Check for
updates

The 4D Space-Time Dimensions of Facial Perception

Adelaide L. Burt* and David P. Crewther

Centre for Human Psychopharmacology, Swinburne University of Technology, Melbourne, VIC, Australia

Facial information is a powerful channel for human-to-human communication. Characteristically, faces can be defined as biological objects that are four-dimensional (4D) patterns, whereby they have concurrently a spatial structure and surface as well as temporal dynamics. The spatial characteristics of facial objects contain a volume and surface in three dimensions (3D), namely breadth, height and importantly, depth. The temporal properties of facial objects are defined by how a 3D facial structure and surface evolves dynamically over time; where time is referred to as the fourth dimension (4D). Our entire perception of another's face, whether it be social, affective or cognitive perceptions, is therefore built on a combination of 3D and 4D visual cues. Counterintuitively, over the past few decades of experimental research in psychology, facial stimuli have largely been captured, reproduced and presented to participants with two dimensions (2D), while remaining largely static. The following review aims to advance and update facial researchers, on the recent revolution in computer-generated, realistic 4D facial models produced from real-life human subjects. We delve in-depth to summarize recent studies which have utilized facial stimuli that possess 3D structural and surface cues (geometry, surface and depth) and 4D temporal cues (3D structure + dynamic viewpoint and movement). In sum, we have found that higher-order perceptions such as identity, gender, ethnicity, emotion and personality, are critically influenced by 4D characteristics. In future, it is recommended that facial stimuli incorporate the 4D space-time perspective with the proposed time-resolved methods.

THE 4D SPACE-TIME PERSPECTIVE OF FACIAL PERCEPTION

What does our visual-perceptual system "*see*" when we view a face? Prima facie, we see a *three-dimensional* form, evolving over time, that enables multifaceted social, cognitive and affective perceptions of one another. From cognitive information such as identity and recognition; to detecting complex patterns of speech or emotion; to gleaning important social cues such as gender, ethnicity, age or health. The human face has been defined by other authors as a multi-dimensional pattern which evolves over time (Zhang et al., 2003; Liu et al., 2012; Marcolin and Vezzetti, 2017). This multi-dimensional, complex pattern is also highly individualized; varying from person to person. Thus, each human face possesses concurrently a unique volumetric structure and surface pattern in three dimensions (or 3D) and a temporal pattern across time in four dimensions (or 4D). The 3D volumetric structure or form of human facial features contains spatial dimensions of breadth, height and width, combined with a unique surface pattern. The 4D temporal pattern of the

human face encompasses all dynamic movement and changes to this 3D spatial form that evolve with time. Little is known however, about how we detect and perceive the 3D spatial or 4D temporal patterns of human faces on two-dimensional computer screens.

Indeed, over the past few decades, the field of facial perception has been largely built upon 2D static images and photographs displayed on screens, which are computationally reproduced with two-dimensions of width and height (for a review see Matsumoto, 1988; Lundqvist et al., 1998; Beaupré et al., 2000; Tottenham et al., 2009; Langner et al., 2010; Krumhuber et al., 2017). 2D dynamic or video-recorded stimuli have likewise made a valuable contribution; demonstrating facial recognition may be enhanced with motion (Kanade et al., 2000; Wallhoff, 2004; Pantic et al., 2005; Lucey et al., 2010; Krumhuber et al., 2013, 2017; Dobs et al., 2018) although this issue remains unresolved (Christie and Bruce, 1998; Krumhuber et al., 2013, 2017). 2D stimuli demonstrate robust and reliable human facial recognition performance across behavioral studies of face perception (Beaupré et al., 2000; Goeleven et al., 2008; Rossion and Michel, 2018). These stimuli have helped characterize individual abilities in facial recognition- from face experts (Tanaka, 2001; Russell et al., 2009) to describing a continuum of individual differences in face recognition abilities (Wilmer et al., 2010; Dennett et al., 2012; Rhodes et al., 2014); and to describe the challenges of face perception exhibited in clinical or atypical populations (Behrmann and Avidan, 2005; Halliday et al., 2014). Some of the most influential neuroimaging models of facial processing have employed 2D faces, discovering spatio-temporal networks and regions of activation (Haxby et al., 2000; O'Toole et al., 2002; Kanwisher and Yovel, 2006). In addition, significant practical advantages are afforded by these stimuli such as; the ability to more precisely control and manipulate stimuli; reduced data size and reduced computational load or complexity (Wallbott and Scherer, 1986; Motley and Camden, 1988; Bruce et al., 1993; Russell, 1994; Elfenbein and Ambady, 2002). Overall, 2D faces are a valuable source of experimental stimuli as they offer a robust level of control, reproducibility and replication.

In recent years, three-dimensional virtual representations of the human face that are more concurrent with our everyday social experiences, are beginning to be recognized as essential stimuli (Blauch and Behrmann, 2019; Zhan et al., 2019). While these newer technologies are not being utilized in psychology, they offer several advantages as naturalistic, ecologically valid representations of human faces. Historically however, creating a simulation of a biological human face on a two-dimensional computer screen in 3D and 4D formats, has proven difficult. Simulating concurrently the unique 3D structure of a face, with the added dimension of movement over time, has made computational recovery a challenging field. More recently, computer science has seen rapid advancement in facial reproduction technology in 3D and 4D formats, enabled through development within gaming and virtual-reality (Blascovich et al., 2002; Bailenson et al., 2004; Bülthoff et al., 2018; Chen et al., 2018; Smith and Neff, 2018). In fact, computer-generated objects, human bodies and faces give us a new sense of virtual dynamics and interaction, producing the visual illusion of being able to be moved, rotated or grasped in a virtual space (Freud et al.,

2018a; Chessa et al., 2019; Mendes et al., 2018; Sasaoka et al., 2019). Evidence is beginning to emerge that it is also critical to actively explore 3D spaces when viewing faces (Bailenson et al., 2004; Bülthoff et al., 2018). Thus, while in the past we have only been able to experiment with facial perception from a largely two-dimensional or static perspective, our overarching goal is to open-up other dimensions of these newer stimuli for researchers. It is likely by viewing human facial modeling from a space-time (4D) perspective, that much will be learned about how naturalistic and ecologically valid facial perception is performed in our interactions with each other.

HOW ARE 3D AND 4D FACIAL DATABASES CAPTURED AND REPRODUCED INTO A DIGITAL STIMULUS?

How do we make _you really_ look like _you_ on a digital screen? Facial databases, comprise sets of facial stimuli which capture and reproduce multiple human subjects or individuals into a digital format. Virtual humans, often termed "virtual busts or avatars," represent a burgeoning area of computational development (Blascovich et al., 2002; Bailenson et al., 2004). While "virtual avatars" are being presented to the viewer on a two-dimensional flat screen, they elicit the illusion of being three-dimensional. These virtual computer-generated human bodies possess impressive anatomical-accuracy, including of both individual body parts and holistic bodies (Allen et al., 2003; Cofer et al., 2010; Ramakrishna et al., 2012; Pujades et al., 2019). Likewise, the reproduction of biological human faces into a virtual model on-screen, has been achieving a sense of realism through the extensive development of several teams (Wang et al., 2004; Vlasic, 2005; Gong et al., 2009; Weise et al., 2009; Sandbach et al., 2012; Bouaziz et al., 2013; Chen et al., 2013; Garrido et al., 2013; Li et al., 2013; Shi et al., 2014; Suwajanakorn et al., 2014; Zhang et al., 2014; Cao et al., 2015; Hsieh et al., 2015; Thies, 2015; Saito et al., 2016; Jeni et al., 2017; Zollhöfer et al., 2018). Digitized facial databases are now available as stimuli across several media formats; including 3D static models or images (Yin et al., 2006; Saito et al., 2016) 4D dynamic movies or videos (Wang et al., 2004; Vlasic, 2005; Gong et al., 2009; Weise et al., 2009; Garrido et al., 2013; Shi et al., 2014; Suwajanakorn et al., 2014; Zhang et al., 2014; Cao et al., 2015; Thies, 2015; Jeni et al., 2017) and as 3D/4D interactive stimuli, such as in virtual reality (Bouaziz et al., 2013) or with eye-gaze tracking (Thies, 2017; Chen et al., 2018).

The capture techniques, reconstruction algorithms and 3D software packages used to develop and display a life-like 3D or 4D human face into a virtual form is a highly individualistic, complex process and is achieved differently by each database. Thus, the different capture and reproduction techniques are comprehensively surveyed and reviewed elsewhere (Gaeta et al., 2006; Beeler et al., 2010; Sandbach et al., 2012; Patil et al., 2015; Feng and Kittler, 2018; Zollhöfer et al., 2018). Briefly however, the databases above demonstrate methodologically how to produce an anatomical head model through mesh, polygon or geometric

structures, using width x height x depth coordinates (Egger et al., 2019). They demonstrate the unique and highly technical challenge of producing surface rendering or skin rendering with shape deformations that are anatomically precise, while capturing environmental lighting, reflectance, pigmentation, shadow and texture (Egger et al., 2019). The 4D databases also exhibit how animation and modeling are combined to produce anatomically accurate rigid and non-rigid head and facial movements, whilst maintaining 3D shape and surface properties (Egger et al., 2019). For a more detailed overview of how facial information is captured with lighting or viewpoint camera arrays within these databases, see **Figure 1**. To see an illustration of the workflow process of 3D modeling, rendering and animation for human faces and its level of complexity and limitations, see Marschner et al. (2000), Murphy and Leopold (2019) for primate faces. To review state-of-the-art advancements over the past 20 of development, as well as the more technical challenges that remain in the field producing 3D animated faces see the comprehensive review by Egger et al. (2019).

Of note, is that the reproduction of 3D facial models still present challenges in stimuli recreation (Marschner et al., 2000; Egger et al., 2019). Further challenges in implementing 3D or 4D facial models are also unique to research fields such as psychology and neuroscience. Faces are a unique, special class of images in human vision, where deviations from anatomically accurate geometry, surfaces or movement can be described as incorrect or "uncanny" by human observers (Marschner et al., 2000; Kätsyri, 2006; Seyama and Nagayama, 2007). Therefore, participant viewing conditions must be well-controlled to ensure models are presented at the same viewing distance and settings (e.g., environment lighting), as original development (Marschner et al., 2000). In addition, future development of 3D or 4D databases for experimental designs, can further improve stimuli by being built as full-volumes and not surfaces. For example, the database presented in **Figure 2B**, experiences a loss of data from the posterior viewpoint of the head, which may impede studies examining VR environments (Bulthoff and Edelman, 1992; Stolz et al., 2019; Lamberti et al., 2020) or multiple viewpoints of a facial stimulus (Bülthoff et al., 2018; Abudarham et al., 2019; Zhan et al., 2019). Overall, despite some remaining challenges, 3D and 4D facial capture and reconstruction has demonstrated that the unique facial features belonging to you and me, such as the cheeks, eyes and nose, can simulated on-screen in a highly realistic manner (see **Figure 2**).

HOW POPULAR ARE 3D AND 4D FACIAL STIMULI?

An initial quantification of the most widely cited 3D stimuli, suggests the static 3D facial database *BU-3DFE* by Yin et al. (2006) and the Bosphorus *3D* database by Savran et al. (2008) are at the forefront. Similarly, the dynamic video-recorded BU4D-S 4D database produced by Zhang et al. (2013, 2014) is the leading 4D database in citation count (**Figure 3**). It is worth highlighting that a selection of 3D and 4D facial databases, have been developed with specific applications across disciplines. Facial identification

and recognition studies, commonly referred to as biometrics, may benefit from the databases developed by Benedikt et al. (2010); Feng et al. (2014), and Zhang and Fisher (2019). A relatively large corpus of datasets has been developed to present stimuli as three-dimensional forms or structures on-screen (built with breadth, height and depth), as static emotional expressions (*FIDENTIS-3D*; Urbanová et al., 2018; *Bosphorus*; Savran et al., 2008; *BU-3DFE*; Yin et al., 2006; *ICT-3DRFE*; Stratou et al., 2011; *ND-2006*; Faltemier et al., 2008; *Casia 3D*; Zhong et al., 2007; *The York 3D*; Heseltine et al., 2008; *GAVDB*; Moreno, 2004; *Texas 3DFRD*; Gupta et al., 2010). To establish 4D dynamical facial stimuli, faces have also been video-recorded and re-built with naturalistic spontaneous or acted emotions; as dynamical 4D objects that produce movement over a series of frames (Gur et al., 2002; *Cam3D*; Mahmoud et al., 2011; *BU-4DFE*; Yin et al., 2006; *BU4D-S*; Zhang et al., 2014; *4DFAB*; Cheng et al., 2018). Studies of facial perceptions which require precise control and manipulation of artificial motion properties, may benefit from the databases developed for morphing paradigms (*D3DFACS*; Cosker et al., 2011; *SIC DB*; Beumier and Acheroy, 1999). In sum, more extensive surveys of 3D and 4D facial databases are available; whereby facial researchers can select stimuli from a range of available databases (*for a review see* Weber et al., 2018; Zhou and Xiao, 2018; Sadhya and Singh, 2019; Sandbach et al., 2012; Smeets et al., 2012). For a comprehensive survey of the publicly facial databases with both 2D and 3D stimuli, see Sadhya and Singh (2019).

At first glance, the citation count pictured in **Figure 4** appears impressive. Yet, we observed that this citation count is largely driven by the development of computer science and engineering; including capturing, reproducing, recognition and tracking of 3D and 4D facial technology. Thus we examined logarithmic growth across academic fields below (**Figure 5**). Notably psychology and neuroscience articles associated with terms such as "3D or 4D faces," in titles, abstracts and keywords, demonstrate reduced logarithmic growth compared to other academic fields. It appears that psychologists and neuroscientists continue to reference more traditional, two-dimensional facial databases in their research. Although not exhaustive, some examples of the most popular, highly cited repositories of 2D image-based facial sets across psychology include; "*The Montreal Set of Face Displays*," (Beaupré et al., 2000) "*The Radboud Face Set*," (Langner et al., 2010) "*The Karolinska Directed Faces*," (Lundqvist et al., 1998) "*The Japanese-Caucasian Facial Affect Set*," (Matsumoto, 1988) and "*The NimStim Set of Facial Expression*s," (Tottenham et al., 2009). More recently, standardized stimulus sets of 2D video-recorded faces have been developed; the "*Cohn-Kanade (CK)*," (Kanade et al., 2000) "*The Extended Cohn-Kanade (CK+) Dataset*," (Lucey et al., 2010) "*Face and Recognition Network-Group*," (Wallhoff, 2004) "*MMI face database*," (Pantic et al., 2005). For a comparative review of the most-cited 2D static and dynamic face stimuli utilized by the research community, please see the review provided by Krumhuber et al. (2017).

Recent experimental studies utilizing 4D dynamical faces, however, have raised issue with whether the human face should be reduced on-screen to a 2D or static pattern (Elfenbein and Ambady, 2002). While these arguments have existed for a lengthy
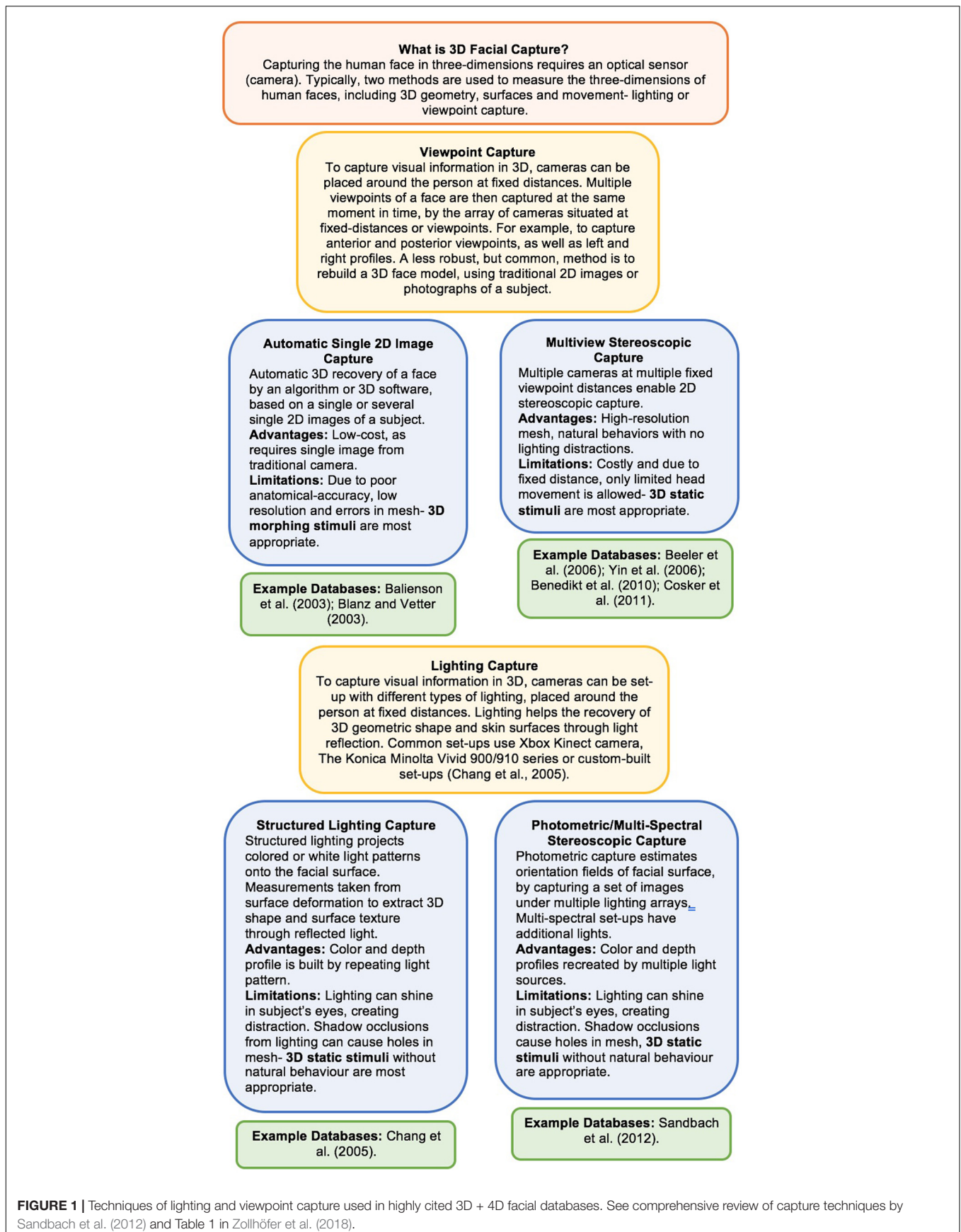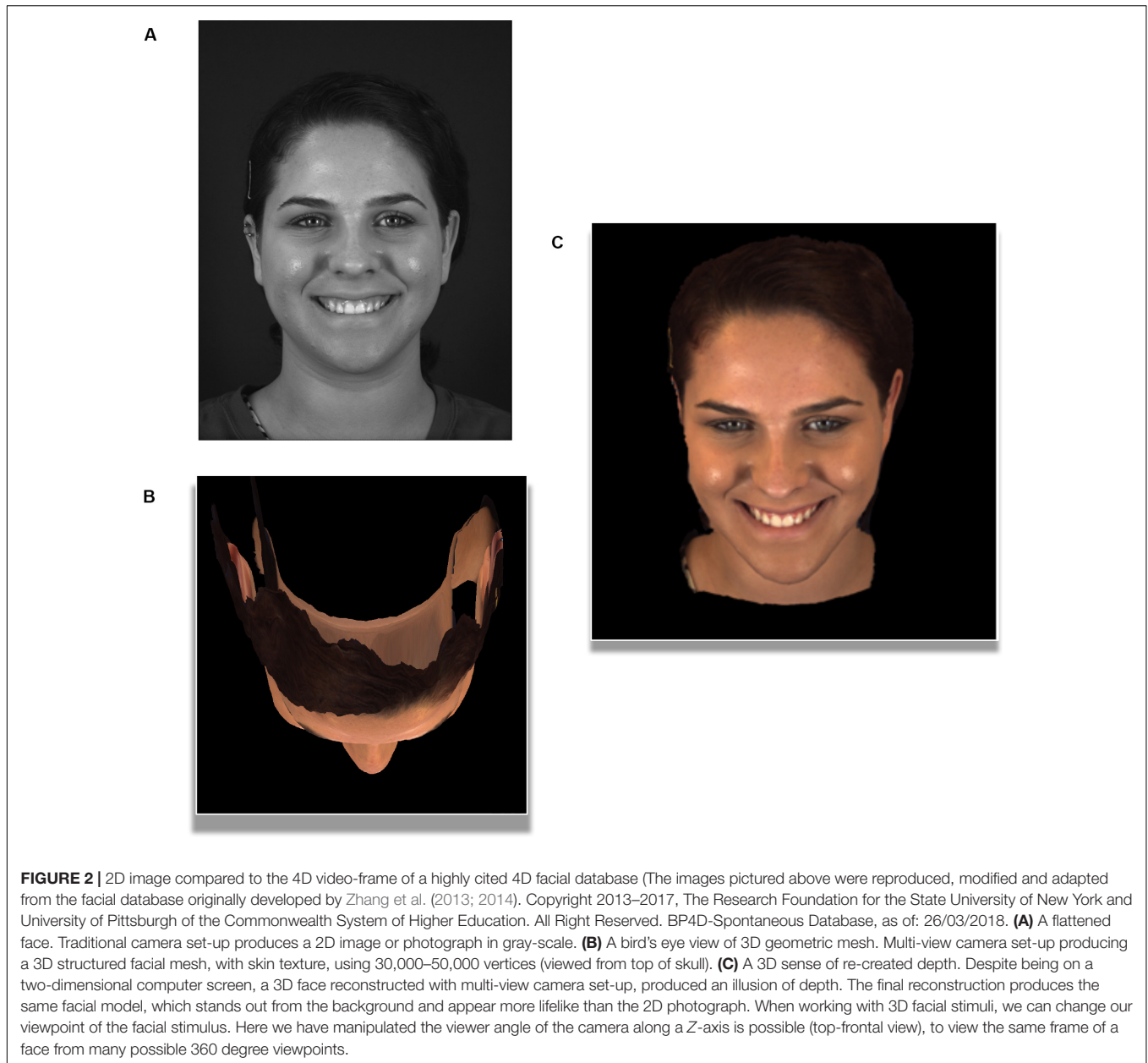
What is 3D Facial Capture?
Capturing the human face in three-dimensions requires an optical sensor (camera). Typically, two methods are used to measure the three-dimensions of human faces, including 3D geometry, surfaces and movement- lighting or viewpoint capture.

Viewpoint Capture
To capture visual information in 3D, cameras can be placed around the person at fixed distances. Multiple viewpoints of a face are then captured at the same moment in time, by the array of cameras situated at fixed-distances or viewpoints. For example, to capture anterior and posterior viewpoints, as well as left and right profiles. A less robust, but common, method is to rebuild a 3D face model, using traditional 2D images or photographs of a subject.

Automatic Single 2D Image Capture
Automatic 3D recovery of a face by an algorithm or 3D software, based on a single or several single 2D images of a subject.
Advantages: Low-cost, as requires single image from traditional camera.
Limitations: Due to poor anatomical-accuracy, low resolution and errors in mesh- **3D morphing stimuli** are most appropriate.

Example Databases: Balienson et al. (2003); Blanz and Vetter (2003).

Multiview Stereoscopic Capture
Multiple cameras at multiple fixed viewpoint distances enable 2D stereoscopic capture.
Advantages: High-resolution mesh, natural behaviors with no lighting distractions.
Limitations: Costly and due to fixed distance, only limited head movement is allowed- **3D static stimuli** are most appropriate.

Example Databases: Beeler et al. (2006); Yin et al. (2006); Benedikt et al. (2010); Cosker et al. (2011).

Lighting Capture
To capture visual information in 3D, cameras can be set-up with different types of lighting, placed around the person at fixed distances. Lighting helps the recovery of 3D geometric shape and skin surfaces through light reflection. Common set-ups use Xbox Kinect camera, The Konica Minolta Vivid 900/910 series or custom-built set-ups (Chang et al., 2005).

Structured Lighting Capture
Structured lighting projects colored or white light patterns onto the facial surface. Measurements taken from surface deformation to extract 3D shape and surface texture through reflected light.
Advantages: Color and depth profile is built by repeating light pattern.
Limitations: Lighting can shine in subject's eyes, creating distraction. Shadow occlusions from lighting can cause holes in mesh- **3D static stimuli** without natural behaviour are most appropriate.

Example Databases: Chang et al. (2005).

Photometric/Multi-Spectral Stereoscopic Capture
Photometric capture estimates orientation fields of facial surface, by capturing a set of images under multiple lighting arrays. Multi-spectral set-ups have additional lights.
Advantages: Color and depth profiles recreated by multiple light sources.
Limitations: Lighting can shine in subject's eyes, creating distraction. Shadow occlusions cause holes in mesh, **3D static stimuli** without natural behaviour are appropriate.

Example Databases: Sandbach et al. (2012).

FIGURE 1 | Techniques of lighting and viewpoint capture used in highly cited 3D + 4D facial databases. See comprehensive review of capture techniques by Sandbach et al. (2012) and Table 1 in Zollhöfer et al. (2018).

FIGURE 2 | 2D image compared to the 4D video-frame of a highly cited 4D facial database (The images pictured above were reproduced, modified and adapted from the facial database originally developed by Zhang et al. (2013; 2014). Copyright 2013–2017, The Research Foundation for the State University of New York and University of Pittsburgh of the Commonwealth System of Higher Education. All Right Reserved. BP4D-Spontaneous Database, as of: 26/03/2018. **(A)** A flattened face. Traditional camera set-up produces a 2D image or photograph in gray-scale. **(B)** A bird's eye view of 3D geometric mesh. Multi-view camera set-up producing a 3D structured facial mesh, with skin texture, using 30,000–50,000 vertices (viewed from top of skull). **(C)** A 3D sense of re-created depth. Despite being on a two-dimensional computer screen, a 3D face reconstructed with multi-view camera set-up, produced an illusion of depth. The final reconstruction produces the same facial model, which stands out from the background and appear more lifelike than the 2D photograph. When working with 3D facial stimuli, we can change our viewpoint of the facial stimulus. Here we have manipulated the viewer angle of the camera along a Z-axis is possible (top-frontal view), to view the same frame of a face from many possible 360 degree viewpoints.

period within the literature, the recent cultural shift in gaming and virtual-reality has enabled us to better envision "*virtual-facial realism*." It is timely, therefore, to review what 3D and 4D dynamic face stimuli can offer by opening other dimensions of facial stimuli for researchers in all fields.

WHAT IS THE 3D SPATIAL PATTERN OF THE HUMAN FACE?

The Influence of 3D Volume and Surfaces in Building Facial Perception

Despite experimental stimuli being displayed on a two-dimensional, flat screen such as a computer monitor, the human visual system perceives 3D structure remarkably efficiently. Three-dimensional objects or models rely upon a geometric structural and surface model presenting accurate reconstruction of breadth, height and *importantly depth*, while still being presented on a two-dimensional screen (Hallinan et al., 1999; Marschner et al., 2000; Egger et al., 2019). Representing detailed facial tissue, bone structure and the musculature system, which make up the human face, has not been an easy feat, yet is now computationally achievable (Zhang et al., 2003; Ekrami et al., 2018). Moreover, these 3D facial models are revealing key differences about both individual recognition, as well as categorization or classification of human faces as objects.

To disseminate which visual elements of 3D geometry contribute to facial recognition in humans, psychophysical
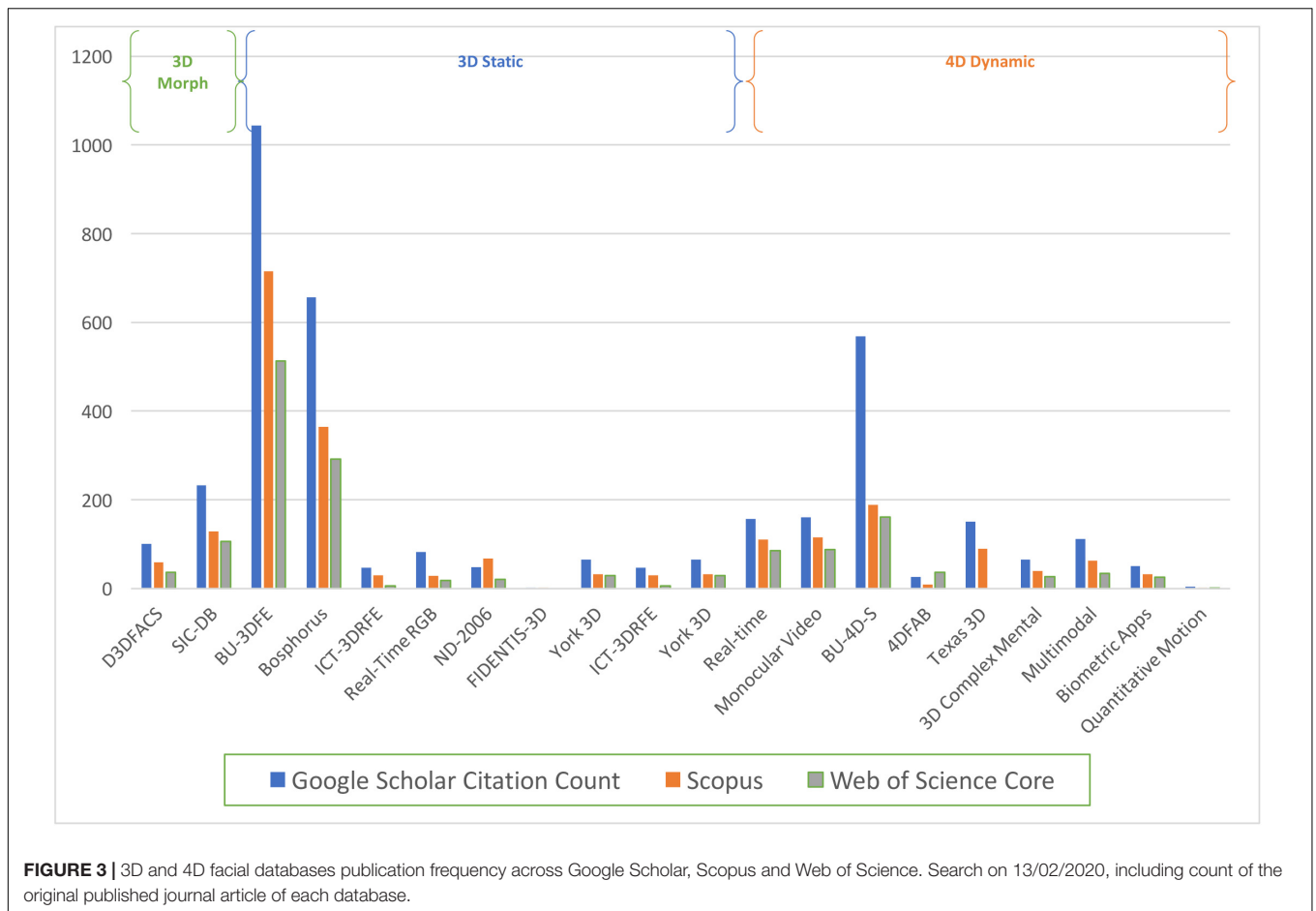
FIGURE 3 | 3D and 4D facial databases publication frequency across Google Scholar, Scopus and Web of Science. Search on 13/02/2020, including count of the original published journal article of each database.

studies have begun to examine volume and curvilinear shape (Hallinan et al., 1999; Chelnokova and Laeng, 2011; Jones et al., 2012; Eng et al., 2017). Psychophysical research comparing 2D and stereoscopic 3D face stimuli are revealing some potential key differences in how we detect and perceive 3D models. Stereoscopic stimuli are used to develop 3D faces from 2D images paired with red-cyan stereoscopic glasses, which elicit depth through binocular disparity cues (Schwaninger and Yang, 2011; Häkkinen et al., 2008; Lambooij et al., 2011). Using these methods, Chelnokova and Laeng (2011) have interpreted the increased fixation time spent on cheeks and noses observed in their study during the 3D viewing condition, as a preference for the more volumetric properties of faces. In the case of unfamiliar faces, recognition of identities were performed better with upright 3D stereoscopic faces, compared to matched 2D stimuli (Chelnokova and Laeng, 2011; Eng et al., 2017). Comparatively, inverted faces elicited no difference in identity recognition between 2D or 3D stereoscopic form, suggesting this is an effect observed using a holistic face template (Eng et al., 2017). Further, reaction times appear significantly faster for identifying 3D upright compared to 3D inverted faces, suggesting it is more efficient to perceive upright faces (Eng et al., 2017). Reaction times to 3D stereoscopic stimuli however, appear to be generally slower than 2D matched-stimuli (Chelnokova

and Laeng, 2011; Eng et al., 2017). Thus, these preliminary studies present initial psychophysical evidence that upright 3D stereoscopic information contained within a face, may provide better accuracy in identification than a matched 2D image. Albeit, with a slower response. The implications of these studies suggest that the structural facial features associated with 3D geometric volume, such as cheeks and noses, may add a critical element during naturalistic facial processing. Overall, these studies provide preliminary psychophysical evidence that stereoscopic volume within facial features may be involved in perception; however, with large standard errors more rigorous investigation is required.

Further studies have explored how we perceive 3D faces with a volumetric structure, through our classification of their curvilinear properties or organic shape. Curvilinear geometry is defined as a shape bound by curved, organic lines as seen in animate objects. Inanimate objects, characteristically include linear geometry or edges bound by straight lines. In fact, facial identity or individuality is often built into 3D computerized models, by using multiple, individualized curves, rather than the straight lines used to build objects (Yin et al., 2006; Liu et al., 2012; Li et al., 2018). Researchers have likewise utilized curvilinear shapes in 2D images to demonstrate that the entire class of animate objects (e.g., faces, bodies, animals) are detected
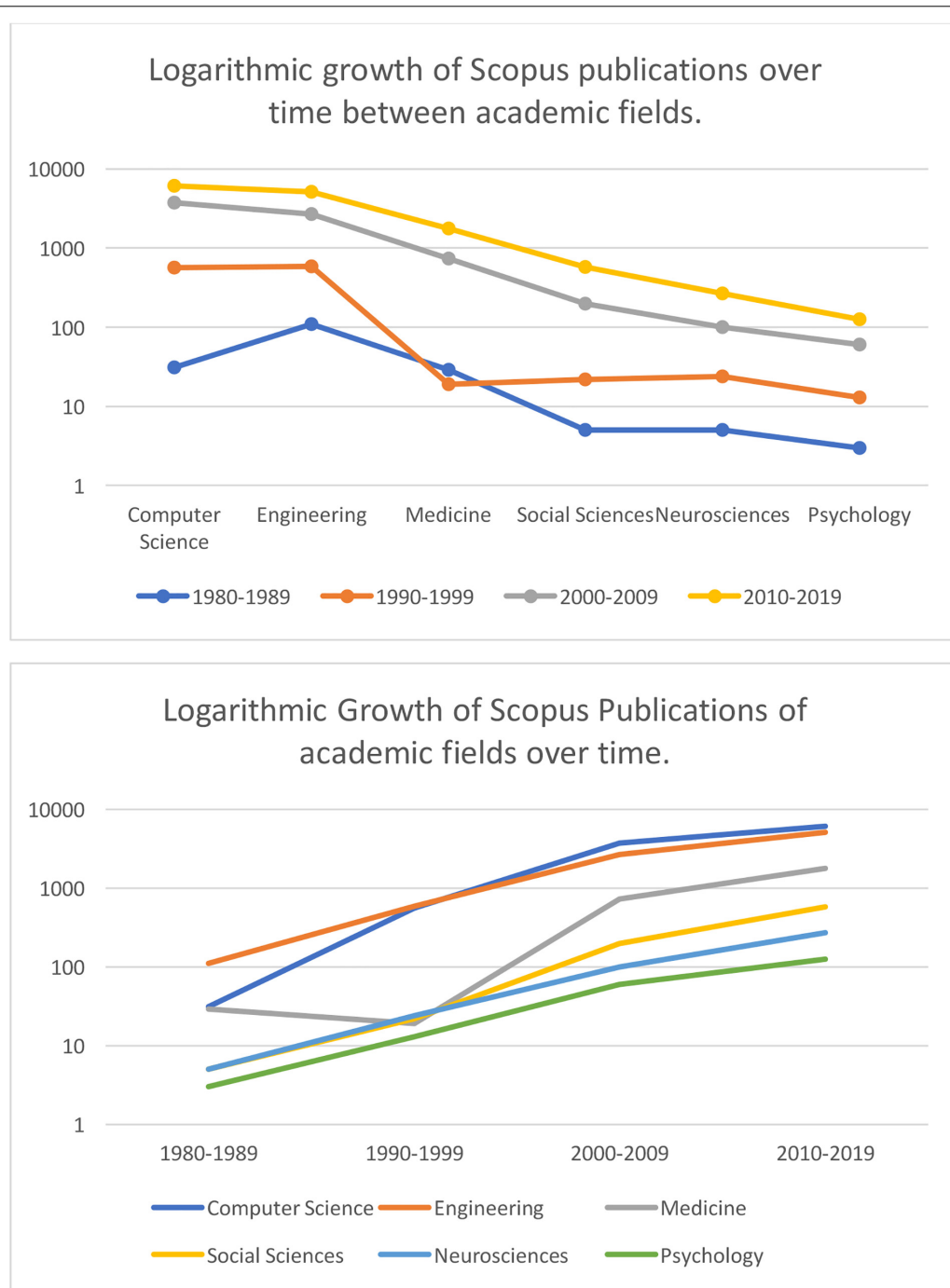
FIGURE 4 | Logarithmic Scale of Scopus publication growth from 1980 to 2019 across fields containing key search terms. Search terms: ("allintitle, abstract or keywords": "3-D face" OR "3D face" OR "three-dimensional face" OR "three-dimensional face" OR "4D").

and classified principally through curvilinear discrimination (Levin et al., 2001; Long et al., 2016; Zachariou et al., 2018). Inanimate object recognition, by comparison, relies mainly on detecting rectilinear shapes and lines, such as that observed in a box or house object (Levin et al., 2001; Long et al., 2016; Zachariou et al., 2018). In terms of human 3D perception

of faces, Jones et al. (2012) have demonstrated that we can perceive some personality traits in others, using exclusively 3D shape or curvilinearity. Presenting 3D facial scans exclusively with curvilinear shape, with all surface properties removed, participants accurately reported above-chance another person's self-reported Big-5 trait-level of "Neuroticism" (Goldberg, 1993;
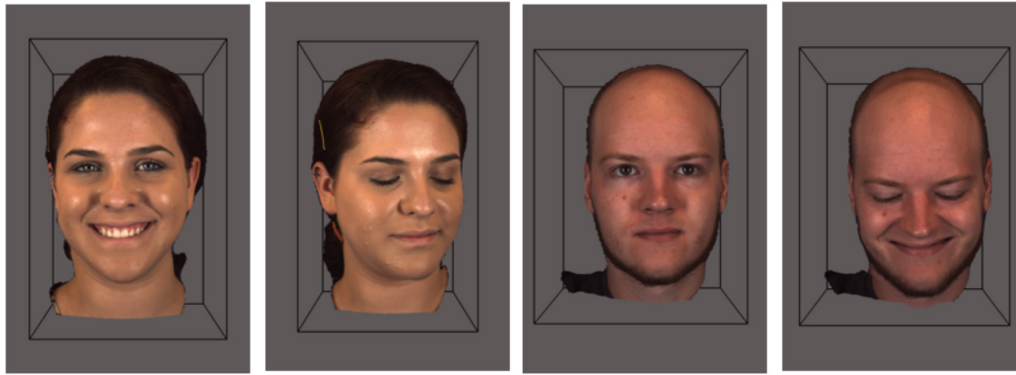
FIGURE 5 | BP-4D-S facial database presented on gray background with bounding box representing a 3D depth plane (The images pictured above were reproduced, modified and adapted from the facial database originally developed by Zhang et al. (2013, 2014). Copyright 2013–2017, The Research Foundation for the State University of New York and University of Pittsburgh of the Commonwealth System of Higher Education. All Right Reserved. BP4D-Spontaneous Database, as of: 26/03/2018. 3D faces are built on-screen as models with a geometric mesh in a three-dimensional plane; presenting an anatomically accurate modeled object with width, height and depth. This is captured by the 3D camera set-up. This reconstruction by computer-scientists is highly individualized, but is typically underlined by a geometric mesh or structure. The 3D surface features of the individual are then rendered. As the face moves around in 3D space, it becomes apparent why having three dimensions maintains facial structure more realistically than a facial model built with only two dimensions, as indicated by the depth plane.

Jones et al., 2012). Thus, curvilinear shape appears to be a visual cue that aids perception of faces, as well as categorizing social or personality traits in others. Although, overall to identify traits, a combination of geometry and surface cues have been found to enhance perception (Jones et al., 2012).

Alongside, a geometric volume or structure, 3D surface rendering is a key process in developing 3D and 4D facial models (*human faces see* Marschner et al., 2000 *or for primate faces* Murphy and Leopold, 2019). The surface of the human face, or epogeneous skin layer, is a highly individualized, changeable feature which aids recognition. Our unique skin surface, including our individualized texture, shading, illumination, reflectance and pigmentation can be reproduced and rendered on a computer screen by using a 3D facial model (Egger et al., 2019). One common skin rendering technique used in these databases, for example, is bump surface models, with anatomically correct models being the ultimate goal (Yin et al., 2006; Zhang et al., 2014). As early as 1991, the psychological or neuroscientific literature was described as "surface primitive" (Bruce et al., 2013) indicating it lacked evidence of how 3D rendered surfaces contributed to our perceptions. Since then 2D facial imagery has helped reveal the contributions of surfaces to perception; where our recognition of faces becomes impaired due to textural or color cues being unexpected, such as in contrast polarity (Hayes et al., 1986; Liu and Chaudhuri, 1997, 2003; Itier and Taylor, 2002; Russell et al., 2006), or when lighting direction is in an unusual or bottom position (Hill and Bruce, 1996; Braje et al., 1998; Johnston et al., 2013).

Similarly, surface properties conveyed by 3D rendering have shown that several factors contribute to our perceptions, such as texture gradients, reflectance, pigmentation and lighting styles. In viewing 3D images, both surface and structure have been considered as key components of our perception of unfamiliar faces or strangers (O'Toole et al., 2005). When examining 3D scanned models of unfamiliar faces, Liu et al. (2005) reveal

texture gradients alone can facilitate recovery of facial identity above-chance, although the performance was poorer than shape-from-shading, suggesting shadows and illumination may be more critical. More recently it has been suggested that when perceiving familiar 3D faces of friends and acquaintances, we do not heavily rely on texture or surface information and instead rely upon 3D structure (Blauch and Behrmann, 2019). Contrastingly, when looking at 2D faces, familiar faces have been suggested to be recognized more so from an individuals' surface information, rather than the two-dimensional shape (Russell and Sinha, 2007). For example, where skin surfaces are removed from an underlying 2D facial shape, familiar faces cannot be recognized (Russell and Sinha, 2007).

As well as texture, 3D rendering also encompasses coloring or skin pigmentation, which has been shown to be a key feature of identification (Yip and Sinha, 2002; Russell et al., 2007). One study of the effect of color and pigmentation cues on 3D face models, describes our accurate ability to detect kinship or genetic relationships of individuals we have not seen previously (Fasolt et al., 2019). For example, we can accurately detect kinship through skin pigmentations exclusively, as well as genetic facial morphology. Interestingly, Fasolt et al. (2019) concluded that we did not require a combination of structure and pigmentation, as high-accuracy was obtained from each of these properties alone. Similarly, physical health is perceived in another, based on skin pigmentation and texture, more so than 3D shape. For example, Jones et al. (2012) demonstrate physical health of a 3D face can be accurately classified using texture exclusively, based on true health in daily living (Jones et al., 2012). Skin surfaces were also suggested to provide accurate information about a person's Big-5 self-reported trait levels of "Agreeableness" and "Extraversion" (Jones et al., 2012).

In summary, it appears the unique, complex geometric forms and surface properties contained within 3D faces, may be a *vital factor* underlying our perceptions of each other. This

effect has been investigated in identity recognition tasks of unfamiliar faces (Levin et al., 2001; Pizlo et al., 2010; Long et al., 2016; Eng et al., 2017; Zachariou et al., 2018), memory of familiar faces (Blauch and Behrmann, 2019; Kinect Xbo,, 2019; Zhan et al., 2019) and our detection of important social cues, such as genetic relatedness (Fasolt et al., 2019) health or personality traits (Jones et al., 2012). While studies examining our perceptions of 3D facial volume and surfaces are relatively few, they suggest potential for future behavioral and neuroimaging studies. Further examining these 3D properties of faces using applications, such as VR or stereoscopic stimuli, will not only inform visual scientists, but also contribute knowledge of how facial perceptions are achieved in social, affective and cognitive psychology. For example, 3D models with volume, symmetry, curvilinear edges, shadows, texture, pigmentation, reflectance and illumination when systematically controlled may influence our perceptions of identity, gender, ethnicity, age, health, beauty, personality and kinship.

The Influence of 3D Depth in Building Facial Perception

Depth is a visual cue in 2D imagery and digital screens, that creates the illusion of having linear perspective. In other words, depth cues refer to all visual features which gives humans the ability to perceive that objects are at a distance from us. When viewing a two-dimensional screen, this requires creating an illusion of depth perception from a flat surface. Without being explicitly a 3D object, depth perception is still recovered or perceived from two-dimensional flat screens, through visual cues such as stereopsis or binocular disparity, linear perspective or alignment, convergence, object constancy, illumination, shadows and textural gradients (Hallinan et al., 1999; Vishwanath and Kowler, 2004; Häkkinen et al., 2008; Lambooij et al., 2011). The illusion of more realistic or naturalistic depth, is often what separates stimuli from being considered 2D, such as we see in traditional photographs, or appearing 3D with applications such as VR (Blascovich et al., 2002; Bailenson et al., 2004) or stereoscopic glasses and monitors (Vishwanath and Kowler, 2004; Nguyen and Clifford, 2019). However, to date the exact similarities and differences between how we utilize depth perception when viewing 2D facial imagery and 3D modeled faces remains unresolved.

Several preliminary studies have investigated how depth cues influence facial recognition with 3D models presented on two-dimensional screens, however, findings so far have generally been mixed (Chang et al., 2005; Farivar et al., 2009; Dehmoobadsharifabadi and Farivar, 2016). For example, Dehmoobadsharifabadi and Farivar (2016) demonstrate that a facial identity can be recognized significantly above-chance with only one depth cue presented within 3D faces; either stereopsis, texture gradients, structure-from-motion and binocular disparity cues. While use a similar method, but identify that shading aids accurate identity recognition of 3D facial models more so than other depth cues. In eye-tracking studies, Chelnokova and Laeng (2011) demonstrate 3D stereoscopic faces alter eye-gaze

attention, directing attention toward the nose and chin, more so than the eyes; an effect not commonly reported with typical 2D photographs of faces (Chelnokova and Laeng, 2011). Atabaki et al. (2015) further demonstrate there is no eye-gaze pattern or angular differences when perceiving a real human or 3D stereoscopic avatar, suggesting that viewing a stereoscopic face on a screen produces comparable attention strategies to viewing physical humans. Neuroimaging studies present evidence of increased dorsal-stream engagement when viewing 3D stimuli containing depth cues, for example stereopsis cues and structure-from motion-cues (Zeki et al., 1991; Tootell et al., 1995; Heeger et al., 1999; Backus et al., 2001; Tsao et al., 2003; Kamitani and Tong, 2006; Freud et al., 2018b). Depth added to unfamiliar faces is largely attributed to increasing both dorsal-ventral stream engagement, when presented with motion (Farivar, 2009). Overall, these initial studies infer that depth cues presented with 3D models and applications, may impact our psychophysical perception, eye-tracking and neural strategies. Comparatively, conflicting studies have also demonstrated no psychophysical improvement when comparing 2D and 3D stereoscopic faces under different task conditions, such as when poses of the face change (Hong et al., 2006) or for unfamiliar faces (Hancock et al., 2000). Liu and Ward (2006) for example, found 3D (stereo) and 2D (without stereo) produce similar recognition, suggesting that depth cues must be presented in the correct combination (Liu and Ward, 2006). Overall, this preliminary evidence suggests both similarities and differences may exist in how we perceive depth from matched 3D and 2D stimuli. Future research therefore, may benefit from directly comparing matched 2D-3D facial stimuli, as presented by Sadhya and Singh (2019) or by implementing 3D depth with VR or stereoscopic vision (Blascovich et al., 2002; Bailenson et al., 2004).

Concluding this section, depth cues such as texture and shading gradients, shape-from-motion and stereoscopic vision or binocular disparity cues, may be beneficial to identity recognition for unfamiliar faces, as well as producing similar eye-gaze attentional strategies as viewing real-life humans (Chelnokova and Laeng, 2011; Dehmoobadsharifabadi and Farivar, 2016). However, there may be no differences in depth perception between viewing 3D facial models and 2D imagery (Hancock et al., 2000; Hong et al., 2006). Depth cues presented by 3D models on-screen, may also underlie reproduction of natural eye gaze patterns toward more structural features, that would typically occur in a physical human–human interaction, and which is limited by 2D stimuli (Vishwanath and Kowler, 2004; Chelnokova and Laeng, 2011; Atabaki et al., 2015). Adding the third-dimension of depth to a facial stimulus, alongside motion, may increase both dorsal and ventral stream engagement (Farivar et al., 2009). Overall, there remains unresolved areas for future research in how depth cues of 3D and 2D models are perceived. Future research can ultimately benefit by determining how important each of these depth cues are in our facial perceptions of both 3D stimuli compared to 2D matched-stimuli (Sadhya and Singh, 2019) and by utilizing 3D models presented in VR or stereoscopic vision (Blascovich et al., 2002; Bailenson et al., 2004; Chelnokova and Laeng, 2011).

WHAT IS THE 4D TEMPORAL PATTERN OF THE HUMAN FACE?

The Influence of 4D Dynamic Viewpoint Changes in Building Facial Perception

Another critical contribution of 3D facial models to understanding facial perceptions, is the ability to present faces at multiple viewpoints due to their dynamical, moving nature. Faces are highly dynamical objects which move freely in space. Thus, to examine the face as a realistic perceptual object on-screen, it must be modeled with unconstrained motion. Unconstrained motion on-screen, simply replicates the natural dynamic movements and changes in viewpoint as a person moves their head. From the observers view- This encompasses changes to our viewpoint through rotation, position, orientation, visual angle, distance, and occlusion. Similarly, this effect can be reproduced on-screen to examine facial perception with different viewpoints of the same face. One of the greatest contributions 3D face databases offer is the inclusion of unconstrained head motion within space (for some examples see **Figure 6** MPI; Kaulard et al., 2012; *BP-4D-S*; Zhang et al., 2014).

Conventionally, viewpoint changes have been highly restricted in experimental research in the past. In the literature, two-dimensional faces are consistently presented at one fixed-viewpoint, which is incongruent with our daily life experience (Trebicky et al., 2018). This viewpoint is termed the "*central-frontal plane*," where a face is positioned directly facing the camera. Indeed, the reasoning behind this decision may be sound. Incremental shifts in camera angle away from the "*central-frontal*" viewpoint of a 2D face, leads to incrementally poorer recognition performance (Lee et al., 2006). Similarly, psychophysical evidence indicates that 2D faces with a different or novel viewpoint, suffer a cost to reaction time and sensitivity (Shepard and Metzler, 1971; Bulthoff and Edelman, 1992; Hill et al., 1997; Wallraven et al., 2002). Of course, where repetitive learning or familiarity with a face-viewpoint is gained (Bulthoff and Edelman, 1992; Logothetis and Pauls, 1995; Tarr, 1995) or when experimental task demands are changed (Vanrie et al., 2001; Foster and Gilson, 2002) improvement can occur.

2D databases are available, where facial viewpoint is systematically changed in orientation or visual angle (*UT Dallas*; O'Toole et al., 2005; *AD-FES*; Van Der Schalk et al., 2011; *GEMEP Core set*; Bänziger et al., 2012; *MMI*; Valstar and Pantic, 2010; *MPI bio*; Kleiner et al., 2004) or alternatively through capturing unconstrained head movement (FG-NET FEEDtum; Wallhoff, 2004). 4D facial databases containing unconstrained head movement in a three-dimensional space have also been developed. For example, see the MPI; Kaulard et al., 2012 or *BP-4D-S*; Zhang et al., 2014. Psychophysical investigation using 3D stimuli has identified viewpoints or angles that may be beneficial for facial identification. For example, a decline in accuracy and efficiency in 3D facial recognition is observed, where viewpoint rotations occur along yaw, pitch and roll axes for 3D simulated stimuli (Favelle et al., 2011, 2017). Optimal viewpoints of unfamiliar faces, by contrast, are gained by presentation at both a frontal plane or yaw rotation (Favelle
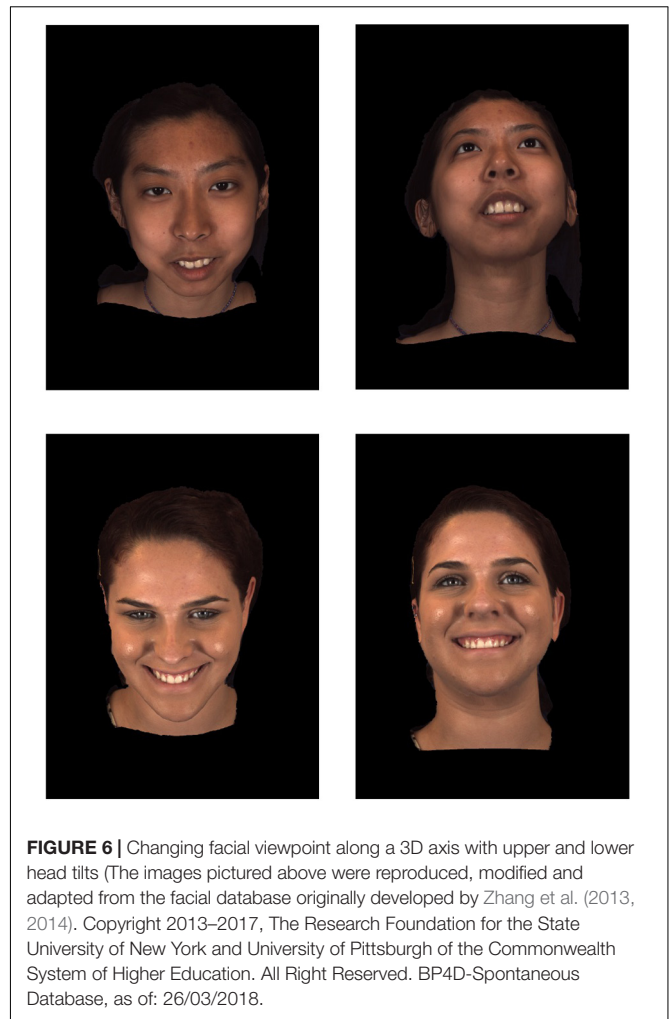

FIGURE 6 | Changing facial viewpoint along a 3D axis with upper and lower head tilts (The images pictured above were reproduced, modified and adapted from the facial database originally developed by Zhang et al. (2013, 2014). Copyright 2013–2017, The Research Foundation for the State University of New York and University of Pittsburgh of the Commonwealth System of Higher Education. All Right Reserved. BP4D-Spontaneous Database, as of: 26/03/2018.

and Palmisano, 2018). Interestingly, recognition of 3D models in one study does not appear affected by viewing the asymmetrical right or left sides of the face, through yaw and roll rotations (Favelle et al., 2011, 2017).

Another study of 3D models in a VR environment, has suggested that the amount of facial information available to the visual system is the factor which most impacts recognition (Bülthoff et al., 2018). For instance, occlusion of prominent facial features or the holistic configuration of features, will reduce performance significantly in a VR environment. Consequently, pitch rotations impact recognition greatly (Bülthoff et al., 2018) – an upward pitch will produce poorer performance followed by downward pitch occlusions of 3D face stimuli (Van der Linde and Watson, 2010; Favelle et al., 2011, 2017). Roll rotations produce the greatest accuracy overall, simply because they do not contain *any* occlusion of facial features, with little disruption to the facial configuration information detected by the visual system (Favelle et al., 2011). Contrastingly, Wallis and Bülthoff (2001) used 3D head models to show that identity is not recognized well, if the identity changes as the head rotates in a temporally smooth sequence suggesting we have a strong temporal association of

facial identity from memory. Thus, it appears that 3D faces with viewpoint changes may both enhance or decrease our recognition performance depending on the task conditions, with substantial contributions to the theories of featural or configural facial processing being made.

In another study, as the viewing angle of 3D model faces changed, information regarding dominance, submission and emotions displayed by another person was gained. More precisely, Mignault and Chaudhuri (2003) describe how a 3D viewpoint of a raised head tilt or upright pitch, is associated with increased personality trait-attributions of dominance and superiority; alongside emotional states of increased contempt, pride and happiness. Contrastingly, a downward head tilt or pitch implies to the viewer increased personality-attribution of submissiveness; alongside emotional states of increased sadness, inferiority, shame, embarrassment, guilt, humiliation and respect. By using 3D dynamical facial models which naturally change viewpoint and position in space, these initial studies have identified how personality and emotion perception can be impacted with viewpoint changes of the head.

Neuroimaging evidence suggests that 2D facial processing relies upon neurons sensitive to changes in viewpoint, in the Superior Temporal Sulcus (STS) and Infero-Temporal Cortex (IT). Unimodal function or exclusive firing for changes in viewpoint is demonstrated in these areas (Perrett et al., 1985, 1992; Oram and Perrett, 1992). For example, significant decrease in firing rates of these regions is observed when facial viewpoint is changed from the "*central-frontal*" plane, such as left and right profile or anterior positions (Logothetis and Pauls, 1995). These detected neurons groups may be responsible for the psychophysical findings of facial viewpoint, as discussed above. More broadly, Pourtois et al. (2005) found the Temporal cortex will respond to viewpoint changes, while the *FFA* does not respond to changes in viewer angle. To pursue this question in three dimensions, Ewbank and Andrews (2008) presented both familiar and unfamiliar 3D faces. When viewing rotation or orientation changes, face-selective voxels required more recovery from adaption (Grill-Spector et al., 1999; Andrews and Ewbank, 2004). Both emotion and viewpoint changes in 3D face stimuli did however, activate FFA (Grill-Spector et al., 1999; Andrews and Ewbank, 2004). While shape or size dimensions did not appear to matter, the FFA appeared to activate category-specific information of viewpoint change. These results are consistent with single-cell recording data of the primate infero-temporal cortex (IT) or the superior temporal sulcus (STS) and fMRI studies with human and primate subjects, which have independently studied the effects of size and viewpoint changes (Perrett et al., 1985, 1992; Sáry et al., 1993; Lueschow et al., 1994; Logothetis and Pauls, 1995; Wang et al., 1996; de Beeck et al., 2001; Pourtois et al., 2005).

In summation, 3D and 4D models have already made a substantial contribution to facial research pertaining to viewpoint change, for both familiar and unfamiliar faces (Wallis and Bülthoff, 2001; Favelle et al., 2011, 2017; Favelle and Palmisano, 2018). Renewed potential for investigations into the featural and configural theories of facial processing, has also been addressed by occlusions through viewpoint change (Van der Linde and Watson, 2010; Favelle et al., 2011, 2017; Bülthoff et al., 2018). Initial investigations into viewpoint-dependent face regions (FFA) in the brain are emerging (Grill-Spector et al., 1999; Andrews and Ewbank, 2004) although require more extensive analysis. Finally, naturalistic, unconstrained viewpoint changes provided by 4D dynamical facial modeling, may be most beneficial in understanding how we perceive emotional state and personality traits of others (Mignault and Chaudhuri, 2003).

The Influence of 4D Dynamic Movement in Building Facial Perception

Historically, facial literature has fixated on static, 2D imagery in experimental design; that is, on-screen stimuli without movement (Kamachi et al., 2013). Overall, inconsistent findings in the literature have been reported for facial recognition relating to 2D dynamic or moving faces that were recorded with 2D capture- namely width and height (for a review see Kätsyri, 2006; Fiorentini and Viviani, 2011; Alves, 2013; Krumhuber et al., 2017). Some psychophysical studies have revealed an advantage in identification, when 2D facial motion is displayed, compared to static counter-parts (Pike et al., 1997; Wehrle et al., 2000; Knappmeyer et al., 2003; Ambadar et al., 2005; Cunningham and Wallraven, 2009) while others have not (Fiorentini and Viviani, 2011; O'Toole et al., 2011). Another significant contribution of 3D models to the facial research community therefore, is the progression of databases into four dimensions (4D). In other words, facial stimuli can now be presented with the additional dimension of dynamic movement to a 3D model. This progression is ultimately necessary when we consider that human faces are "*highly dynamical*" by nature or rarely "*still*" objects and yet are also three-dimensional. Indeed, precise variation in facial motion displayed in three dimensions, allows us to portray and communicate a wealth of information as a social species (Jack and Schyns, 2017).

Communication, or our wide-ranging emotions, mental states and speech, is conveyed to other humans principally through two forms of movement which are built into 3D digitized models. Facial movement can be modeled as rigid and non-rigid motion (for review see Smeets et al., 2012; Chrysos et al., 2018). Evidence suggests anatomically accurate rigid and non-rigid facial movements are essential to facial perception (Trautmann et al., 2009; Johnston et al., 2013) and is achievable computationally (Zhang et al., 2003; Zhang and Fisher, 2019). Deformable facial modeling, relates to movement we produce through voluntary, striated musculature or facial muscles. For example, when we talk, raise our eyebrows or smile, we activate this musculature system. Rigid head modeling, relates to the constrained-motion produced by the head, which is controlled through the neck musculature. An example of rigid motion is when we change the posture of our head, such as when shaking our head in disagreement or nodding in agreement. Both deformable modeling and rigid head modeling help reproduce naturalistic movements on-screen. Briefly, some stand-out examples of facial databases

with motion include the *BU4D-S* (Zhang et al., 2014) and the *D3DFACS* (Cosker et al., 2011). However, also see (Wang et al., 2004; Vlasic, 2005; Gong et al., 2009; Weise et al., 2009; Garrido et al., 2013; Shi et al., 2014; Suwajanakorn et al., 2014; Zhang et al., 2014; Cao et al., 2015; Thies, 2015; Jeni et al., 2017). Comprehensive reviews comparing the static (3D) and dynamic (4D) datasets are also available, as evaluated by Sandbach et al. (2012) and Smeets et al. (2012). However, a recent shift toward dynamic, moving facial databases has also been observed in the 2D literature and is worth mentioning (*for a review see* Krumhuber et al., 2013; Krumhuber et al., 2017).

The emerging field of 4D facial research is therefore likely to offer answers into how naturalistic social, cognitive and affective perceptions are achieved when observing facial motion. Deformable movement is the more popular facial research area, encompassing studies of speech, language and emotional expressions. Deformable movement of our facial musculature in the literature is defined popularly by the Facial Action Coding System (Ekman and Friesen, 1978; Ekmans and Rosenberg, 1997). FACS is a highly cited, objective system of measuring facial movement activated by groups of facial muscles, defined as "action units." (Ekman and Friesen, 1978; Ekmans and Rosenberg, 1997 44 action units, the FACS system groups facial muscles which activate collectively. These activations are ultimately what produce deformable expression and speech output. For example, an emotion pattern such as a disgusted response, can be quantified into an action unit or series of muscle activations (AUs), such as a "grimace" of the mouth and "screwing up" of the eyes. Exemplars of different FACS coded muscle activations, illustrating smiles, frowns and grimaces of deformable facial features are displayed in **Figure 7**.

Recently, Savran et al. (2012) presented a detailed review to compare how the FACS system aids perception of 2D and 3D faces, suggesting emotion recognition with the FACS system is superior in 3D faces, particularly with low intensity emotions. While the FACS system has been extensively used with 2D dynamic faces displaying emotional expressions, for example, 2D images (Pantic and Rothkrantz, 2004; Cohn and De la Torre, 2014; Baltrušaitis et al., 2015) and 2D dynamic videos (Donato et al., 1999; Pantic and Rothkrantz, 2004; Bartlett et al., 2006) this study indicates 3D/4D dynamic faces may be more important for low intensity emotions. The FACS system has likewise achieved action unit detection in 3D facial videos or scans (Cosker et al., 2011) for successful emotion detection (Sun et al., 2008; Tsalakanidou and Malassiotis, 2010; Berretti et al., 2013; Reale et al., 2013; Tulyakov et al., 2015; Danelakis et al., 2018); and even more socially related morphologies such as dominance, trustworthiness and attractiveness (Gill et al., 2014) or the influences of ethnicity and gender on how we display emotional expressions differently from one another (Jack et al., 2014; Jandova and Urbanova, 2018).

Deformable modeling has made a further advancement in facial stimuli design, by digitizing naturalistic, spontaneous movements, rather than the synthesized, artificial movements produced in the past (Wehrle et al., 2000; Tcherkassof et al., 2007). For example, high classification accuracy using the 3D FACS system has been demonstrated for spontaneous emotions displayed in these databases (96%; Tarnowski et al., 2017). Spontaneous deformation provides computer scientists the ability to track anatomically accurate emotional displays from humans in-situ (Bartlett et al., 2006; Zhang et al., 2014). This is of consequence to emotion and language research, as motion or timing of deformable changes, is considered important (Bassili, 1978; Ekman, 1982; Wehrle et al., 2000; Kamachi et al., 2013). This is exemplified in the case of "*false smiles.*" For example, while the FACS activated in any smile is highly similar, it is the *timing* of activation which reveals the difference between a deceptive smile and a smile of genuine happiness, as can the consistency between the eye and mouth FACS (Ekman, 1982; Cohn and Schmidt, 2003; Guo et al., 2018). 4D faces have also helped delineate *when* we can evaluate an emotional expression. For example, most expressions begin the same, with the 6 basic emotional expressions becoming revealed temporally later in Western Caucasians, as shown through the FACS system (Jack et al., 2014). More culturally specific, East Asians have demonstrated more overlap and less specificity in the FACS displaying surprise, fear, disgust and anger, compared to Western Caucasians (Jack et al., 2014). Overall, deformable modeling of 4D faces combined with 3D FACS, offer researchers a promising methodology of speech, language and emotion which use deformable facial musculature.

Albeit less studied, rigid modeling of 4D face datasets is also of key importance to include in dynamic stimuli presentation. Rigid head movement is defined as that constrained by the neck, involving changes to the position or posture of the head. Rigid motion in the literature is defined and measured along 3D axes, including changes in head tilt, pitch, yaw and roll. Using 3D neutral facial models, emotional states or personality traits associated with head tilt include a downward head angle as guilt, sadness, embarrassment, gratitude, fear, surprise, shame, guilt as submissive displays upward head angle as pride, contempt, and scorn as dominance (Mignault and Chaudhuri, 2003). For example, rigid head motion alone has demonstrated the ability to determine an individuals' identity (Hill and Johnston, 2001). 3D head posture alone, with neutral expression, indicates a wide array of information (Mignault and Chaudhuri, 2003), including gender-based displays of dominance and aggression, or alternatively submission. Many traits are perceived through head tilt; including a lowered head indicating sadness, inferiority and submissiveness (Otta et al., 1994). Overall, these initial studies of unconstrained movement, while relatively rare, have provided methodologies whereby 4D faces can be used to improve studies of language, emotion and personality traits, alongside social factors such as gender and ethnicity. Thus, it is likely a growing number of studies in future will utilize facial stimuli with unconstrained head motion, as presented by deformable and rigid head modeling. Overall, both rigid and deformable modeling of 4D faces combined with the 3D FACS system, offer psychologists a promising methodology for all areas of speech, language and emotion research.

FIGURE 7 | Video-recorded dynamic 4D facial stimulus captured as still frame-by-frame images, displaying happy and disgusted FACS coded emotion (The images pictured above were reproduced, modified and adapted from the facial database originally developed by Zhang et al. (2013, 2014). Copyright 2013–2017, The Research Foundation for the State University of New York and University of Pittsburgh of the Commonwealth System of Higher Education. All Right Reserved. BP4D-Spontaneous Database, as of: 26/03/2018. This exemplar presents the frame-by-frame stills from a video-recording of a 4D face, moving from a happy expression to a disgusted expression.

RECOMMENDATIONS

In this review, we discuss the nature of 4D face stimuli that are modeled over space and time. The review introduces the available 3D/4D face models and databases for researchers, while considering how these are captured and reconstructed for experimental design. Finally, we review studies which incorporate 3D spatial and 4D temporal properties to reveal how volume, surfaces, depth, viewpoint and movement affect our perceptions of others. In sum, we recommend the implementation of 4D stimulus models built from real-life human faces with anatomically accurate spatial and temporal dimensions, to resolve many unanswered questions of facial perception.

Future research in the vision sciences can ultimately benefit by determining how important 3D structural and depth properties are in our facial perceptions of 3D models compared to 2D matched-stimuli (Sadhya and Singh, 2019). Secondly, at the start of this revolution, we will begin to understand how facial perception is realized in increasingly naturalistic settings, which capture and reproduce the complex interaction of 3D volume, surface, depth, viewpoint and movement properties (Bulthoff and Edelman, 1992; Bailenson and Yee, 2006; Stolz et al., 2019; Lamberti et al., 2020). For example, 3D/4D facial models enable complex visual cues and 3D depth within naturalistic social environments, once integrated with applications such as VR and stereoscopic technology (Bailenson et al., 2004; Stolz et al., 2019; Lamberti et al., 2020). By assimilating complex models of face stimuli into future research, we can resolve long-standing questions about how we identify familiar compared to unfamiliar faces, especially from different viewpoints (Wallis and Bülthoff, 2001; Abudarham et al., 2019; Zhan et al., 2019). We may also improve understanding of how we identify social characteristics of individuals; such as gender, ethnicity or personality traits. We can better explore social perceptions with 3D facial features and surfaces by investigating many unresolved questions, such as what the nose or eyes reveal about ethnicity and gender? (Lv et al., 2020) or how do we identify race, health and age from skin surfaces? (Jones et al., 2012). Likewise, we can examine the decoding of more complicated emotional expressions, speech and language with 3D FACS over time (Lu et al., 2006; Jack et al., 2014). Finally, we may better understand the reasons why emotional expressions, inversion or movement, present challenges for specific clinical subpopulations during facial perception tasks, such as in Autism Spectrum Disorder (Weigelt et al., 2012; Anzalone et al., 2014) or Schizophrenia (Onitsuka et al., 2006; Wang et al., 2007; Marosi et al., 2019).

To meet the challenges of integrating complex 4D face models in psychological methodology, more complex computational methods of analysis will be required. It is recommended that future experiments utilize multivariate time-resolved methods of analysis. Thus, as a facial stimulus evolves over time, methods of analysis should simultaneously capture this dynamical information. Existing methods such as psychophysical behavioral measures (eye-tracking, reaction-time, inspection time, accuracy) or neuroimaging measures (decoding, multivariate pattern analysis or time-frequency analysis) are recommended.

CONCLUSION

In conclusion, this in-depth review highlights the effects of presenting on-screen facial stimuli with both 3D visual properties (volume, surface and depth) and 4D dynamical properties (viewpoint, movement). We conclude that *the human face is inherently defined as a 4D dynamic* perceptual object. Ignoring this definition in experimental stimuli remains a widespread issue which confounds our understanding of naturalistic human facial perception is performed. To date, the emerging use of 4D dynamic models in facial research has so far revealed an amazing richness in facial perceptions of identity, gender, ethnicity, personality traits, speech and emotional expressions.

AUTHOR CONTRIBUTIONS

AB wrote the manuscript including synthesis of literature search, performed manuscript writing and editing. DC performed the editing. Both authors contributed to the article and approved the submitted version.

REFERENCES

Abudarham, N., Shkiller, L., and Yovel, G. (2019). Critical features for face recognition. *Cognition* 182, 73–83. doi: 10.1016/j.cognition.2018.09.002

Allen, B., Curless, B., Curless, B., and Popović, Z. (2003). "The space of human body shapes: reconstruction and parameterization from range scans," in *Paper Presented at the ACM Transactions on Graphics (TOG)*, New York, NY.

Alves, N. T. (2013). Recognition of static and dynamic facial expressions: a study review. *Estudos Psicol.* 18, 125–130. doi: 10.1590/s1413-294x2013000100020

Ambadar, Z., Schooler, J. W., and Cohn, J. F. (2005). Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychol. Sci.* 16, 403–410. doi: 10.1111/j.0956-7976.2005.01548.x

Andrews, T. J., and Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage* 23, 905–913. doi: 10.1016/j.neuroimage.2004.07.060

Anzalone, S. M., Tilmont, E., Boucenna, S., Xavier, J., Jouen, A.-L., Bodeau, N., et al. (2014). How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3D+time) environment during a joint attention induction task with a robot. *Res. Autism Spectr. Disord.* 8, 814–826. doi: 10.1016/j.rasd.2014.03.002

Atabaki, A., Marciniak, K., Dicke, P. W., and Thier, P. (2015). Assessing the precision of gaze following using a stereoscopic 3D virtual reality setting. *Vision Res.* 112, 68–82. doi: 10.1016/j.visres.2015.04.015

Backus, B. T., Fleet, D. J., Parker, A. J., and Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *J. Neurophysiol.* 86, 2054–2068. doi: 10.1152/jn.2001.86.4.2054

Bailenson, J. N., Beall, A. C., Blascovich, J., and Rex, C. (2004). Examining virtual busts: Are photogrammetrically generated head models effective for person identification? *Presence* 13, 416–427. doi: 10.1162/1054746041944858

Bailenson, J. N., and Yee, N. (2006). A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence* 15, 699–716. doi: 10.1162/pres.15.6.699

Baltrušaitis, T., Mahmoud, M., and Robinson, P. (2015). "Cross-dataset learning and person-specific normalisation for automatic action unit detection," in *Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Ljubljana.

Bänziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion* 12, 1161–1179. doi: 10.1037/a0025827

Bartlett, M. S., Littlewort, G., Frank, M. G., Lainscsek, C., Fasel, I. R., and Movellan, J. R. (2006). Automatic recognition of facial actions in spontaneous expressions. *J. Multimed.* 1, 22–35.

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 373–379. doi: 10.1037/0096-1523.4.3.373

Beaupré, M., Cheung, N., and Hess, U. (2000). *The Montreal Set of Facial Displays of Emotion [Slides]*. Montreal: University of Quebec.

Beeler, T., Bickel, B., Beardsley, P., Sumner, B., and Gross, M. (2010). High-quality single-shot capture of facial geometry. *Paper Presented at the ACM Transactions on Graphics (TOG)*, New York, NY.

Behrmann, M., and Avidan, G. (2005). Congenital prosopagnosia: face-blind from birth. *Trends Cogn. Sci.* 9, 180–187. doi: 10.1016/j.tics.2005.02.011

Benedikt, L., Cosker, D., Rosin, P. L., and Marshall, D. (2010). Assessing the uniqueness and permanence of facial actions for use in biometric applications. *Syst. Man Cybern. Part A Syst. Hum. IEEE Trans.* 40, 449–460. doi: 10.1109/tsmca.2010.2041656

Berretti, S., Del Bimbo, A., and Pala, P. (2013). Automatic facial expression recognition in real-time from dynamic sequences of 3D face scans. *Vis. Comput.* 29, 1333–1350. doi: 10.1007/s00371-013-0869-2

Beumier, C., and Acheroy, M. (1999). "SIC DB: multi-modal database for person authentication," in *Proceedings 10th International Conference on Image Analysis and Processing*, Venice.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., and Bailenson, J. N. (2002). TARGET ARTICLE: immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* 13, 103–124. doi: 10.1207/s15327965pli1302_01

Blauch, N., and Behrmann, M. (2019). Representing faces in 3D. *Nat. Hum. Behav.* 3, 776–777. doi: 10.1038/s41562-019-0630-6

Bouaziz, S., Wang, Y., and Pauly, M. (2013). Online modeling for realtime facial animation. *ACM Trans. Graph.* 32, 1–10.

Braje, W., Kersten, D., Tarr, M., and Troje, N. (1998). Illumination effects in face recognition. *Psychobiology* 26, 371–380.

Bruce, V., Coombes, A., and Richards, R. (1993). Describing the shapes of faces using surface primitives. *Image Vis. Comput.* 11, 353–363. doi: 10.1016/0262-8856(93)90014-8

Bruce, V., Healey, P., Burton, M., Doyle, T., Coombes, A., and Linney, A. (2013). Recognising facial surfaces. *Perception* 42, 1200–1214. doi: 10.1068/p200755n

Bulthoff, H. H., and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl. Acad. Sci. U.S.A.* 89, 60–64. doi: 10.1073/pnas.89.1.60

Bülthoff, I., Mohler, B. J., and Thornton, I. M. (2018). Face recognition of full-bodied avatars by active observers in a virtual environment. *Vision Res.* 157, 242–251. doi: 10.1016/j.visres.2017.12.001

Cao, C., Bradley, D., Zhou, K., and Beeler, T. (2015). Real-time high-fidelity facial performance capture. *ACM Trans. Graph.* 34, 1–9. doi: 10.1145/2766943

Chang, Y., Vieira, M., Turk, M., and Velho, L. (2005). "Automatic 3D facial expression analysis in videos," in *Proceedings of the International Workshop on Analysis and Modeling of Faces and Gestures* (Berlin: Springer), 293–307. doi: 10.1007/11564386_23

Chelnokova, O., and Laeng, B. (2011). Three-dimensional information in face recognition: an eye-tracking study. *J. Vis.* 11:27. doi: 10.1167/11.13.27

Chen, S. Y., Gao, L., Lai, Y. K., Rosin, P. L., and Xia, S. (2018). "Real-time 3D face reconstruction and gaze tracking for virtual reality," in *Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Tuebingen: IEEE), 525–526.

Chen, Y., Wu, H., Shi, F., Tong, X., and Chai, J. (2013). "Accurate and robust 3D facial capture using a single RGBD camera," in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, Sydney.

Cheng, S., Kotsia, I., Pantic, M., and Zafeiriou, S. (2018). "4dfab: a large scale 4d database for facial expression analysis and biometric applications," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, UT.

Chessa, M., Maiello, G., Klein, L. K., Paulun, V. C., and Solari, F. (2019). "Grasping objects in immersive Virtual Reality," in *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Houston, TX.

Christie, F., and Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Mem. Cogn.* 26, 780–790. doi: 10.3758/bf03211397

Chrysos, G. G., Antonakos, E., Snape, P., Asthana, A., and Zafeiriou, S. (2018). A comprehensive performance evaluation of deformable face tracking "In-the-wild". *Int. J. Comput. Vis.* 126, 198–232. doi: 10.1007/s11263-017-0999-5

Cofer, D., Cymbalyuk, G., Reid, J., Zhu, Y., Heitler, W. J., and Edwards, D. H. (2010). AnimatLab: a 3D graphics environment for neuromechanical simulations. *J. Neurosci. Methods* 187, 280–288. doi: 10.1016/j.jneumeth.2010.01.005

Cohn, J. F., and De la Torre, F. (2014). "Automated face analysis for affective," in *The Oxford Handbook of Affective Computing*, eds R. Calvo, S. D'Mello, J. Gratch, and A. Kappas, (New York, NY: Oxford University Press), 131–150.

Cohn, J. F., and Schmidt, K. (2003). "The timing of facial motion in posed and spontaneous smiles," in *Proceedings of the 2003 International Conference on Active Media Technology*, Chongqing, 57–69.

Cosker, D., Krumhuber, E., and Hilton, A. (2011). "A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling," in *Proceedings of the 2011 International Conference on Computer Vision*, Berlin, 2296–2303.

Cunningham, D. W., and Wallraven, C. (2009). Dynamic information for the recognition of conversational expressions. *J. Vis.* 9, 7–17. doi: 10.1167/9.13.7

Danelakis, A., Theoharis, T., and Pratikakis, I. (2018). Action unit detection in 3D facial videos with application in facial expression retrieval and recognition. *Multimed. Tools Appl.* 77, 24813–24841. doi: 10.1007/s11042-018-5699-9

de Beeck, H. O., Wagemans, J., and Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat. Neurosci.* 4, 1244–1252. doi: 10.1038/nn767

Dehmoobadsharifabadi, A., and Farivar, R. (2016). Are face representations depth cue invariant? *J. Vis.* 16:6. doi: 10.1167/16.8.6

Dennett, H. W., McKone, E., Edwards, M., and Susilo, T. (2012). Face aftereffects predict individual differences in face recognition ability. *Psychol. Sci.* 23, 1279–1287. doi: 10.1177/0956797612446350

Dobs, K., Bülthoff, I., and Schultz, J. (2018). Use and usefulness of dynamic face stimuli for face perception studies—A review of behavioral findings and methodology. *Front. Psychol.* 9. doi: 10.3389/fpsyg.2018.01355

Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., and Sejnowski, T. J. (1999). Classifying facial actions. *IEEE Trans. Pattern Anal. Mach. Intell.* 21, 974–989. doi: 10.1109/34.799905

Egger, B., Smith, W., Tewari, A., Wuhrer, S., Zollhöfer, M., Beeler, T., et al. (2019). 3D Morphable Face Models – Past, Present and Future. *arXiv* 1909.01815. Available online at: https://arxiv.org/abs/1909.01815 (accessed June 20, 2020).

Ekman, P. (1982). What emotion categories or dimensions can observers judge from facial behavior? *Emot. Hum. Face* 39–55.

Ekman, P., and Friesen, W. V. (1978). *Manual for the Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press.

Ekmans, P., and Rosenberg, E. L. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford: Oxford University Press.

Ekrami, O., Claes, P., White, J. D., Zaidi, A. A., Shriver, M. D., and Van Dongen, S. (2018). Measuring asymmetry from high-density 3D surface scans: an application to human faces. *PLoS One* 13:e0207895 doi: 10.1371/journal.pone.0207895

Elfenbein, H. A., and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychol. Bull.* 128, 203–235. doi: 10.1037/0033-2909.128.2.203

Eng, Z., Yick, Y., Guo, Y., Xu, H., Reiner, M., Cham, T., et al. (2017). 3D faces are recognized more accurately and faster than 2D faces, but with similar inversion effects. *Vision Res.* 138, 78–85. doi: 10.1016/j.visres.2017.06.004

Ewbank, M. P., and Andrews, T. J. (2008). Differential sensitivity for viewpoint between familiar and unfamiliar faces in human visual cortex. *Neuroimage* 40, 1857–1870. doi: 10.1016/j.neuroimage.2008.01.049

Faltemier, T. C., Bowyer, K. W., and Flynn, P. J. (2008). Using multi-instance enrollment to improve performance of 3D face recognition. *Comput. Vis. Image Underst.* 112, 114–125. doi: 10.1016/j.cviu.2008.01.004

Farivar, R. (2009). Dorsal–ventral integration in object recognition. *Brain Res. Rev.* 61, 144–153. doi: 10.1016/j.brainresrev.2009.05.006

Farivar, R., Blanke, O., and Chaudhuri, A. (2009). Dorsal–ventral integration in the recognition of motion-defined unfamiliar faces. *J. Neurosci.* 29, 5336–5342. doi: 10.1523/jneurosci.4978-08.2009

Fasolt, V., Holzleitner, I. J., Lee, A. J., O'Shea, K. J., and DeBruine, L. M. (2019). Contribution of shape and surface reflectance information to kinship detection in 3D face images. *J. Vis.* 19:9. doi: 10.1167/19.12.9

Favelle, S., Hill, H., and Claes, P. (2017). About face: matching unfamiliar faces across rotations of view and lighting. *Iperception* 8:2041669517744221.

Favelle, S., and Palmisano, S. (2018). View specific generalisation effects in face recognition: front and yaw comparison views are better than pitch. *PLoS One* 13:e0209927. doi: 10.1371/journal.pone.0209927

Favelle, S. K., Palmisano, S., and Avery, G. (2011). Face viewpoint effects about three axes: the role of configural and featural processing. *Perception* 40, 761–784. doi: 10.1068/p6878

Feng, G., Zhao, Y., Tian, X., and Gao, Z. (2014). A new 3-dimensional dynamic quantitative analysis system of facial motion: an establishment and reliability test. *PLoS One* 9:e113115. doi: 10.1371/journal.pone.0113115

Feng, Z.-H., and Kittler, J. (2018). Advances in facial landmark detection. *Biom. Technol. Today* 2018, 8–11. doi: 10.1016/s0969-4765(18)30038-9

Fiorentini, C., and Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *J. Vis.* 11:17. doi: 10.1167/11.3.17

Foster, D. H., and Gilson, S. J. (2002). Recognizing novel three–dimensional objects by summing signals from parts and views. *Proc. R. Soc. Lond. B Biol. Sci.* 269, 1939–1947. doi: 10.1098/rspb.2002.2119

Freud, E., Macdonald, S. N., Chen, J., Quinlan, D. J., Goodale, M. A., and Culham, J. C. (2018a). Getting a grip on reality: grasping movements directed to real objects and images rely on dissociable neural representations. *Cortex* 98, 34–48. doi: 10.1016/j.cortex.2017.02.020

Freud, E., Robinson, A. K., and Behrmann, M. (2018b). More than action: the dorsal pathway contributes to the perception of 3-d structure. *J. Cogn. Neurosci.* 30, 1047–1058. doi: 10.1162/jocn_a_01262

Gaeta, M., Iovane, G., and Sanguineto, E. (2006). A 3D geometric approach to face detection and facial expression recognition. *J. Discrete Math. Sci. Cryptogr.* 9, 39–53. doi: 10.1080/09720529.2006.10698059

Garrido, P., Valgaert, L., Wu, C., and Theobalt, C. (2013). Reconstructing detailed dynamic face geometry from monocular video. *ACM Trans. Graph.* 32, 1–10. doi: 10.1145/2508363.2508380

Gill, D., Garrod, O. G. B., Jack, R. E., and Schyns, P. G. (2014). Facial movements strategically camouflage involuntary social signals of face morphology. *Psychol. Sci.* 25, 1079–1086. doi: 10.1177/0956797614522274

Goeleven, E., De Raedt, R., Leyman, L., and Verschuere, B. (2008). The Karolinska directed emotional faces: a validation study. *Cogn. Emot.* 22, 1094–1118. doi: 10.1080/02699930701626582

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *Am. Psychol.* 48, 26–34. doi: 10.1037/0003-066x.48.1.26

Gong, B., Wang, Y., Liu, J., and Tang, X. (2009). "Automatic facial expression recognition on a single 3D face by exploring shape deformation," in *Proceedings of the 2009 ACM International Conference on Multimedia*, Beijing.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24, 187–203. doi: 10.1016/s0896-6273(00)80832-6

Guo, H., Zhang, X. H., Liang, J., and Yan, W. J. (2018). The dynamic features of lip corners in genuine and posed smiles. *Front. Psychol.* 9:202. doi: 10.3389/fpsyg.2018.00202

Gupta, S., Castleman, K. R., Markey, M. K., and Bovik, A. C. (2010). "Texas 3D face recognition database," in *Proceedings of the 2010 IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI)*, Austin, TX, 97–100.

Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., et al. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *J. Neurosci. Methods* 115, 137–143. doi: 10.1016/s0165-0270(02)00006-7

Häkkinen, J., Kawai, T., Takatalo, J., Leisti, T., Radun, J., Hirsaho, A., et al. (2008). "Measuring stereoscopic image quality experience with interpretation based quality methodology," in *Proceedings of the SPIE Image Quality and System Performance V*, eds S. P. Fernand, and F. Gaykema, (San Jose, CA: SPIE).

Halliday, D. W. R., MacDonald, S. W. S., Sherf, S. K., and Tanaka, J. W. (2014). A reciprocal model of face recognition and Autistic traits: evidence from an individual differences perspective. *PLoS One* 9:e94013. doi: 10.1371/journal.pone.0094013

Hallinan, P. W., Gordon, G., Yuille, A. L., Giblin, P., and Mumford, D. (1999). *Two-and Three-Dimensional Patterns of the Face*. Natick, MA: AK Peters.

Hancock, P. J., Bruce, V., and Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends Cogn. Sci.* 4, 330–337. doi: 10.1016/s1364-6613(00)01519-9

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/s1364-6613(00)01482-0

Hayes, T., Morrone, M. C., and Burr, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception* 15, 595–602. doi: 10.1068/p150595

Heeger, D. J., Boynton, G. M., Demb, J. B., Seidemann, E., and Newsome, W. T. (1999). Motion opponency in visual cortex. *J. Neurosci.* 19, 7162–7174. doi: 10.1523/jneurosci.19-16-07162.1999

Heseltine, T., Pears, N., and Austin, J. (2008). Three-dimensional face recognition using combinations of surface feature map subspace components. *Image Vis. Comput.* 26, 382–396. doi: 10.1016/j.imavis.2006.12.008

Hill, H., and Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 986–1004. doi: 10.1037/0096-1523.22.4.986

Hill, H., and Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Curr. Biol.* 11, 880–885. doi: 10.1016/s0960-9822(01)00243-3

Hill, H., Schyns, P. G., and Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition* 62, 201–222. doi: 10.1016/s0010-0277(96)00785-8

Hong, L. C., Ward, J., and Young, A. W. (2006). Transfer between two-and three-dimensional representations of faces. *Vis. Cogn.* 13, 51–64. doi: 10.1080/13506280500143391

Hsieh, P.-L., Ma, C., Yu, J., and Li, H. (2015). "Unconstrained realtime facial performance capture," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA.

Itier, R. J., and Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. *Neuroimage* 15, 353–372. doi: 10.1006/nimg.2001.0982

Jack, R. E., Garrod, O. G., and Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Curr. Biol.* 24, 187–192. doi: 10.1016/j.cub.2013.11.064

Jack, R. E., and Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annu. Rev. Psychol.* 68, 269–297. doi: 10.1146/annurev-psych-010416-044242

Jandova, M., and Urbanova, P. (2018). Sexual dimorphism in human facial expressions by 3D surface processing. *Homo J. Comp. Hum. Biol.* 69, 98–109. doi: 10.1016/j.jchb.2018.06.002

Jeni, L. A., Cohn, J. F., and Kanade, T. (2017). Dense 3d face alignment from 2d video for real-time use. *Image Vis. Comput.* 58, 13–24. doi: 10.1016/j.imavis.2016.05.009

Johnston, P., Mayes, A., Hughes, M., and Young, A. W. (2013). Brain networks subserving the evaluation of static and dynamic facial expressions. *Cortex* 49, 2462–2472. doi: 10.1016/j.cortex.2013.01.002

Jones, A. L., Kramer, R. S., and Ward, R. (2012). Signals of personality and health: the contributions of facial shape, skin texture, and viewing angle. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 1353. doi: 10.1037/a0027078

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., and Akamatsu, S. (2013). Dynamic properties influence the perception of facial expressions. *Perception* 42, 1266–1278. doi: 10.1068/p3131n

Kamitani, Y., and Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.* 16, 1096–1102. doi: 10.1016/j.cub.2006.04.003

Kanade, T., Cohn, J. F., and Tian, Y. (2000). "Comprehensive database for facial expression analysis," in *Proceedings of the 2000 Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, Grenoble, 46–53.

Kanwisher, N., and Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 2109–2128. doi: 10.1098/rstb.2006.1934

Kätsyri, J. (2006). *Human Recognition of Basic Emotions From Posed and Animated Dynamic Facial Expressions*. Ph.D. thesis, Helsinki University of Technology, Espoo.

Kaulard, K., Cunningham, D. W., Bülthoff, H. H., and Wallraven, C. (2012). The MPI facial expression database—a validated database of emotional and conversational facial expressions. *PLoS One* 7:e32321. doi: 10.1371/journal. pone.0032321

Kinect Xbo, (2019). *Xbox Family of Sites*. Available at: https://www.xbox.com/en-GB/kinect (accessed August 30, 2019).

Kleiner, M., Wallraven, C., and Bülthoff, H. H. (2004). *The MPI VideoLab-A System for High Quality Synchronous Recording of Video and Audio from Multiple Viewpoints*. Tübingen: MPI.

Knappmeyer, B., Thornton, I. M., and Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Res.* 43, 1921–1936. doi: 10.1016/s0042-6989(03)00236-0

Krumhuber, E. G., Kappas, A., and Manstead, A. S. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Krumhuber, E. G., Skora, L., Küster, D., and Fou, L. (2017). A review of dynamic datasets for facial expression research. *Emot. Rev.* 9, 280–292. doi: 10.1177/1754073916670022

Lamberti, F., Cannavo, A., and Montuschi, P. (2020). Is immersive virtual reality the ultimate interface for 3D animators? *Computer* 53, 36–45. doi: 10.1109/mc. 2019.2908871

Lambooij, M., IJsselsteijn, W., Bouwhuis, D. G., and Heynderickx, I. (2011). Evaluation of stereoscopic images: beyond 2D quality. *IEEE Trans. Broadcast.* 57, 432–444. doi: 10.1109/tbc.2011.2134590

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., and Van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cogn. Emot.* 24, 1377–1388. doi: 10.1080/02699930903485076

Lee, Y., Matsumiya, K., and Wilson, H. R. (2006). Size-invariant but viewpoint-dependent representation of faces. *Vision Res.* 46, 1901–1910. doi: 10.1016/j. visres.2005.12.008

Levin, D. T., Takarae, Y., Miner, A. G., and Keil, F. (2001). Efficient visual search by category: specifying the features that mark the difference between artifacts and animals in preattentive vision. *Percept. Psychophys.* 63, 676–697. doi: 10.3758/bf03194429

Li, H., Yu, J., Ye, Y., and Bregler, C. (2013). Realtime facial animation with on-the-fly correctives. *ACM Trans. Graph.* 32, 1–10.

Li, Y., Wang, Y., Liu, J., and Hao, W. (2018). Expression-insensitive 3D face recognition by the fusion of multiple subject-specific curves. *Neurocomputing* 275, 1295–1307. doi: 10.1016/j.neucom.2017.09.070

Liu, C. H., and Chaudhuri, A. (1997). Face recognition with multi-tone and two-tone photographic negatives. *Perception* 26, 1289–1296. doi: 10.1068/p261289

Liu, C. H., and Chaudhuri, A. (2003). Face recognition with perspective transformation. *Vision Res.* 43, 2393–2402. doi: 10.1016/s0042-6989(03)00429-2

Liu, C. H., Collin, C. A., Farivar, R., and Chaudhuri, A. (2005). Recognizing faces defined by texture gradients. *Percept. Psychophys.* 67, 158–167. doi: 10.3758/bf03195019

Liu, C. H., and Ward, J. (2006). The use of 3D information in face recognition. *Vision Res.* 46, 768–773. doi: 10.1016/j.visres.2005.10.008

Liu, F., Van Der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., et al. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet.* 8:e1002932. doi: 10.1371/journal. pgen.1002932

Logothetis, N. K., and Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cereb. Cortex* 5, 270–288. doi: 10.1093/cercor/5.3.270

Long, B., Konkle, T., Cohen, M. A., and Alvarez, G. A. (2016). Mid-level perceptual features distinguish objects of different real-world sizes. *J. Exp. Psychol. Gen.* 145, 95–109. doi: 10.1037/xge0000130

Lu, X., Chen, H., and Jain, A. K. (2006). "Multimodal facial gender and ethnicity identification," in *Advances in Biometrics. ICB 2006. Lecture Notes in Computer Science*, Vol. 3832, eds D. Zhang, and A. K. Jain, (Berlin: Springer).

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. *Paper Presented at the Computer Vision and Pattern Recognition Workshops (CVPRW), 2010*, San Francisco, CA.

Lueschow, A., Miller, E. K., and Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cereb. Cortex* 4, 523–531. doi: 10.1093/cercor/4.5.523

Lundqvist, D., Flykt, A., and Öhman, A. (1998). *The Karolinska Directed Emotional Faces (KDEF)*. Stockholm: Karolinska Institutet.

Lv, C., Wu, Z., Wang, X., Dan, Z., and Zhou, M. (2020). Ethnicity classification by the 3D discrete landmarks model measure in Kendall shape space. *Pattern Recognit. Lett.* 129, 26–32. doi: 10.1016/j.patrec.2019.10.035

Mahmoud, M., Baltrušaitis, T., Robinson, P., and Riek, L. D. (2011). "3D corpus of spontaneous complex mental states," in *Affective Computing and Intelligent Interaction. ACII 2011. Lecture Notes in Computer Science*, Vol. 6974, eds S. D'Mello, A. Graesser, B. Schuller, and J. C. Martin, (Berlin: Springer).

Marcolin, F., and Vezzetti, E. (2017). Novel descriptors for geometrical 3D face analysis. *Multimed. Tools Appl.* 76, 13805–13834. doi: 10.1007/s11042-016-3741-3

Marosi, C., Fodor, Z., and Csukly, G. (2019). From basic perception deficits to facial affect recognition impairments in schizophrenia. *Sci. Rep.* 9:8958.

Marschner, S. R., Guenter, B., and Raghupathy, S. (2000). Modeling and rendering for realistic facial animation. *Paper Presented at the Rendering Techniques 2000*, Vienna.

Matsumoto, D. (1988). *Japanese and Caucasian Facial Expressions of Emotion (JACFEE) and Neutral Faces (JACNeuF)*. San Francisco, CA: San Francisco State University.

Mendes, D., Caputo, F. M., Giachetti, A., Ferreira, A., and Jorge, J. (2018). A survey on 3d virtual object manipulation: from the desktop to immersive virtual environments. *Comput. Graph. Forum* 38, 21–45. doi: 10.1111/cgf.13390

Mignault, A., and Chaudhuri, A. (2003). The many faces of a neutral face: head tilt and perception of dominance and emotion. *J. Nonverb. Behav.* 27, 111–132.

Moreno, A. (2004). "GavabDB: a 3D face database," in *Proceedings of the 2nd COST275 Workshop on Biometrics on the Internet*, Vigo.

Motley, M. T., and Camden, C. T. (1988). Facial expression of emotion: a comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *West. J. Speech Commun.* 52, 1–22. doi: 10.1080/10570318809389622

Murphy, A. P., and Leopold, D. A. (2019). A parameterized digital 3D model of the Rhesus macaque face for investigating the visual processing of social cues. *J. Neurosci. Methods* 324:108309. doi: 10.1016/j.jneumeth.2019.06.001

Nguyen, A. T. T., and Clifford, C. W. G. (2019). Gazing into space: systematic biases in determining another's fixation distance from gaze vergence in upright and inverted faces. *J. Vis.* 19:5. doi: 10.1167/19.11.5

Onitsuka, T., Niznikiewicz, M. A., Spencer, K. M., Frumin, M., Kuroki, N., Lucia, L. C., et al. (2006). Functional and structural deficits in brain regions subserving face perception in schizophrenia. *American Journal of Psychiatry* 163, 455–462. doi: 10.1176/appi.ajp.163.3.455

Oram, M., and Perrett, D. (1992). Time course of neural responses discriminating different views of the face and head. *J. Neurophysiol.* 68, 70–84. doi: 10.1152/jn. 1992.68.1.70

O'Toole, A., Roark, D., and Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn. Sci.* 6, 261–266. doi: 10.1016/s1364-6613(02)01908-3

O'Toole, A. J., Harms, J., Snow, S. L., Hurst, D. R., Pappas, M. R., Ayyad, J. H., et al. (2005). A video database of moving faces and people. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 812–816. doi: 10.1109/tpami.2005.90

O'Toole, A. J., Phillips, P. J., Weimer, S., Roark, D. A., Ayyad, J., Barwick, R., et al. (2011). Recognizing people from dynamic and static faces and bodies: dissecting identity with a fusion approach. *Vision Res.* 51, 74–83. doi: 10.1016/j.visres. 2010.09.035

Otta, E., Lira, B. B. P., Delevati, N. M., Cesar, O. P., and Pires, C. S. G. (1994). The effect of smiling and of head tilting on person perception. *J. Psychol.* 128, 323–331. doi: 10.1080/00223980.1994.9712736

Pantic, M., and Rothkrantz, L. J. (2004). Facial action recognition for facial expression analysis from static face images. *IEEE Trans. Syst. Man Cybern. Part B* 34, 1449–1461. doi: 10.1109/tsmcb.2004.825931

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. (2005). "Web-based database for facial expression analysis," in *Proceedings of the 2005 IEEE International Conference on Multimedia and Expo*, Amsterdam.

Patil, H., Kothari, A., and Bhurchandi, K. (2015). 3-D face recognition: features, databases, algorithms and challenges. *Artif. Intell. Rev.* 44, 393–441. doi: 10. 1007/s10462-015-9431-0

Perrett, D. I., Hietanen, J. K., Oram, M. W., and Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. R. Soc. Lond. B* 335, 23–30. doi: 10.1098/rstb.1992.0003

Perrett, D. I., Smith, P., Potter, D., Mistlin, A., Head, A., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. Lond. B* 223, 293–317. doi: 10.1098/rspb.1985.0003

Pike, G. E., Kemp, R. I., Towell, N. A., and Phillips, K. C. (1997). Recognizing moving faces: the relative contribution of motion and perspective view information. *Vis. Cogn.* 4, 409–438. doi: 10.1080/713756769

Pizlo, Z., Sawada, T., Li, Y., Kropatsch, W. G., and Steinman, R. M. (2010). New approach to the perception of 3D shape based on veridicality, complexity, symmetry and volume. *Vision Res.* 50, 1–11. doi: 10.1016/j.visres.2009.09.024

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., and Vuilleumier, P. (2005). View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity: an event-related fMRI study. *Neuroimage* 24, 1214–1224. doi: 10.1016/j.neuroimage.2004.10.038

Pujades, S., Mohler, B., Thaler, A., Tesch, J., Mahmood, N., Hesse, N., et al. (2019). The virtual caliper: rapid creation of metrically accurate avatars from 3D measurements. *IEEE Trans. Visual. Comput. Graph.* 25, 1887–1897. doi: 10.1109/tvcg.2019.2898748

Ramakrishna, V., Kanade, T., and Sheikh, Y. (2012). Reconstructing 3d human pose from 2d image landmarks. *Paper Presented at the European Conference on Computer Vision*, Amsterdam.

Reale, M., Zhang, X., and Yin, L. (2013). "Nebula feature: a space-time feature for posed and spontaneous 4D facial behavior analysis," in *Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Shanghai.

Rhodes, G., Jeffery, L., Taylor, L., Hayward, W. G., and Ewing, L. (2014). Individual differences in adaptive coding of face identity are linked to individual differences in face recognition ability. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 897–903. doi: 10.1037/a0035939

Rossion, B., and Michel, C. (2018). Normative accuracy and response time data for the computerized Benton Facial Recognition Test (BFRT-c). *Behav. Res. Methods* 50, 2442–2246.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol. Bull.* 115, 102–141. doi: 10.1037/0033-2909.115.1.102

Russell, R., Biederman, I., Nederhouser, M., and Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision Res.* 47, 157–165. doi: 10.1016/j.visres.2006.11.002

Russell, R., Duchaine, B., and Nakayama, K. (2009). Super-recognizers: people with extraordinary face recognition ability. *Psychon. Bull. Rev.* 16, 252–257. doi: 10.3758/pbr.16.2.252

Russell, R., and Sinha, P. (2007). Real-world face recognition: the importance of surface reflectance properties. *Perception* 36, 1368–1374. doi: 10.1068/p5779

Russell, R., Sinha, P., Biederman, I., and Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from Contrast Negation. *Perception* 35, 749–759. doi: 10.1068/p5490

Sadhya, D., and Singh, S. K. (2019). A comprehensive survey of unimodal facial databases in 2D and 3D domains. *Neurocomputing* 358, 188–210. doi: 10.1016/j.neucom.2019.05.045

Saito, S., Li, T., and Li, H. (2016). "Real-time facial segmentation and performance capture from RGB input," in *ECCV, 2016. LNCS*, Vol. 9912, eds B. Leibe, J. Matas, N. Sebe, and M. Welling, (Cham: Springer)

Sandbach, G., Zafeiriou, S., Pantic, M., and Yin, L. (2012). Static and dynamic 3D facial expression recognition: a comprehensive survey. *Image Vis. Comput.* 30, 683–697. doi: 10.1016/j.imavis.2012.06.005

Sáry, G., Vogels, R., and Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* 260, 995–997. doi: 10.1126/science.8493538

Sasaoka, T., Asakura, N., and Inui, T. (2019). Ease of hand rotation during active exploration of views of a 3-D object modulates view generalization. *Exp. Brain Res.* 237, 939–951. doi: 10.1007/s00221-019-05474-6

Savran, A., Alyüz, N., Dibeklioǧlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., et al. (2008). "Bosphorus database for 3D face analysis," in *Biometrics and Identity Management. BioID 2008. Lecture Notes in Computer Science*, Vol. 5372, eds B. Schouten, N. C. Juul, A. Drygajlo, and M. Tistarelli, (Berlin: Springer)

Savran, A., Sankur, B., and Bilge, M. T. (2012). Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units. *Pattern Recogn.* 45, 767–782. doi: 10.1016/j.patcog.2011.07.022

Schwaninger, A., and Yang, J. (2011). The application of 3D representations in face recognition. *Vision Res.* 51, 969–977. doi: 10.1016/j.visres.2011.02.006

Seyama, J. I., and Nagayama, R. (2007). The Uncanny Valley: effect of realism on the impression of artificial human faces. *Presence* 16, 337–351. doi: 10.1162/pres.16.4.337

Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171, 701–703. doi: 10.1126/science.171.3972.701

Shi, F., Wu, H.-T., Tong, X., and Chai, J. (2014). Automatic acquisition of high-fidelity facial performances using monocular videos. *ACM Trans. Graph.* 33, 1–13. doi: 10.1145/2661229.2661290

Smeets, D., Claes, P., Hermans, J., Vandermeulen, D., and Suetens, P. (2012). A comparative study of 3-D face recognition under expression variations. *IEEE Trans. Syst. Man. Cybern. Part C* 42, 710–727. doi: 10.1109/tsmcc.2011.2174221

Smith, H. J., and Neff, M. (2018). "Communication behavior in embodied virtual reality," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, QC.

Stolz, C., Endres, D., and Mueller, E. M. (2019). Threat-conditioned contexts modulate the late positive potential to faces—A mobile EEG/virtual reality study. *Psychophysiology* 56, e13308. doi: 10.1111/psyp.13308

Stratou, G., Ghosh, A., Debevec, P., and Morency, L. P. (2011). "Effect of illumination on automatic expression recognition: a novel 3D relightable facial database," in *Proceedings of the 9th international conference on Automatic Face and Gesture Recognition*, Santa Barbara, CA, 611–618.

Sun, Y., Reale, M., and Yin, L. (2008). "Recognizing partial facial action units based on 3d dynamic range data for facial expression recognition," in *Proceedings of the 2008 8th IEEE International Conference on Automatic Face Gesture Recognition*, Amsterdam, 1–8.

Suwajanakorn, S., Kemelmacher-Shlizerman, I., and Seitz, S. M. (2014). "Total moving face reconstruction," in *Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*, Vol. 8692, eds D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, (Cham: Springer).

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *J. Exp. Psychol. Gen.* 130, 534–543. doi: 10.1037/0096-3445.130.3.534

Tarnowski, P., Kołodziej, M., Majkowski, A., and Rak, R. J. (2017). Emotion recognition using facial expressions. *Procedia Comput. Sci.* 108, 1175–1184.

Tarr, M. J. (1995). Rotating objects to recognize them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychon. Bull. Rev.* 2, 55–82. doi: 10.3758/bf03214412

Tcherkassof, A., Bollon, T., Dubois, M., Pansu, P., and Adam, J. M. (2007). Facial expressions of emotions: a methodological contribution to the study of spontaneous and dynamic emotional faces. *Eur. J. Soc. Psychol.* 37, 1325–1345. doi: 10.1002/ejsp.427

Thies, R. (2015). Real-time expression transfer for facial reenactment. *ACM Trans. Graph.* 34, 1–14. doi: 10.1145/2816795.2818056

Thies, R. (2017). Demo of FaceVR: real-time facial reenactment and eye gaze control in virtual reality. *Paper Presented at the ACM SIGGRAPH 2017 Emerging Technologies*, Los Angeles, CA.

Tootell, R. B., Reppas, J. B., Dale, A. M., Look, R. B., Sereno, M. I., Malach, R., et al. (1995). Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature* 375, 139–141. doi: 10.1038/375139a0

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res.* 168, 242–249. doi: 10.1016/j.psychres. 2008.05.006

Trautmann, S. A., Fehr, T., and Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res.* 1284, 100–115. doi: 10.1016/j.brainres.2009.05.075

Trebicky, V., Fialova, J., Stella, D., Sterbova, Z., Kleisner, K., and Havlicek, J. (2018). 360 degrees of facial perception: congruence in perception of frontal portrait, profile, and rotation photographs. *Front. Psychol.* 9:2405. doi: 10.3389/fpsyg. 2018.02405

Tsalakanidou, F., and Malassiotis, S. (2010). Real-time 2D+3D facial action and expression recognition. *Pattern Recogn.* 43, 1763–1775. doi: 10.1016/j.patcog. 2009.12.009

Tsao, D. Y., Vanduffel, W., Sasaki, Y., Fize, D., Knutsen, T. A., Mandeville, J. B., et al. (2003). Stereopsis activates V3A and caudal intraparietal areas in macaques and humans. *Neuron* 39, 555–568. doi: 10.1016/s0896-6273(03)00459-8

Tulyakov, S., Vieriu, R.-L., Sangineto, E., and Sebe, N. (2015). "Facecept3d: real time 3d face tracking and analysis," in *Proceedings of the 2015 IEEE International Conference on Computer Vision Workshops*, Santiago.

Urbanová, P., Ferková, Z., Jandová, M., Jurda, M., Èerrı, D., and Sochor, J. (2018). Introducing the FIDENTIS 3D Face Database. *Anthropol. Rev.* 81, 202–223. doi: 10.2478/anre-2018-0016

Valstar, M., and Pantic, M. (2010). "Induced disgust, happiness and surprise: an addition to the MMI facial expression database," in *Proceedings of the 3rd International Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect*, Valletta, 65.

Van der Linde, I., and Watson, T. (2010). A combinatorial study of pose effects in unfamiliar face recognition. *Vision Res.* 50, 522–533. doi: 10.1016/j.visres.2009. 12.012

Van Der Schalk, J., Hawk, S. T., Fischer, A. H., and Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion* 11, 907–920. doi: 10.1037/a0023853

Vanrie, J., Willems, B., and Wagemans, J. (2001). Multiple routes to object matching from different viewpoints: mental rotation versus invariant features. *Perception* 30, 1047–1056. doi: 10.1068/p3200

Vishwanath, D., and Kowler, E. (2004). Saccadic localization in the presence of cues to three-dimensional shape. *J. Vis.* 4, 445–458.

Vlasic, D. (2005). Face transfer with multilinear models. *ACM Trans. Graph.* 24, 426–433. doi: 10.1145/1073204.1073209

Wallbott, H. G., and Scherer, K. R. (1986). Cues and channels in emotion recognition. *J. Pers. Soc. Psychol.* 51, 690–699. doi: 10.1037/0022-3514.51.4.690

Wallhoff, F. (2004). *Fgnet-Facial Expression and Emotion Database*. München: Technische Universität München.

Wallis, G., and Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4800–4804. doi: 10.1073/ pnas.071028598

Wallraven, C., Schwaninger, A., Schuhmacher, S., and Bülthoff, H. H. (2002). "View-based recognition of faces in man and machine: Re- visiting inter-extra-ortho," in *Proceedings of the International Workshop on Biologically Motivated Computer Vision*, (Berlin: Springer), 651–660. doi: 10.1007/3-540-361 81-2_65

Wang, G., Tanaka, K., and Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272, 1665–1668. doi: 10.1126/science.272.5268.1665

Wang, P., Kohler, C., Barrett, F., Gur, R., Gur, R., and Verma, R. (2007). "Quantifying facial expression abnormality in schizophrenia by combining 2D and 3D features," in *Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN.

Wang, Y., Huang, X., Lee, C., Zhang, S., Li, Z., Samaras, D., et al. (2004). High resolution acquisition, learning and transfer of dynamic 3D facial expressions. *Comput. Graph. Forum* 23, 677–686. doi: 10.1111/j.1467-8659.2004.00 800.x

Weber, R., Soladié, C., and Séguier, R. (2018). "A survey on databases for facial expression analysis," in *Proceedings of the 13th h International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 5: VISAPP*, Cesson-Sévigné.

Wehrle, T., Kaiser, S., Schmidt, S., and Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *J. Pers. Soc. Psychol.* 78, 105–119. doi: 10.1037/0022-3514.78.1.105

Weigelt, S., Koldewyn, K., and Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: a review of behavioral studies. *Neurosci. Biobehav. Rev.* 36, 1060–1084. doi: 10.1016/j.neubiorev.2011.12.008

Weise, T., Li, H., Van Gool, L., and Pauly, M. (2009). "Face/off: Live facial puppetry," in *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, New York, NY.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., et al. (2010). Human face recognition ability is specific and highly heritable. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5238–5241. doi: 10.1073/pnas.0913053107

Yin, L., Wei, X., Sun, Y., Wang, J., and Rosato, M. J. (2006). "A 3D facial expression database for facial behavior research," in *Proceedings of the 2006 7th International Conference on Automatic Face and Gesture Recognition*, Southampton.

Yip, A. W., and Sinha, P. (2002). Contribution of color to face recognition. *Perception* 31, 995–1003. doi: 10.1068/p3376

Zachariou, V., Del Giacco, A. C., Ungerleider, L. G., and Yue, X. (2018). Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization. *J. Vis.* 18:3. doi: 10.1167/18.12.3

Zeki, S., Watson, J., Lueck, C., Friston, K. J., Kennard, C., and Frackowiak, R. (1991). A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.* 11, 641–649. doi: 10.1523/jneurosci.11-03-00641.1991

Zhan, J., Garrod, O. G. B., van Rijsbergen, N., and Schyns, P. G. (2019). Modelling face memory reveals task-generalizable representations. *Nat. Hum. Behav.* 3, 817–826. doi: 10.1038/s41562-019-0625-3

Zhang, J., and Fisher, R. B. (2019). 3D Visual passcode: speech-driven 3D facial dynamics for behaviometrics. *Signal Process.* 160, 164–177. doi: 10.1016/j. sigpro.2019.02.025

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., et al. (2013). "A high-resolution spontaneous 3d dynamic facial expression database," in *Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Shanghai.

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., et al. (2014). Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image Vis. Comput.* 32, 692–706. doi: 10.1016/j.imavis. 2014.06.002

Zhang, Y., Prakash, E. C., and Sung, E. (2003). Modeling and animation of individualized faces for 3D facial expression synthesis. *Int. J. Imaging Syst. Technol.* 13, 42–64. doi: 10.1002/ima.10041

Zhong, C., Sun, Z., and Tan, T. (2007). "Robust 3D face recognition using learned visual codebook," in *Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN.

Zhou, S., and Xiao, S. (2018). 3D face recognition: a survey. *Hum. Cent. Comput. Inf. Sci.* 8:35.

Zollhöfer, M., Thies, J., Garrido, P., Bradley, D., Beeler, T., Pérez, P., et al. (2018). State of the art on monocular 3D face reconstruction, tracking, and applications. *Comput. Graph. Forum* 37, 523–550. doi: 10.1111/cgf.13382