, spiders, and bots become more active than actually human users on the web, the facial depictions that machines exchange can be seen as a case of transhuman representation (we will come back to this observation in the following sections).

3.1. Geometry-based representations

Early systems were based on geometrical measures estimated from physical human faces. Computer scientist Michael Kelly elaborated the method called “planning” which showed successful results to detect edges of facial features (Fig. 1). Another important contribution was initiated by Takeo Kanade who introduced spatial parameters to detect regions and to calculate rounded contours, like chins (Fig. 2).

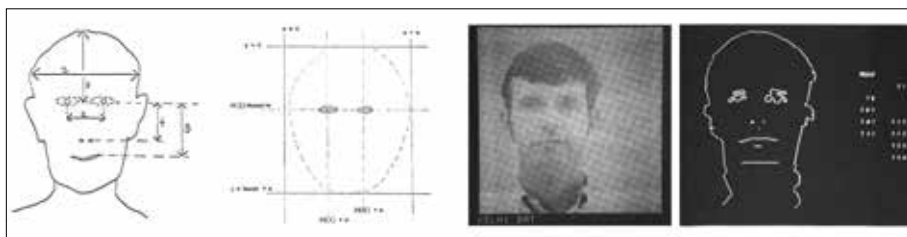


Figure 1. Geometrical representations by Kelly (1970).

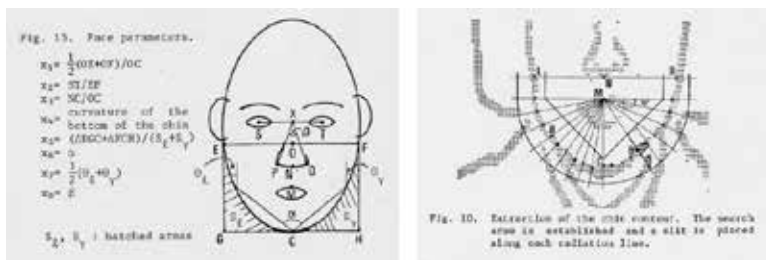


Figure 2. Geometrical representations by Kanade (Sakai, Nagao, and Kanade 1972).

3.2. Signal-based representations

Improvements in hardware and software allowed the visual computing field to start dealing with images using signal processing techniques. In 1987, the well-known eigenfaces approach was envisioned by Lawrence Sirovich and Michael Kirby. It applied Principal Component Analysis (PCA) to the quantification of a series of images. The result was a new image that represented the average of the whole set or it could also render the delta image (or caricature) based on a sample face (Fig. 3).



Figure 3. Eigenfaces by Sirovich and Kirby (1987).

Eigenfaces were adapted by Matthew Turk and Alex Pentland for face detection. Such implementation was possible by storing visual values as vectors (hence named eigenvectors) and comparing new images with the ones previously quantified. The process also relied on PCA but Turk and Pentland not only showed the average and delta images but also other eigenfaces resulting from different combinations of components (Fig. 4).

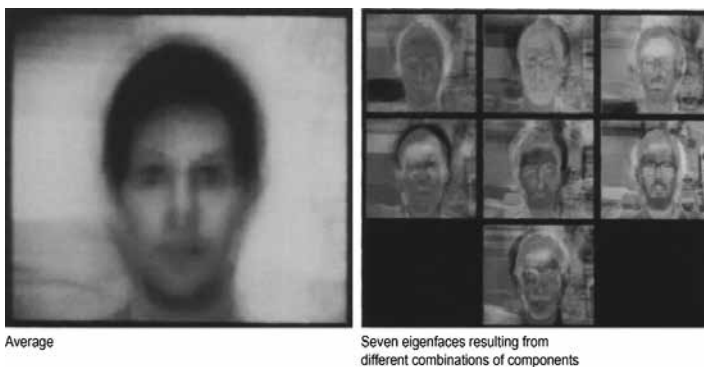


Figure 4. Eigenfaces by Turk and Pentland (1991).

Another derivative development that emerged from eigenfaces are the Fisherfaces (Belhumeur, Hespanha and Kriegman 1997) which are based on the statistical approach to taxonomic classification by Sir Ronald Fisher. In this case, it uses a class-specific linear method for dimensionality reduction, for example the class “glasses” can help to determine if a person is wearing them or not. The linear method was systematically defined as Linear Discriminant Analysis (LDA) by Kamran Etemad and Rama Chellappa (1997). Using a series of wavelet transforms, LDA also implied to configure an eigentemplate as the discriminant component.

Understanding visual information as signal has been a fundamental piece to the development of contemporary face recognition software based on machine learning. Because images can be analyzed as numerical units, a vast amount of techniques have been created to segment and compute correlations. While the first image measures included distances, angles, and pixel coordinates, now we talk about features, filters, descriptors and classifiers to handle metrics, ranks and rates.

Fig. 5 exemplifies different types of signal preprocessing applied to images for face recognition. Of course, the amount of graphical representations resulting from variations, adjustments, and repurposes of these methods is too large to be surveyed here. We have only selected some examples, trying to find the earliest possible depiction before the model was adopted widely. The interested reader can refer to already existing reviews in (Zhao 2003).

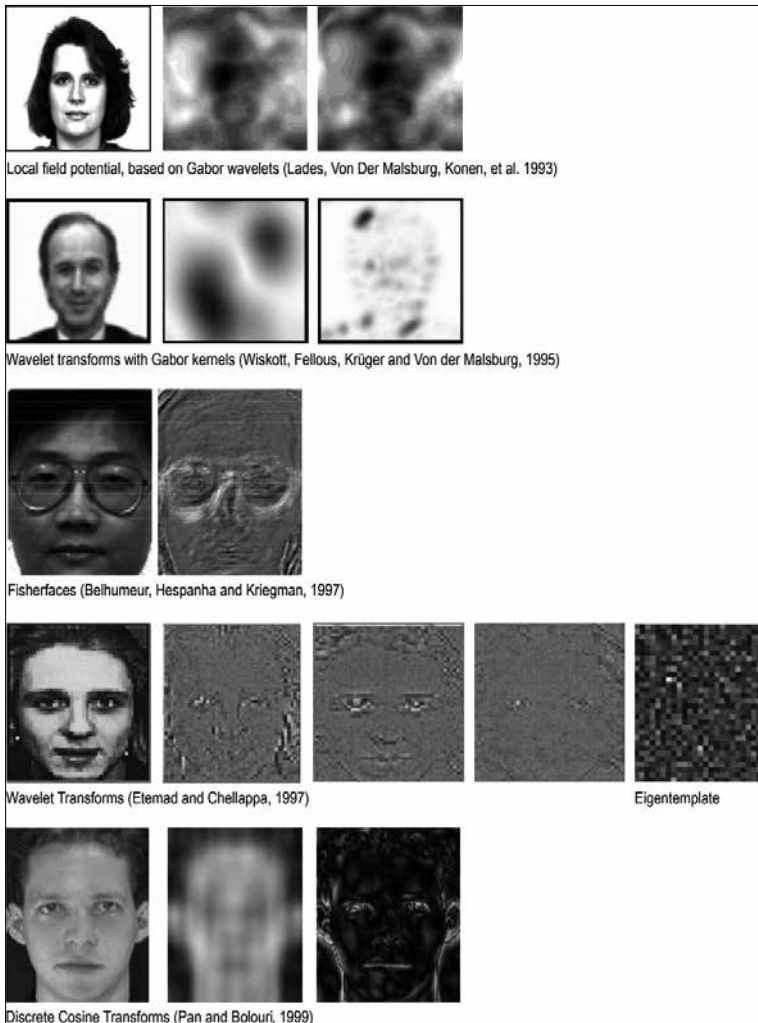


Figure 5. Different types of signal processing applied to face images.

Regarding the location of facial features, an early mechanism consisted on placing a rectangular grid on top of the face image. The method was exemplified as an image graph by Buhmann, Lades and von der Malsburg (1990). In their first version, the nodes were labeled using Gabor wavelet coefficients, but the technique was rapidly enhanced to locate facial landmarks (Wiskott, Fellous, Krüger and von der Malsburg 1997) (Fig. 6).

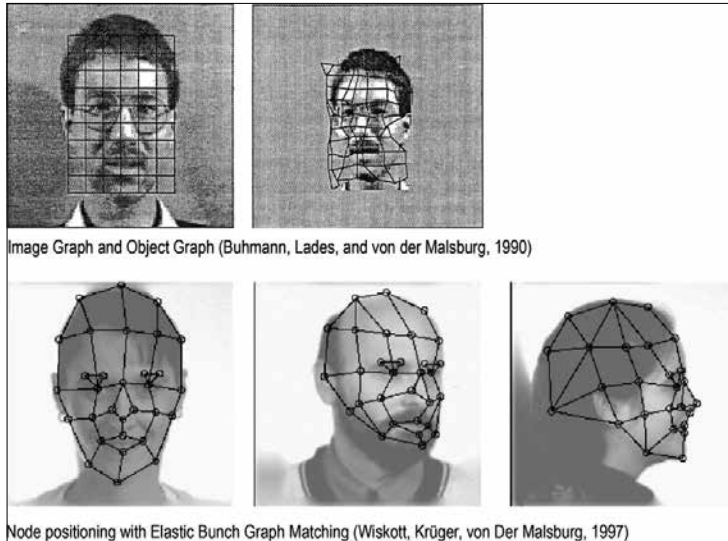


Figure 6. Location of facial features as nodes.

To conclude this section, we highlight that a more robust study should reserve a place to other visualizations that are constructed in multidisciplinary domains. We are thinking about infrared and thermal imagery, fractal methods and 3D face recognition, just to mention a few (Fig. 7).

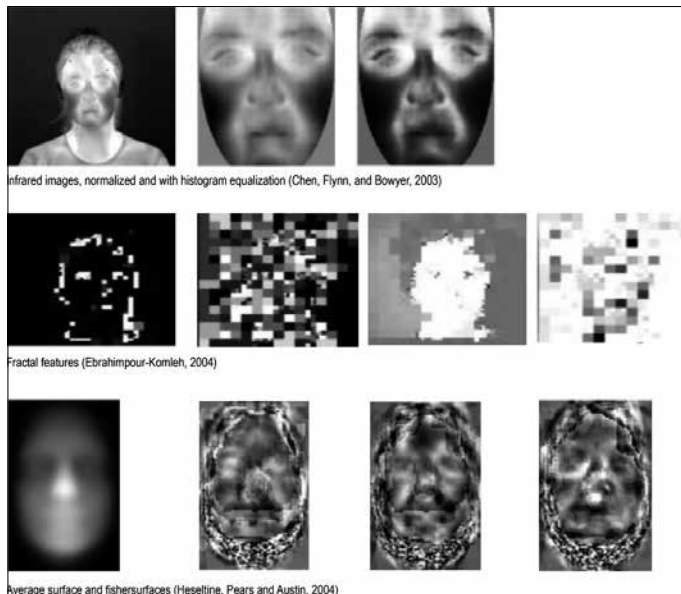


Figure 7. Infrared, fractal and 3D face representations.

4. Using face recognition software

With the increasing popularity of face recognition technology in software applications that we use on an everyday basis, it is important to distinguish different practices that are being shaped. Although it is not our objective to elaborate a typology of software uses, in this section we identify some uses within the context of interactive gestures and graphical user interfaces.

As an entry point, the graphical interfaces that appear on the electronic screen can be described in terms of its constitutive parts: how is it made and how does it establish a communicative relationship between users and developers? This is paramount from a semiotic perspective because digital images host, locate and define modalities of interaction: to believe, to do, and to know. It is a process of enunciation where actors choose among a series of possibilities that will complete the meaning.

Following a narrative schema (Greimas and Courtès 1979), a person seen as a potential user will be first enacted within a narrative quest to believe in the software values (the functions but also the expected rewards). This modality is located before the performance, when the user has a virtual idea of what will happen upon clicking on a button. The kinds of discourse in which we commonly notice the modality of believe are those prior to installing a software application, typically the software download page. PhotoLab, Snapchat or FaceApp, to mention only some examples, signify and communicate precisely their face recognition values in the App Store or Google Play.

We can think about a different mode of presentation based on doing. In this modality, the potential user becomes a realized subject at the end of the quest, testing factually the software functionalities. To give an example, let's consider a few demo tools. Google Cloud's Vision API⁵ allows any visitor to try the image recognition module from different landing pages: the Vision AI project and the Cloud Vision Documentation. While the text sentences in the former page are clearly targeted to decision-takers and managers, the latter addresses a more technical user. However, in this demo, just like in EyeEm or Betaface, the software functionalities can be performed by any kind of user profile, by simply dragging/dropping an image and seeing immediately the results of the process.

5. <https://cloud.google.com/vision> (last accessed 18 September 2020).

A third modality refers to knowing. In this mode of existence, the person has become a user, i.e. an actualized subject that has tested, understood, and is now able to produce new configurations. Following our typology, we situate here software applications like RunwayML⁶ or programming libraries like *cmltrackr*⁷ that are oriented towards more experienced users but emphasizing, nonetheless, the value of “easiness-to-use”. In these cases, the user has often access to code chunks, parameters, and technical reference.

As we can appreciate, the elements and configurations of graphical interfaces vary from one modality to another. In a previous work, dedicated to software applications for visual computing (Reyes 2017), we argued that, after a period of socio-cultural adoption, interface configurations give shape to patterns combining spatial and visual assemblies. We call these patterns “interface logics”, building on the notion of media theorist Noah Wardrip-Fruin, who uses the term “operational logics” to refer to “patterns in the interplay of data, process, surface, interaction, author, and audience” (2009, p. 13).

Patterns help to identify models that can be adopted in numerous situations. In visual computing software that we have examined, the logic of “feedback” refers to simple graphical means that indicate the state of the program (e.g., alert messages, progress bar); “domain crossings”, recalls that interfaces combine different domains and fields. Most obviously, this can be appreciated in the language and names of menus, and less evident are the use of techniques forged in special domains (e.g. filters from signal processing adapted to images); and, “dynamic and live data” shows that some information that an interface depicts only exists upon request of the user (requests to a server, real-time updates).

Following the same line, perhaps a more refined category such as “facial logics” needs to be elucidated for face description and representation. A thorough analysis of face recognition software that associates their graphical interfaces and the narrative schema is an endeavor that we are interested to pursue in a near future. We also consider necessary to situate this work in the long run: recording how the enunciation marks imprinted by the sender evolve throughout software versions.

6. <https://runwayml.com/> (last accessed 18 September 2020).

7. <https://github.com/auduno/cmltrackr> (last accessed 18 September 2020).

5. Experiments with facial data

At first sight, the face visualizations in section 2 might evoke interpretations referring to science/technology but, from another standpoint, also to occlusion/distortion values. The fact that those images are not necessarily targeted for human consumption strengthens their illustrative purpose when they appear as supporting material in research publications. However, those images do exist in everyday life. They are situated at the middle of an inter-communicative process that connects a user with the creator of a digital tool. In other words, every time we run a facial recognition program, the processes happening under the hood of the graphical interface are based on those kinds of facial descriptions.

In social and human sciences, facial occlusion and deformation have a long tradition (see for example Eco 2007; Leone 2018). Although the main media that have attracted the reflection on faces remain painting, photography and cinema, other forms such as album covers, graffiti, and comics, can also be included in a multidisciplinary scenario combining visual semiotics and cultural analytics. This is of course possible thanks to the digitization of material supports: from physical substances to pixel/vector-based formats.

In our research we have already conducted a series of experiments including popular music and their visual forms (Reyes and Manovich 2020). Extending this motivation, we assembled a dataset containing 542 images of album covers with occasion of our talk at the symposium “Transhuman Visages: Artificial Faces in Arts, Science and Society” organized by the semiotics scholar Massimo Leone at the Polish Institute of Advanced Studies in January 2020. The experimental approach that we follow consists on applying our methodology to our own corpus of analysis and in discussing some visualization prototypes that support our findings related to strategies of face occlusion that challenge research on automatic face detection.

Our dataset contains the albums released in 1998 and in 2018 according to data gathered from Wikipedia⁸ (Tab. 1). We chose these two years in order to obtain samples from two different moments in media history: on

8. https://en.wikipedia.org/wiki/1988_in_music and https://en.wikipedia.org/wiki/2018_in_music

the one hand, the contemporary and, on the other hand, the time before software for digital image manipulation became massive. Indeed, by 1987 Adobe had just released the first version of Illustrator and in early 1990 they would start distributing Photoshop.

	1988	2018	Total
Number of images	227	315	542
One face	95	157	252
Two faces	11	21	32
Three faces	14	7	21
Four faces	11	10	21
Five faces	8	5	13
More than 5	11	7	18
Illustration style	16	29	45
Occluded faces	14	18	32

Table 1. Data set of album covers faces, 1988 and 2018.

Our first classification distinguishes if a face is present in an image and how many are detected. This process can be achieved easily with Google Cloud Vision API which recently increased the number of detected faces in its free try version. The second classification is visual style. By taking a look at the automatic labeling, many images included classes like “typography”, “art”, “logo”, “photography”. To simplify, we filtered images with the label “illustration”. The third category that we use is occlusion. The fact that many images contained distorted or partly masked faces were labeled as “person”, “headgear”, or “headwear”. This property allowed us to create a larger cluster for occlusion strategies.

Overall, our dataset shows predominantly album covers with only one face, being 2018 the more accentuated year in comparison to 1988. From an interdisciplinary perspective, we could ask questions related to this trend such as: Does it mean that today there are more single-membered bands than there were in 1988? Or are there any artists that followed a solo

career from bands in 1988? Are there any artists who released an album in both years? While the first two questions need the consideration of a larger span of time and documentation, the third is easier to answer from the data table.

Regarding our second classification, visual style, we distinguish several observations. First, the more persistent visual framing is the portrait, mostly in the mid and large shot formats, and rarely as close-ups. This observation is coherent with the number of single faces detected in the album cover. Second, in relation to the category “illustration”, it remains the preferred graphic style for hard rock & roll and heavy metal bands in both years. Third, it appears there is an increase use of digital visual effects and gradients in the 2018 set. To verify these intuitions, it would be time consuming to evaluate every image in search of alterations mainly because visual styles and techniques become interrelated at many different layers. For example, an illustration could start as a pencil sketch, then traced as vector image, then modeled as 3D shape, the texturized as cartoon drawing and finally stylized with some ray-tracing filters for ambient and illumination.

Another aspect concerning visual style can be approached from the aesthetic angle. Tools like EyeEm include an aesthetic analysis⁹ module that calculates a score for visual appealing of images. Whereas the tool is meant to feed a deep learning model with photographs that share a similar visual identity, we tested the demo with images from our collection. Not surprisingly, higher scores were more regular in 2018 than in 1988 as their visual style is closer to the photographs uploaded and curated by the contemporary creative community, on which the neural model is trained¹⁰.

Finally, regarding the classification of occlusion in both years, our filter gave us a cluster of 32 images. We then manually organized the images into four different types of covering the face: by wearing a mask, by visual effects, by the pose of the body, and by the frame of the image (Fig. 8).

9. <https://www.eyeem.com/eyeem-vision>.

10. <https://developer.nvidia.com/blog/understanding-aesthetics-deep-learning/>.

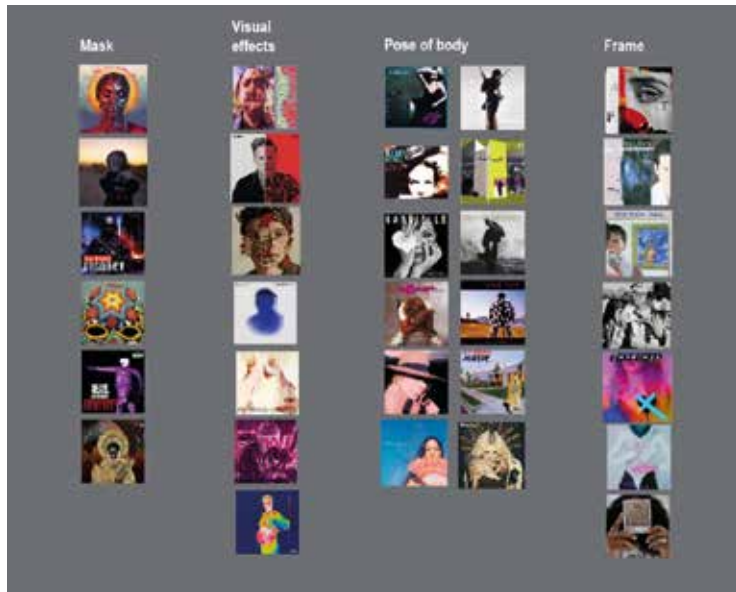


Figure 8. Album covers filtered by occlusion types. Images were obtained from Wikipedia, individual copyrights belong to the labels or the artists.

A deeper observation of occlusion leads to ask not only how to occlude but why. In this respect, the image in itself is not always enough to determine the reason and, in a medium like album covers, we have to turn to the music, lyrics, temporal context, artist's ideology / subjectivity, and other factors. Broadly speaking, we suggest that values in our set range from mystery to over-exhibition and from intimacy to anonymity. In any case, artistic images invite to dynamic and multiple readings that could inspire other criteria of analysis. In an exercise to define a simple tensive map, we locate on one axis a range from shown faces to hidden faces and, on the vertical axis, from distorted to sharp depiction (Fig. 9).

We believe this kind of speculative manipulation of facial data, blending automatic and manual techniques, is useful in the search for different meaning values. In the particular case of occlusion in album covers, they are hardly associated to the canonical properties of a prototypical album cover. Instead, they appear as ideological and critical cultural statements.

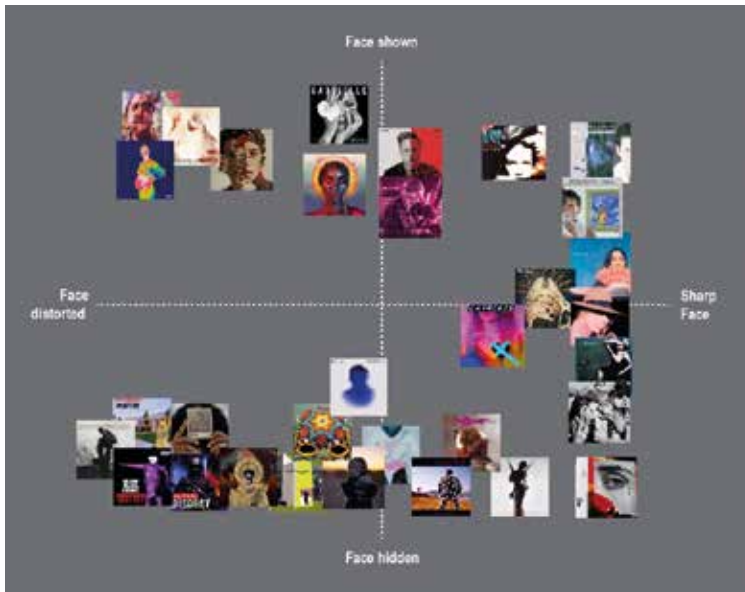


Figure 9. Map of images according to four axes values.

6. Conclusions

In this contribution, facial data is understood as angles, distances, pixel coordinates, features, filters, descriptors, and classifiers that are employed to describe and to represent faces. More specifically, facial datasets can contain data structures such as metrics, ranks, rates, and semantic categories. The variety of facial data is related to the areas of interest in face recognition as a field of study. We discussed some pioneering works in face analysis and face representation. We observed how the technical aspects of the field are more or less present or transformed at the surface level of face recognition software. For this matter, we regarded contemporary software applications from a semiotic and communicative perspective. We proposed to conceive graphical interface elements as patterns that follow “interface logics”, but that could be better refined through a dedicated “facial logics” category.

Our experimental part brought together technical faces in scientific research and cultural faces in the domain of music and graphical design. From an interpretative reading, we observed that values of occlusion/ deformation are common to both types of representation. Then, using

these variables we sketched a couple of data visualizations that support the design of an online version.

In future work we plan to continue our study of how interfaces work in order to produce, create and speculate about new graphical models; not only as concepts but also as objects, in the form of technical sketches or prototypes. In this scenario, we are developing an interactive visualization to explore our dataset but also to put in practice different interface logics.

Bibliographic references

- BELHUMEUR P., J. HESPANHA, and D. KRIEGMAN (1997) "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19,7: 711-20.
- ECO U. (2007) *Historia de la fealdad*, Lumen, Barcelona.
- EKMAN P. and W. FRIESEN (1977) *Manual for the Facial Action Coding System*, Consulting Psychologists Press, Palo Alto.
- ETEMAD K. and R. CHELLAPPA (1997) *Discriminant Analysis for Recognition of Human Face Images*, "Josa" a 14, 8: 1724-33.
- GREIMAS A. and J. COURTÉS (1979) *Sémiotique: Dictionnaire raisonné de la théorie du langage*, Hachette, Paris.
- KELLY M. (1970) *Visual Identification of People by Computer*, Doctoral dissertation, Department of Computer Science, Stanford University.
- KROON B., A. HANJALIC, and S. BOUGHORBEL (2007) "Comparison of Face Matching Techniques under Pose Variation", *Proceedings of the 6th ACM International Conference on Image and Video Retrieval (CIVR '07)*, ACM Press, New York: 272-9.
- LEONE M. (2018) *The Semiotics of the Face in the Digital Era*, "Perspectives", 17: 27-9.
- LYONS M., J. BUDYNEK, and S. AKAMATSU (1999) "Automatic Classification of Single Facial Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21, 12: 1357-62.
- REYES E. (2017) *The Image-Interface: Graphical Supports for Visual Information*, Wiley, Hoboken, NJ.
- _____. and L. MANOVICH (2020) "Cultural Viz: An Aesthetic Approach to Cultural Analytics", *Leonardo*, 53, 4: 408-14.

- SAKAI T., M. NAGAO, and T. KANADE (1972) "Computer Analysis and Classification of Photographs of Human Faces", *First USA-Japan Computer Conference*, Kyoto University, 2-7.
- SAMAL A. and P. IYENGAR (1992) *Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey*, "Pattern Recognition", 25, 1: 65-77.
- SHAVER P., et al. (1987) *Emotion Knowledge: Further Exploration of a Prototype Approach*, "Journal of Personality and Social Psychology", 52, 6: 1061-86.
- SIROVICH L. and M. KIRBY (1987) *Low-Dimensional Procedure for the Characterization of Human Faces*, "Journal of the Optical Society of America", 4, 3: 519-24.
- TURK M. and A. PENTLAND (1991) *Eigenfaces for Recognition*, "Journal of Cognitive Neuroscience", 3, 1: 71-86.
- VAPNIK V. (2013) *The Nature of Statistical Learning Theory*, Springer, Berlin
- WARDROP-FRUIIN N. (2009) *Expressive Processing: Digital fictions, computer games, and software studies*, MIT Press, Cambridge, MA.
- ZHAO W. et al. (2003) *Face Recognition: A Literature Survey*, "ACM Computing Surveys", 35, 4: 399-58.