SN Computer Science (2021) 2:236
https://doi.org/10.1007/s42979-021-00604-w

ORIGINAL RESEARCH

The X-Faces Behind the Portraits of No One

João Correia[1] · Tiago Martins[1] · Sérgio Rebelo[1] · João Bicker[1] · Penousal Machado[1]

Received: 31 July 2020 / Accepted: 20 March 2021

Abstract
This paper presents our work on the computational creation of photorealistic face images with a focus on how we transformed our generative and evolutionary system *X-Faces* into an interactive Media Art installation entitled *Portraits of No One*. The *X-Faces* system resorts to Computer Vision and Computer Graphics to automatically create new face images by recombining facial parts extracted from existing examples, along with Evolutionary Computation and Machine Learning to automatically explore the vast space of composite faces that can be generated. This system consistently generates face images with great photorealism and value, for example, for Data Augmentation or to assess and improve the performance of face detectors. In this paper, we describe how we explored the capabilities of *X-Faces* in a Media Art context. The result, the interactive installation *Portraits of No One*, synthesis and displays facial portraits on the borderline between the real and artificial using the facial features captured from its audience. The photorealism of the faces displayed in the installation space invoked the capabilities of Artificial Intelligence to generate content that makes people question its veracity. With the installation *Portraits of No One*, we allow people to engage with *X-Faces* and to be part of them.

Keywords Artificial intelligence · Computer graphics · Computer vision · Evolutionary computation · Image generation · Media art

Introduction

Recent advances in Computer Vision (CV) supported the emergence of detection and recognition tools that are increasingly more capable. This enabled the emergence of Artificial Intelligence (AI) techniques that allow the quick and easy generation of images that achieve a high level of photorealism. This paper presents our research on the computational generation of face images. This research has its genesis in our goal to expand existing datasets of face images by using Evolutionary Computation (EC) to evolve new face examples consisting of recombinations of the original ones. We developed a generative process based on CV and Computer Graphics (CG) techniques that enables the automatic recombination, or exchange, of the facial parts (i.e. eyebrows, eyes, noses and mouths) from existing face images and, this way, generate new composite faces. To fully explore this generation process, we employed EC, combined with an automated fitness function based on the response of the classifier, to automatically find generated faces that are photorealistic and put existing face classifiers to the test [1]. We called the generated faces *X-Faces* (after eXploit Faces). Afterwards, we successfully used this approach to perform Data Augmentation (DA) on Face Detection (FD) datasets and retrain the classifier to improve its performance [2].

More recently, we explored the capabilities of the developed approach in a more artistic context. The result is the interactive Media Art installation *Portraits of No One* [3]. This installation generates and displays portraits of human

This article is part of the topical collection "Evolution, the New AI Revolution" guest edited by Anikó Ekárt and Anna Isabel Esparcia-Alcázar.

✉ João Correia
jncor@dei.uc.pt

✉ Tiago Martins
tiagofm@dei.uc.pt

✉ Sérgio Rebelo
srebelo@dei.uc.pt

João Bicker
bicker@dei.uc.pt

Penousal Machado
machado@dei.uc.pt

1 Department of Informatics Engineering, Centre for Informatics and Systems of the University of Coimbra, University of Coimbra, Coimbra, Portugal

Published online: 24 April 2021

Fig. 1 Environment of the installation *Portraits of No One*. Photo by José Paulo Ruas/DGPC 2019

faces by recombining the facial parts that are retrieved from the faces of its audience (see Fig. 1). Whenever users give their faces to the installation, they will immediately see parts of their faces contaminating the ever-changing portraits that are projected on the walls. In a way, one could say that the installation feeds on the faces of people who interact with it. The photorealism of the portraits makes people question themselves about the veracity of what they are seeing, which is located somewhere on the borderline between the real and artificial. This installation was designed for Sonae Media Art Award 2019, selected as a finalist artwork and, consequently, exhibited at the National Museum of Contemporary Art, in Lisbon, Portugal. In this paper, we describe the application of the *X-Faces* system in a Media Art scenario to create the installation *Portraits of No One*.[1]

The remainder of this paper is organised as follows. Section 2 overviews related work both on the computational generation of human face images and generated faces in visual art. Section 3 presents our approach to generate face images. Section 4 describes the application of this approach to create an interactive media art installation. Finally, Sect. 5 presents our conclusions and directions for future work.

Related Work

In the context of this work, we are focused on the study of the computational generation of human faces, developed with two main purposes: (i) to generate new face images with DA purpose; and (ii) to creatively use artificial generated human faces in visual artworks, especially in a Media Art scenario.

Computational Generation of Face Images

The computational generation of new images instances, especially faces, to augment the baseline training datasets has been explored using multiple techniques and by multiple authors. The techniques available range from simple transformations, e.g. rotation, scale, flip horizontally or vertically; to more complex ones, e.g. adding noise, filters, distortions [4]. However, these techniques are generic image processing operations which do not exploit knowledge of the underlying problem domain to generate new instances [5]. However, some approaches take advantage of the problem domain. Jiang et al. [6] proposed a 3D reconstruction methodology to augment the existing face images in different poses, illuminations and expressions. Mohammadzade and Hatzinakos [7] used subspace projection of facial expressions to generate new facial expressions. The use of a Deep Neural Network (DNN) in the process is also explored. Seyyedsalehi and Seyyedsalehi [8] used DNN to extract under unconstrained conditions, the expression, the pose and identity of a set of face images, generating new instances based on an input image. With the same objective, Nirkin et al. [9] employed a DNN for face swapping under unconstrained conditions showing that they can generate a considerable amount of face images that improve the base model on face recognition. More recently, Lv et al. [10] presented different face augmentation algorithms, including landmark perturbation, and synthesis methods to generate face images with different glasses, haircuts, poses and illuminations. They show that the models trained with generated face images achieve a similar performance when compared to deep models trained with a much larger dataset. More recently, with the introduction of Generative Adversarial Networks (GANs), several variations of this algorithm and network architectures were reported. The state of the art algorithm is the StyleGAN, able to generate high-resolution images on a variety of datasets [11, 12]. One of the most notable deployments of this model is the *This Person Does Not Exist* project [13] that

[1] Supplementary materials and information about this installation Portraits of No One are available at https://cdv.dei.uc.pt/portraits-of-no-one/.

shows photorealistic faces generated by StyleGAN. These are impressive results but at this point in time it is computationally expensive to train and deploy a model with the characteristics of StyleGAN to have control of the faces that are generated.

The employment of EC under DA of face image datasets was explored by different authors. The following works have a crucial factor in common: using the output of the classifier to assign fitness. Chen et al. [14] used a Genetic Algorithm (GA) to generate faces, where the encoding of the individuals is a vector of integers representing the pixel intensity values. The variation operators are image filters and segmentation of predefined parts (e.g. forehead, eye, nose and mouth). Machado et al. [15] used Genetic Programing (GP) to evolve images for the dataset of negative examples, i.e. instances that do not contain a face. The individuals are encoded as expression trees that generate images. The evolved images that were incorrectly classified as positive instances were selected to improve the negative dataset. The augmented dataset was used to train the classifier and, this way, improve its performance.

Generated Faces in Visual Art

In the Media Art context, the exploration and generation of human faces as a subject for visual art is widely explored, mainly throughout the exploration of tools such as video capturing, FD techniques and/or custom-made software. Early developed artwork includes for instance the installations *Video Narcissus* (1987) by Shaw [16], *Interfaces* (1990) by Kac [17] or *Solitary* (1992) by Biggs [18]. This subject has become even more popular recently due to the technological advances in CV. This way, visual artists are exploring these technologies with three main purposes in mind: (i) to create more engaging artistic experiences that enable the users to participate in the artwork; (ii) to create interfaces that enable the users to interact with each other; and (iii) to generate artificial human faces.

CV technologies are employed to gather audience features to be used in the creation of artworks and, at the same time, to create a more engaging artistic experience for their artwork. The interactive installation *Reface* (2007) by Levin and Lieberman [19] generates group portraits of its audience employing this principle. The group portraits are generated through the combination of videos' excerpts of mouths, eyes and brows from different users that interact with the installation. Also, the users may interact with the artwork using eye blinks. In 2015, Howorka [20] developed a digital mirror that displays an average face of all people who have been in front of the installation. In 2019, Al Tawil [21] presented the installation *IDEMixer* that generates live portraits of the participants by blending pairs of faces captured by two different cameras. In the same year, Suhr [22] presented the interactive multimedia performance *I, You, We*. During this performance, users are invited to take a photo. The photographs are, subsequently, rendered and combined between them, and with other visuals, in a live video that runs in the background. Simultaneously, the performer changes the colours of the video through improvisation with a violin and/or a cello.

Some artworks explore these technologies, using the faces of the users as a visual artefact to promote novel ways for the audience to interact with the artwork and with each other, in a direct or indirect way. The installation *Sharing Faces* developed by McDonald, in 2013 [23], behaves as a mirror between two locations: Anyang, in South Korea, and Yamaguchi, in Japan. When users approach the installation, in one of the locations, it tracks their facial expressions and matches them, in real-time, with facial expressions of other people who have already stood in front of the installation in the other location. The artists Lancel et al. [24] developed the installation *Saving Face,* in various geographical contexts, between 2012 and 2016. In this artwork, portraits of users are blended and displayed on public screens. The intensity of each blend is determined by the number of times users made gestures of "take care" on their faces. In 2015, Lozano-Hemmer [25] designed the interactive installation *Level of Confidence*, which uses a face recognition system trained to find similarities between a given face and the faces of forty-three disappeared students in a cartel-related kidnapping in Iguala, Mexico, in 2014. Whenever users stand in front of the installation, it captures their faces and unveils the most similar student and how accurate this match is. In a similar way, Daniele [26] developed the installation *This is not Private,* which enables people to see some of their facial features in the faces of someone else. When users stand in front of a screen watching an interview, their faces slowly merge with the face of the person who is being interviewed. The intensity of the merge is determined by the similarity of the expressions between the user and the interviewee.

On the other hand, some artworks are created only with the goal of generating artificial portraits. *Portraits of Imaginary People* series, developed by Tyka [27], in 2017, presents a series of portraits generated with a GAN trained with thousands of photos of human faces. In 2017, Moura and Ferreira-Lopes [28] explored the creation of imagined faces, which resemble ghost faces, based on false-positive detections on visual randomness. In 2018, Klingemann [29] developed the *Neural Glitch* technique that consists in the random manipulation of the trained weights of GANs to generate new images including artificial portraits.

The state-of-the-art approaches to the generation of new artificial faces mostly use DNNs and GANs. Although these approaches are able to achieve photorealist results, they are computationally expensive and time-consuming to train models that increase the control of the faces that are generated. As a result, these approaches have not been explored to generate face images in the context of participatory media art, which requires real-time generation of distinctive photorealist faces while using an ever-changing dataset of face examples. The *X-Faces* system [1, 2] addresses the generation of faces by automatically recombining facial parts from previous existing faces, which results in a faster, scalable and less expensive approach Therefore, *X-Faces* presents as a suitable approach for computationally creating facial portraits in participatory media art installations.

X-Faces: Generating Faces

At the time of the creation of the *X-Faces* system we required a system that was able to respond to the following requirements: (i) generate automatically, and in a consistent way, a large number of photorealistic faces; (ii) have control over the generative process of the composite faces and scale to rapidly expand the set of face composites. To the best of our knowledge, there were no approaches able to deliver on these requirements. For that matter, we have created a computational system to generate photorealistic face images from parts of existing examples [1, 2]. The system generates new faces by performing the following steps: (i) decomposition of facial parts from multiple images; (ii) assembly of composite images from the separated facial parts; and (iii) exploration of different face composite images. The system combines several techniques ranging from CV, CG and Machine Learning (ML) to generate the faces, converging in an EC approach to generate sets of new face images. In the first iteration, we used EC to generate faces that were not classified as such by classifiers, but from a subjective perspective, they were recognisable as faces [1]. Since the faces generated by the system were misclassified by the detectors as faces, hence considered as exploits, we called them *eXploit Faces* and the developed engine *X-Faces* system. In the second iteration, we invested in making the approach fully automated and suitable for unconstrained conditions, *i.e.* being able to deal with diverse occlusions, poses, variations of scale and appearance. We used the resulting approach for DA to improve the performance of deep face classifiers [2]. In this Section, we describe the iterations and the steps of the *X-Faces* system.

Face Decomposition

In the first iteration of *X-Faces*, the decomposition step was performed by manual annotation of facial parts using a custom annotation tool [1]. This tool was developed to provide annotations for the different parts of the face images: face, left eye, right eye, nose, mouth, left eyebrow and right eyebrow. At this stage, we were in a controlled environment where the faces were in a frontal position with clean backgrounds and lighting conditions, which provided a suitable testbed for a first version of the system. However, the manual process for decomposition would hinder the scalability of this system for a dataset with face images under unconstrained conditions. Thus, in the second iteration, we automated the annotation of the faces by using a face detector and a landmark classifier [2]. Thus, given an input dataset containing face images, we extract some data necessary to generate new faces, namely the facial landmark points and the head pose of each face found in the dataset. For each image in the input dataset, we proceed as follows. As detailed in [2], first, we detect possible face regions in the image using a pre-trained face detector. This detector can search an image for faces and determine their locations and areas on the image. Then, inside each detected face region, we find facial landmarks using a facial landmark predictor that is also pre-trained. This predictor places a sequence of 68 points along the contours of the following facial structures: jaw, eyebrows, nose, eyes, and mouth. Once we have the landmark points, we use them to estimate the head pose, *i.e.* the orientation of the face in space; and we save to file the following data: (i) facial landmark points; (ii) pose information; and (iii) the image path. The output of the stage is a set of files, one per face found in the input dataset. Therefore, the dataset only needs to be preprocessed once, regardless of the number of new instances to be generated. The improvements from the first iteration to the second enable the system to work with different head poses, scales, skin tone variations; in summary, dealing with the faces in an unconstrained manner. Moreover, since we are working with facial landmarks, we only require to annotate or identify the face, in contrast with the first iteration where we would have to identify and manually annotate the face and the parts of the face.

Assembly of Composite Images

Once we have the landmarks and pose calculated for all faces contained in the input dataset, we can proceed to the assemblage of new instances. The creation of a new face instance consists of selecting a face from the dataset and then replacing its elementary parts (eyebrows, eyes, nose and mouth) with parts from other faces. Therefore, each new face consists of a recombination, or a blend, of different faces. However, the automatic recombination of facial parts, especially in an unconstrained environment, demands special care with their alignment to attain natural and seamless facial blends.

The process of generating a new face instance is illustrated in Fig. 2. Given the data extracted during the preprocessing of the input dataset, we begin by randomly selecting one face among all the faces preprocessed. This face will be our *target* face, i.e. the face whose parts will be replaced

Fig. 2 Generation of a new face image. From left to right: (1) input face images; (2) facial key points defined based on the detected facial landmarks; (3) Delaunay triangulation of the key points; (4) triangles with different colours showing their correspondence between faces; (5) input faces warped to the *target* face; (6) facial parts to be blended; and (7) the output face image

with others from other faces. Then, we randomly select one additional face for each part we want to replace. These faces will be our *source* faces, *i.e.* the faces that will provide the parts to be blended onto the *target*. The selection of the *source* faces takes into account the pose of the selected *target* face. That is, the *source* faces must have a head pose similar to the one of the *target* face.

Before blending parts of the *source* faces onto the *target* face, we first need to align each *source* with the *target*, more specifically, the facial landmarks of each *source* with the landmarks of the *target*. Since we know the position of the 68 landmark points in the faces (step 1 in Fig. 2), we can use this information to warp each *source* to match the *target*. To do so, we use OpenCV to calculate the

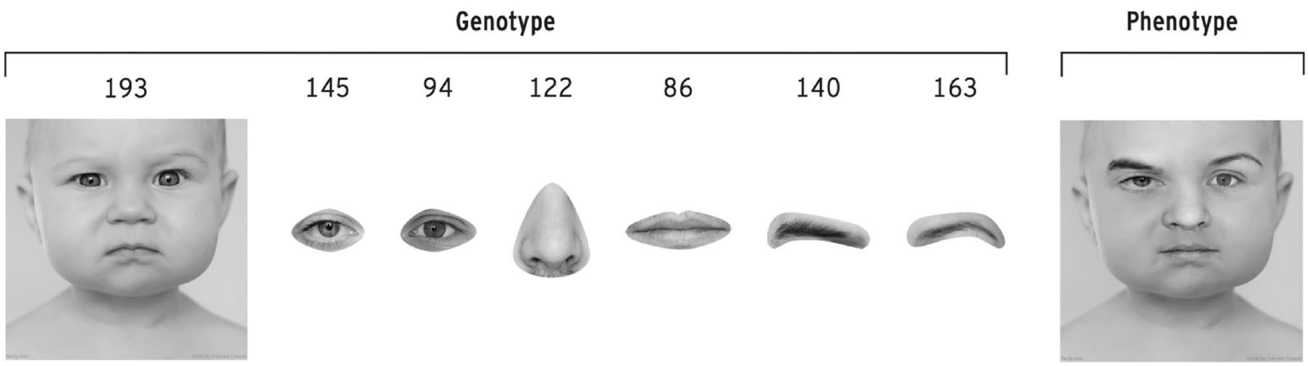

Fig. 3 Genotype and phenotype of an individual. The genotype consists of a tuple of integers (face, left eye, right eye, nose, mouth, left eyebrow, right eyebrow). Each integer encodes an index of an annotated example. The phenotype consists of a composite of the face parts encoded in the genotype

Delaunay triangulation of each face using specific points of its landmarks, plus a few more points (two points on both sides of the nose and six points above the eyebrows) that are calculated in relation to the landmarks (step 2). The Delaunay triangulation of these points allows us to divide each face image into triangles (step 3). The resulting triangles cover corresponding facial regions between the faces (step 4). Therefore, since we know which triangle in each *source* face corresponds to which triangle in the *target* face, the next step is to warp each *source* face triangle to the corresponding *target* face triangle. We use OpenCV to calculate, for each *source* face triangle, the affine transform that maps its vertices to the vertices of the corresponding triangle in the *target* face. Then we use this affine transform to transform the pixels of the *source* face inside the triangle (before transformed) to the *target* face. This results in the *source* faces transformed such that they fully align with the *target* face (step 5).

Once each *source* face is aligned with the *target* face, the next step is to blend them in a natural fashion. We accomplish this by using the seamless cloning algorithm provided by OpenCV, which consists of an implementation of the work by Perez et al. [30]. We use this algorithm to seamlessly copy one region of each warped *source* face and place it on top of the *target* image to produce a new photorealistic face. To delimit the regions of the warped *source* faces that should be blended into the *target* face, we create masks that delimit the different facial parts: left eyebrow, right eyebrow, left eye, right eye, nose, and mouth (step 6). We use specific vertices of the Delaunay triangles to calculate the mask of each facial part. Each mask includes a preset margin, which causes the mask to contain a larger area than the one defined by the vertices used to create it. The purpose of this margin is to expand the area of each shape to cover some skin around each facial part, which is useful for the smooth assemblage of parts to generate new faces. This process results in one image, a composite face, that combines the parts of the different images.

Evolutionary Exploration of Composite Faces

The decomposition processes and assemblage processes allow us to have a system capable of generating faces from parts of different face images. We can see this system as a parametric one, where the input parameters are the faces being recombined, and the output is a new face image that results from that recombination. This way, we can represent each possible recombination of facial parts as a vector of values, identifying the *target* face image and the *source* face images. The value that represents each image corresponds to its index in the containing dataset. With this representation, we have $\#D^n$ different possible faces to generate, where #D is the number of instances in the image dataset and n is the number of face parts. For this particular application involving faces, we would have $n=7$. As the number of #D grows, the number of different faces we can generate grows at a polynomial rate. The face image datasets usually have thousands of examples, so the solution space tends to be large.

Due to the large solution space, there is a need to search for a composite that is suitable for the face detection task. Most face detectors tend to have limited samples of faces, and there have been recent reports that they are even biased. With the generative part of *X-Faces*, we can start with a dataset and expand the number of combinations, but most of them will be repetitive or redundant to have on models that perform face detectors. EC enables the *X-Faces* system to find variations that can be suitable for any given task by providing an appropriate fitness function. Based on the aspects mentioned above, we employ an evolutionary approach to evolve face composites that are not detected by trained face detectors.

We employ a conventional GA using an automatic fitness assignment scheme to explore and choose interesting individuals. In terms of representation, an individual is a set of numerical indexes mapped to a photorealistic part that has been previously collected and annotated on the dataset. Each set of indexes is translated to a composite face, a face made from other faces' parts. Figure 3 explains the genotype of each individual and its corresponding phenotype. Each genotype is mapped into a phenotype creating a composite face where the face parts encoded in the genotype are placed over a *target* face encoded in the genotype (*e.g.* the first gene).

In both iterations [1, 2], fitness is calculated automatically using the confidence value given by the face detectors. The fitness functions used in both iterations were designed to promote composite faces that minimised the face detector response to a point where the face detector could no longer detect the face. This approach for fitness assignment enables the automatic exploration and optimisation of face images that are not detected by the face detectors.

The overall process of generating new face images consists of (i) processing the input dataset of face images to extract data about them; and (ii) creating new faces from the input dataset based on the data extracted. Each new face image consists of an assemblage of facial parts that are inserted into a *target* face. The evolutionary process implemented through a GA empowered the *X-Faces* system's exploratory capabilities by providing means to automatically generate and evaluate suitable solutions drawn from a vast solution space, optimising a given fitness function. In the first work [1], we were able to consistently generate face images that the classifier, used to provide fitness, could not detect as such. From a subjective perspective, the generated images maintained the photorealism of human faces. Moreover, the generated faces revealed uncommon facial combinations that provoked the uncanny valley problem, a phenomenon where an artificial artefact becomes so real that it is unsettling from a human perspective. Based on the findings, in the second iteration [2], we demonstrated that the evolved face composites, when added to the training dataset, improved the performance of the deep learning face detectors that were used (for further information, refer to [2]). The second iteration differs from the first in the following aspects: (i) generation of face composite in an unconstrained matter (*e.g.* different scale variations, lighting conditions, pose); (ii) automatic fitness assignment scheme based on the response of deep face detectors; and (iii) after the evolutionary process occurs, the generated composites are used to retrain the deep face detectors, showing that evolved face composites can augment and improve the dataset and consequently the performance of the detector trained on that expanded dataset.

Portraits of No One: Interacting with Generated Faces

We explored the creative capabilities of the *X-Faces* system through the interactive Media Art installation *Portraits of No One*. This installation employs the generative process described in the previous Section to generate and present photorealistic artificial portraits based on the captured facial features of the audience. This way, the installation is able to invoke the capabilities of AI to generate content that makes people question the veracity of what they are observing. Figure 4 depicts the environment of the installation.

The room of the installation was specifically made for the purpose, 6 m long by 4 m wide. The space of the installation comprises an input area and an output area. The input area, which is located next to the entrance to the room, contains the capturing box attached to one side of a pillar. The capturing box allows users to capture their faces and, this way, to feed the installation. The output area is located at the back of the room, after the input area, and is surrounded by an immersive video projection of twelve meters wide on three walls. This projection contains an array of portraits that are continuously generated by the installation. Each portrait is displayed with a size similar to that of a real face. This area also includes a speaker attached to the ceiling of the room that reproduces an intertwined combination of sounds that are recorded during the interactions of the users with the capturing box. This audio promotes the creation of a more lifelike and reflective environment for the installation with an effect similar to crowd noise, where only a few words are creating a sensation that the portraits are talking with each other. This sonic environment fosters the viewers to engage with the identities embodied by the artificial portraits they are seeing on the walls of the installation.

A typical user interaction with the installation can be described as follows: (i) the user enters the space of the installation; (ii) the user approaches the capturing box and observes her/his face on the screen, in a mirror-like fashion; (iii) if the face is recognised by the system, the button becomes white and blinks fast; otherwise, the button remains red and blinks slowly; (iv) when the button is white, and the reflected image pleases the user, he/she presses the button to capture the face; (v) a countdown of three seconds begins and then the face is captured; (vi) the capturing box pauses for a couple of seconds to avoid frequent captures and the button turns off until a new face can be captured; and, finally, (vii) the user steps into the output area to see new portraits being generated using her/his facial parts.

The installation *Portraits of No One* was a finalist artwork of the *Sonae Media Art Award 2019* and presented

Fig. 4 Environment of the Installation *Portraits of No One*. Photo by José Paulo Ruas/DGPC 2019

in a collective exhibition at the National Museum of Contemporary Art—Museu do Chiado, in Lisbon, Portugal. During the exhibition, from late November 2019 to early February 2020, the capturing box was used about 5.000 times. Figure 5 shows typical portraits generated by the installation.

The creation of this interactive installation involved the development of hardware and software. The computer is the core component of this installation, which is responsible for connecting all hardware and software parts of the installation to manage the inputs (i.e. the user's faces and their interaction sound) and outputs (i.e. the array of portraits and the ambient sound). We also use a microcontroller to allow the computer to communicate with hardware contained in the capturing box and to trigger some software events using this hardware. The following Subsections comprehensively describe the installation's hardware and software.

Hardware

The hardware of the installation is responsible for collecting data from the installation's audience, or users, displaying audiovisual outputs generated by the installation's system, and providing feedback to the users. Figure 6 presents the main hardware components and their data flow.

The users interact with the system using the capturing box. This box contains hardware to collect the necessary data from the users for the installation to operate, i.e. the images of users' faces and the sounds produced by users during their interaction. This capturing box integrates several pieces of hardware,[2] including a video camera to capture the faces, a LED ring that fits around the camera to provide even illumination with few shadows visible in the captured faces, an LCD display to show users a live preview of their faces before capturing, a microphone to record some seconds of audio during each interaction, and a push-button with an RGB LED that allows users to trigger the face capturing while providing colour feedback. The video camera and the LED ring are positioned on the top of the display. The microphone is positioned on one side of the video camera, oriented forward to pick up sounds produced in front of the

[2] More details about the hardware of the installation *Portraits of No One* and its development process is available at https://cdv.dei.uc.pt/portraits-of-no-one/.

Fig. 5 Typical portraits generated in the installation *Portraits of No One*

capturing box. The push-button is placed on the bottom of the screen.

The installation uses three video projectors and one loudspeaker to provide an ever-changing audiovisual environment composed of images of generated portraits and ambient sounds generated from data collected from the audience. The video projectors are attached to the ceiling of the room, each one facing a different wall, to create a single continuous projection. The result is a surrounding video projection of twelve meters wide, on three walls, with a total image size of 5760 by 1080 pixels. This image size allows the presentation of a large array of portraits with good quality. The loudspeaker is also attached to the ceiling of the room, in a central position, facing down. This way, we can fill the room

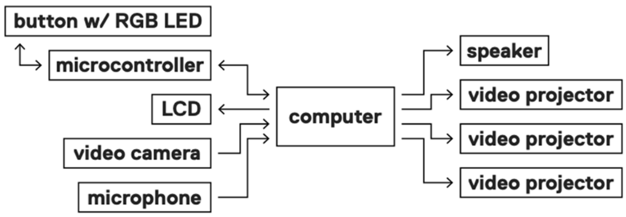

Fig. 6 Diagram of the main hardware components and their data flow. The components on the left are located inside the capturing box, while the ones on the right are attached to the ceiling of the room. The computer is hidden in a space next to the room

with an ambient sound that complements the installation environment.

Software

The installation software is mostly implemented in Java, using the open-source libraries Processing, OpenCV and Dlib. We also use Max for sound recording, processing, and reproduction. The microcontroller that establishes the communication between the computer and other hardware is an Arduino.

A key feature of the installation is the software that implements the real-time detection of the users' faces in the capturing box (see Fig. 7). We resort to off-the-shelf solutions from OpenCV and Dlib libraries to analyse each frame of the video camera feed and detect faces of users, and their facial landmarks, when they are positioned in front of the capturing box. In the context of this work, we require neutral frontal poses for the portraits. This is not only advantageous to the accurate extraction of the facial features but also to ensure visual coherence between different portraits. To achieve this, we implemented a set of detection filters that determine whether a detected face should be considered as valid.

When a valid detection occurs, the system draws a rectangle around the face on the live preview shown on the screen. When more than one valid detection occurs, the system focuses on one face based on two preference criteria: (i) faces horizontally centred in the captured image; and (ii) faces with large bounds. The rectangle that is drawn around the detected face indicates the area of the captured image that will be cropped and used to generate new portraits. The aspect ratio of the crop rectangle is always 3:4 (vertical). The size and position of the rectangle are determined in relation to the detected facial landmarks. This way, we ensure the faces shown in the portraits are always roughly aligned with each other. In addition to drawing a rectangle around the valid face detection, the system also invites the user to capture the face by pressing the push-button.

When the user presses the push-button, the system begins a countdown of three seconds, shown on the top of the screen close to the video camera. When the countdown ends, the system captures the face and saves the related data needed to generate new portraits. At this moment, the system turns off the LED of the push-button, for a couple of seconds, until a new face can be captured. This is part of a delay mechanism that we implemented to avoid bursts of photos.

The installation is continuously generating and displaying new artificial faces from the captured faces of the people who interacted with it. To do so, we employ the *X-Faces* generative system, described earlier in Sect. 3, to create artificial faces from the facial features extracted from the users' faces. This way, images of the captured faces are passed to the *X-Faces* system, which then automatically extracts their facial parts and recombines them to synthesise new ones. In this process, we require the generated faces to satisfy different requirements, namely the matching of facial pose, the scaling and alignment of the facial parts. We are able to generate and display a grid with 225 portraits on the walls of the installation room. Each

Fig. 7 Photographs of the installation *Portraits of No One* showing someone interacting with the capturing box. On the background of the photographs, one can see walls crowded with faces generated and displayed by the installation system. Photo by José Paulo Ruas / DGPC 2019

Fig. 8 Photograph of the installation *Portraits of No One* with someone observing the walls crowded with faces. Photo by José Paulo Ruas/DGPC 2019

Fig. 9 Photograph of the installation *Portraits of No One* with people observing the walls crowded with faces. On the right, one can see the capturing box. Photo by José Paulo Ruas/DGPC 2019

portrait being projected on the walls has a random lifespan between 10 to 20 s. After this lifespan is over, the *X-Faces* system creates another artificial portrait, and replaces the old portrait by the new one. This process is repeated whenever the lifespan of a portrait is over. The set of face images used by the *X-Faces* system to generate the portraits is continuously updated with new faces captured from the users. This enables the installation to quickly adapt to new face images captured by the users, so whenever a new face is captured, its facial parts will contaminate the artificial portraits shown on the walls. The smooth transition between portraits makes users aware of the subtle changes on the walls of the installation that are populated with numerous portraits. Also, the installation implements a mechanism that places a few real faces, selected at random, in the array of artificial faces to make users question the veracity of the portraits they are observing. Figures 8 and 9 show the environment of the installation with users observing the walls populated with numerous portraits, real and artificial.

During the capture of a face, the installation also records a sound sample from a few seconds before the face being captured to a few seconds later. These sound recordings, collected from all users who captured their faces, are then used to create an ambient sound that is reproduced using the loudspeaker. The sound is generated by intertwining a total of twelve samples randomly selected from the set of all recorded samples of users interaction. Most of the interactions occur in silence, so we decided to play a background sound sample, with a volume lower than the main samples. This background sound consists of a sample, in loop, of crowd noise that was recorded previously. The frequency and pitch of the recorded sound samples are digitally attenuated and made homogeneous through single-pole low-pass sound filters. Generally, the users inside the room mostly hear imperceptible sounds resulting from the mixing of captured noise. However, sometimes, they can recognise some words or even sentences. This creates an environment where the users have the perception that the people whose faces are projected on the walls are talking to each other.

Conclusions and Future Work

In this paper, we presented our work on the computational creation of new photorealistic faces images based on the recombination, or exchange, of the facial parts from existing faces. We described the *X-Faces* system which is the basis of this work and employs EC and ML techniques to automatically evolve face images. This system implements a GA to automatically recombine facial parts from different faces and, this way, create new face composites. The evolutionary process of *X-Faces* is fully automatic, thus enabling the comprehensive exploration of face images and optimisation towards an objective function, unveiling suitable and diverse photorealistic face images. The photorealism of the resulting images not only put state-of-the-art face classifiers to the test but also make people question their veracity. Also, the diversity of images that are generated is notable. All these features reveal the great potential in different areas, namely in the expansion of face detection datasets through the generation of new examples and the search for examples that exploit the vulnerabilities of the face detectors [1, 2], along with Media Art through the generation of artificial portraits based on the audience faces [3].

We explored the *X-Faces* system with the creation of the interactive Media Art installation *Portraits of No One* to create an ever-changing audiovisual environment composed of photorealistic human portraits constructed with the facial parts captured from its audience and an ambient sound generated by intertwining sound samples of users' interactions. Whenever users give their faces to the installation, they will immediately see their facial parts contaminating the portraits that are displayed on the walls. As the installation takes advantage of users' facial traits to feed itself and grow, the users' facial parts become part of the artwork. This aspect of the installation creates an engaging experience for users by allowing them to directly participate in the generation of new artificial identities in the form of portraits.

The artificial portraits generated by the installation encourage critical thinking about how recent technological advances are changing our relationship with others and the world. The photorealism of the generated portraits places the audience at the borderline between the real and artificial. As a result, users tend to question the veracity of the faces that they are seeing. Besides, this opens a window of opportunity to discuss the veracity of the images that we see in other contexts and environments and whether we should consider them as unquestionable proofs of the truth. In the installation, one can also observe varied relationships and interactions between users, as well as between the users and the artificial identities generated by the installation. Most of the users see the portraits on the walls as other people, starting to judge and talk about their visual appearance. However, when they recognise themselves, or a familiar face part, in a portrait of no one, they change the behaviour. At this moment, users engender empathy by this artificial person and by the others who, like them, are blended to generate their portraits. The presented installation was designed for the Sonae Media Art Award 2019 and exhibited at the National Museum of Contemporary Art, in Lisbon, Portugal.

Future work will consider: (i) using EC to enhance the interaction of the installation, for example, evolve and display portraits that are not detected by face detectors, or evolve portraits that are similar to a given user's face but made only of facial parts of other people, or to evolve faces with similar expressions, or emotions, as the last face captured; (ii) exploring other approaches to display the generated portraits; (iii) exploring the animation of generated portraits and their reactiveness to external data such as the presence of people in the room; (iv) experimenting with the blending of audio recorded from the audience with computationally generated sounds; (v) setting up *Portraits of No One* in other locations; and (vi) extending the idea behind the installation beyond an exhibition room, for example, by making the system online.

Acknowledgements The third author is funded by Foundation for Science and Technology (FCT) under the Grant SFRH/BD/132728/2017. This work is supported by national funds through the FCT, Portugal, within the scope of the project UID/CEC/00326/2019 and by European Social Fund, through the Regional Operational Program Centro 2020.

Declearations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1 Correia J, Martins T, Martins P, Machado P. X-Faces: the eXploit is out there. In: Pachet F, Cardoso A, Corruble V, Ghedini F, (eds) Proceedings of the Seventh International Conference on Computational Creativity (ICCC 2016), pp 164–182. Paris, France. 2016.
2 Correia J, Martins T, Machado P. Evolutionary data augmentation in deep face detection. In: Proceedings of the Genetic and evolutionary computation conference companion (GECCO '19). Association for Computing Machinery, New York, NY, USA, pp. 163–164. 2019. https://doi.org/10.1145/3319619.3322053.
3. Martins T, Correia J, Rebelo S, Bicker J, Machado P. Portraits of no one: an interactive installation. In: Romero J, Ekárt A, Martins T, Correia J, (eds) Artificial intelligence in music, sound, art and design. Cham: Springer International Publishing, pp. 104–117. 2020. https://doi.org/10.1007/978-3-030-43859-3_8.
4. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks Commun. ACM. 2017;60(6):84–6. https://doi.org/10.1145/3065386.
5. Masi I, Trãn AT, Hassner T, Leksut JT, Medioni G. Do we really need to collect millions of faces for effective face recognition? In: Leibe B, Matas J, Sebe N, Welling M, (eds) Computer vision—ECCV 2016. Lecture notes in computer science, Cham: Springer International Publishing, pp. 579–596. 2016. https://doi.org/10.1007/978-3-319-46454-1_35.
6. Jiang D, Hu Y, Yan S, Zhang L, Zhang H, Gao W. Efficient 3D reconstruction for face recognition. Pattern Recogn. 2005;38(6):787–11. https://doi.org/10.1016/j.patcog.2004.11.004.
7. Mohammadzade H, Hatzinakos D. Projection into expression subspaces for face recognition from single sample per person. IEEE Trans Affect Comput. 2013;4(1):69–13. https://doi.org/10.1109/T-AFFC.2012.30.
8. Seyyedsalehi SZ, Seyyedsalehi SA. Simultaneous learning of nonlinear manifolds based on the bottleneck neural network. Neural Process Lett. 2014;40(2):191–18. https://doi.org/10.1007/s11063-013-9322-9.
9. Nirkin Y, Masi I, Tran Tuan A, Hassner T, Medioni G. On face segmentation, face swapping, and face perception. In: 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), pp. 98–105. 2018. https://doi.org/10.1109/FG.2018.00024.
10. Lv J-J, Shao X, Huang J-S, Zhou X-D, Zhou X. Data augmentation for face recognition. Neurocomputing. 2017;230:184–12. https://doi.org/10.1016/j.neucom.2016.12.025.
11. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 4396–4405. 2019. https://doi.org/10.1109/CVPR.2019.00453.
12. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T. Analyzing and improving the image quality of StyleGAN. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 8107–8116. 2020. https://doi.org/10.1109/CVPR42600.2020.00813.
13. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T. This person does not exist. 2020. https://thispersondoesnotexist.com/. Accessed 28 Jul 2020.
14. Chen J, Wang R, Yan S, Shan S, Chen X, Gao W. Enhancing human face detection by resampling examples through manifolds. IEEE Trans Syst Man Cybern Part A Syst Hum. 2007;37(6):1017–11. https://doi.org/10.1109/TSMCA.2007.906575.
15. Machado P, Correia J, Romero J. Improving face detection. In: Moraglio A, Silva S, Krawiec K, Machado P, Cotta C (eds) In: 15th european conference on genetic programming (EuroGP), pp. 73–84. 2012. https://doi.org/10.1007/978-3-642-29139-5_7.
16. Shaw J. In: Video narcissus. 1987. https://www.jeffreyshawcompendium.com/portfolio/video-narcissus/. Accessed 28 Jul 2020.
17. Kac E. Interfaces 1990. http://www.ekac.org/sstv.html. Accessed 28 Jul 2020.
18. Biggs S. Solitary. 1992. http://littlepig.org.uk/installations/solitary/solitary.htm. Accessed 28 Jul 2020.
19. Levin G, Lieberman Z. Reface—Portrait Sequencer. 2007. http://www.flong.com/projects/reface/. Accessed 28 Jul 2020.
20. Howorka S. Average Face Mirror. 2015. http://www.sarahhoworka.at/projects/average-face-mirror. Accessed 28 Jul 2020.
21. Al Tawil S. IDEMixer. 2019. http://samehaltawil.com/portfolio/idemixer/. Accessed 28 Jul 2020.
22. Suhr HC, You I. We: exploring interactive multimedia performance. In: Proceedings of the 27th ACM international conference on multimedia. New York: ACM, pp. 1147–1155. 2019. https://doi.org/10.1145/3343031.3355706.
23. McDonald K. Sharing Faces. 2013. https://github.com/kylemcdonald/sharingfaces. Accessed 28 Jul 2020.
24. Lancel K, Maat H, Brazier F. Saving face: playful design for social engagement, in public smart city spaces. In: Brooks AL, Brooks E, Sylla C (eds) Interactivity, game creation, design, learning, and innovation. Cham: Springer International Publishing, pp. 296–305. 2019. https://doi.org/10.1007/978-3-030-06134-0_34.
25. Lozano-Hemmer R. Level of Confidence. 2015. http://www.lozano-hemmer.com/artworks/level_of_confidence.php. Accessed 28 Jul 2020.
26. Daniele A. This is not private. 2015. http://www.letitbrain.it/letitbrain/?port=this-is-not-private. Accessed 28 Jul 2020.
27. Tyka M. Portraits of imaginary people. 2017: http://www.miketyka.com/. Accessed 28 Jul 2020.
28. Moura JM, Ferreira-Lopes P. Generative face from random data, on "how computers imagine humans." In: Proceedings of the 8th international conference on digital arts, pp. 85–91. 2017. https://doi.org/10.1145/3106548.3106605.
29. Klingemann M. Neural Glitch. 2018. http://underdestruction.com/2018/10/28/neural-glitch/. Accessed 28 Jul 2020.
30. Pérez P, Gangnet M, Blake A. Poisson image editing. ACM Trans Graph (TOG). 2003. https://doi.org/10.1145/882262.882269