Displays 46 (2017) 25–36

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa

User interface for a better eye contact in videoconferencing ☆

Aleš Jaklič, Franc Solina, Luka Šajn *

Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 5 June 2016
Received in revised form 17 October 2016
Accepted 16 December 2016
Available online 25 December 2016

Keywords:
Videoconferencing
Eye contact
Human-computer interface

ABSTRACT

When people talk to each other, eye contact is very important for a trustful and efficient communication. Video-conferencing systems were invented to enable such communication over large distances, recently using mostly Internet and personal computers. Despite low cost of such solutions, a broader acceptance and use of these communication means has not happened yet. One of the most important reasons for this situation is that it is almost impossible to establish eye contact between distant parties on the most common hardware configurations of such videoconferencing systems, where the camera for face capture is usually mounted above the computer monitor, where the face of the correspondent is observed. Different hardware and software solutions to this problem of missing eye contact have been proposed over the years. In this article we propose a simple solution that can improve the subjective feeling of eye contact, which is based on how people perceive 3D scenes displayed on slanted surfaces, and offer some experiments in support of the hypothesis.

1. Introduction

Videoconferencing enables people to communicate face-to-face over remote distances using video and audio telecommunication technology. First videoconferencing systems appeared already in the late 1930s when television was invented. In 1970 AT&T launched the Picturephone service convinced that a video phone would be a huge commercial success. But the Picturephone was a large customer failure due in part to dedicated big phone sets, high cost and, most of all, by making the experience uncomfortably intrusive since the Picturephone was by default always on [1]. However, teleconferencing became widespread in the 1990s with the advent of computer technology when Internet protocol-based videoconferencing made this functionality available to a much larger public as a service on existing equipment. In the 2000s free Internet videoconferencing services such as Skype became available to most users as applications on personal and laptop computers. Now, hand-held mobile devices offer similar videoconferencing services.

Videoconferencing is an important alternative to other means of interpersonal communication in many different application areas when live conversation is needed and when non-verbal or visual information is an important component of conversation. People with deafness or hearing disorder and people with speech disorders or mutism can now use videoconferencing to communicate with each other using sign language.

Despite these large technical advances, mass adoption of videoconferencing has not materialized yet, although futurists have predicted for almost a century that telephone conversations in the future will proceed as face-to-face meetings using audio and video [2]. The first reason for this conservative attitude is that voice communication is often sufficient in many situations while the second reason is that videoconferencing technology still does not offer the same experience as face-to-face communication in real, physical space. One of the major unresolved issues of videoconferencing concerning the user experience is the loss of eye contact between participants of a teleconferencing session [2–4]. The other reasons are that systems for videoconferencing are still complex and different systems are not inter-compatible. Some users are also appearance shy and being on camera hinders their communication capability. The most serious issue, however, is that it is normally very difficult if not impossible to establish eye contact between the users of a videoconferencing system.

1.1. Eye contact

People are very sensitive to the direction of the eye gaze of other people. We are able to determine very accurately if somebody is actually looking at us. Our eyes express our emotions and intentions and they help us direct attention [5]. Cultural norms in different societies dictate when, for how long and in what

☆ This paper was recommended for publication by Richard H.Y. So.

* Corresponding author.
E-mail addresses: ales.jaklic@fri.uni-lj.si (A. Jaklič), franc.solina@fri.uni-lj.si (F. Solina), luka.sajn@fri.uni-lj.si (L. Šajn).

http://dx.doi.org/10.1016/j.displa.2016.12.002
0141-9382/

situations it is appropriate to gaze into another person's eyes. In Japanese society, for example, eye contact can be considered rude, especially eye contact with a superior [6]. Eye contact plays an important role in the broader context of maintaining one's personal space, even in virtual environments [7]. Nontheless, eye contact is very important, especially in communication between individuals, because avoiding eye contact when communicating face-to-face can be associated with deception [8]. The inability of establishing and maintaining eye contact in videoconferencing systems is one of the most serious limitations of such systems [3,4] since the relationship between the ability to accurately perceive eye contact and the efficiency of communication has been firmly established [9].

Therefore, the problem of establishing eye contact may have affected a wider adoption of videoconferencing systems. Gaze patterns, which in general provide an extremely important and rich set of serial signals in face-to-face communication, should therefore be taken into account for videoconferencing design [10].

Why is the establishment of eye contact difficult over videoconferencing systems? During conversation, a person's eyes are usually directed at the center of the computer screen where the videoconferencing partner's face is displayed. The cameras for capturing the video signal, however, are usually mounted or even built-in above the display screen. Therefore, using average sized desktop computer displays at a normal viewing distances, the angle between the viewing direction and the optical axis of the camera is usually between 15 and 20 degrees [11]. If the angle between the line from the camera to the eyes and the line from the eyes to the screen is more than 5 degrees the loss of eye contact is noticeable, and in the case of 15–20 degrees the loss of eye contact is inevitable [12]. An example of this parallax in videoconferencing is shown in Figs. 1 and 2.

The same problem arises also if the face of a video conference correspondent is displayed in a smaller window on a large screen and if the angle between the correspondent's face in that window and the position of the camera is larger than 5 degrees. The problem with small handheld devices with a built-in camera is normally not as severe, since at the normal viewing distance the angle between the lines towards the image of the face and the camera is smaller than 5 degrees. Indeed, studies have shown that videoconferencing where the camera is mounted above the computer monitor, and thus disabling eye contact, is less trusted than a centrally mounted camera, where eye contact is enabled, but even less trusted than just a voice connection or email [8]. Eye contact is simply a nearly ubiquitous method of affirming trust when people communicate face to face. If several people are involved in a teleconferencing session even more issues related to communicating important non-verbal information between participants arise in comparison to a dyadic video mediated communication [2,13]. An excellent overview of gaze perception and the problem of eye contact during videoconferencing is presented by Bohannon et al. [4].

Out of this frustration with existing videoconferencing systems the idea of 3D tele-immersion has arisen [14]. Tele-immersion is an emerging technology that enables users to collaborate remotely by generating realistic 3D avatars in real time and rendering them inside a shared virtual space [15]. A tele-immersive environment thus provides a venue not only for talking between users as in tele-conferencing systems but also for collaborative work on 3D models and even remote teaching of physical activities. Such tele-immersive systems need, on the one end, 3D video capture technologies and, on the other end, a virtual reality display, making these experimental systems at present still unreachable to the general public.

In this article, we propose a solution for obtaining a better perception of establishing eye contact in a typical dyadic, video mediated face-to-face communication scenario, based on psychophysical properties of the human visual perceptual system. We have proposed this solution initially in a conference article in 2011, but without much analysis and discussion about the explanation of the observed effect [16]. We have further described this solution as one of possible applications of dynamic anamorphosis in [17]. In this article we give a broad overview of the problem of missing eye contact in videoconferencing systems and all possible solutions proposed to solve this problem. Additional experiments to illuminate and confirm our proposal are described in this article as well as a more in-depth analysis and a discussion of our proposed solution of the problem of missing eye contact in videoconferencing systems using results from psychophysical literature.

The structure of the rest of the article is as follows: Section 2 is an overview of previous attempts of solving the problem of missing eye contact in videoconferencing systems and on the background of human visual perception related to the issue of eye contact. Section 3 defines our proposal for solving the problem of missing eye contact in videoconferencing systems and describes three experiments that we made in order to back up our proposal. In Section 4 we discuss possible underlaying psychophysical mechanisms that come into play in the context of our proposal to improve the eye contact. Conclusions are in Section 5.

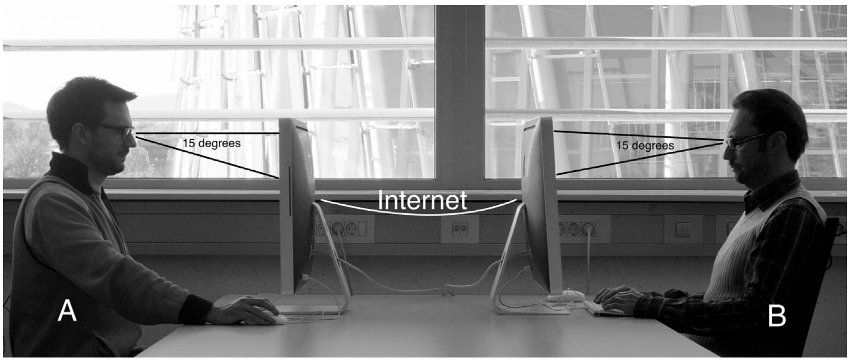
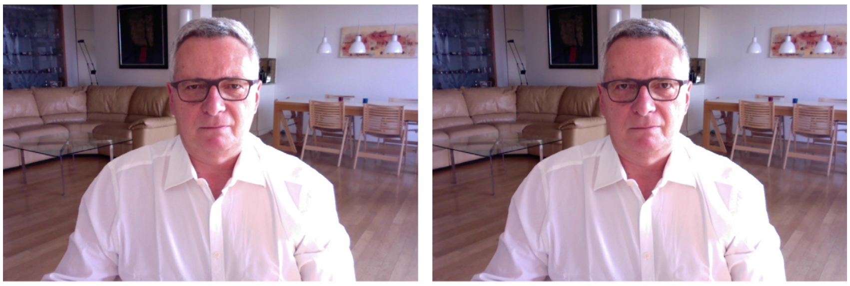


Fig. 1. A typical situation while communicating over a videoconference system over the Internet: person A is looking at the image of person B in the middle of the computer screen in front of him, while on the other end, person B is looking at the image of person A in the middle of his computer screen. Because cameras on both ends are mounted above the computer screens, the eyes of the faces imaged on both screens are looking downwards and thus the eye contact between both parties A and B cannot be established. When using a large computer monitor, the angle between the viewing direction of the user and the line between his eyes and the camera is typically about 15 degrees.



(a) (b)

Fig. 2. Images of a user captured by the camera mounted above a 27-in. computer display at a normal working distance: (a) in the left image, the user is looking into the middle of the display, (b) in the right image, the user is looking straight into the camera. Image (a) on the left illustrates the problem of the missing eye contact in teleconferencing: when the user is looking towards the center of the display, where the eyes of his teleconferencing partner are displayed, his partner on the other side sees the image (a), where the user appears to be looking downwards and not directly into the partner's eyes as in image (b).

2. Background

The problem of the difficulty of establishing eye contact in videoconferencing systems has been known and addressed for many years. The solution to this problem seems obvious, the viewing direction towards the face of our videoconferencing partner must be aligned as closely as possible with the direction of the camera optical axis which is recording our face.

Initially, this problem was addressed with more sophisticated hardware equipment. Later, with more powerful and faster computer processors, software solutions were considered, based primarily on computer vision methods to manipulate the face images. Recently, advances in technology of manufacturing computer display screens seem to offer the best solution in the future.

2.1. Hardware approaches

Initial hardware solutions to the problem of aligning the viewing direction towards the displayed face image with the camera optical axis tried to employ optical devices, such as half mirrors and beamsplitters, to bring the camera behind and into the center of the computer screen [18]. However, such specialized systems were bulky, expensive and generally not accessible to the majority of videoconferencing users.

In recent years, two computer companies that manufacture their own hardware, Apple [19,20] and Sony [21] filed for patents for a new type of LCD computer display that integrates camera image sensors between LCD display cells. In this way, the camera is hidden in or behind the computer screen. This solution has not appeared yet on the market, although, other sensors such as fingerprint scanners, have already been integrated into computer display screens.

2.2. Computer vision approaches

With increased computer power, software solutions for aligning the viewing axis and the camera optical axis, based on the manipulation of the image itself, using computer vision methods, were proposed [22]. To generate a virtual camera view from the middle of the computer screen, the use of anywhere from two [23] to eight cameras have been proposed [24]. These systems are based on the principles of stereo matching and image morphing methods. Since users now expect to use videoconferencing on their own desktop or portable computers using the single built-in camera above the computer display screen, all these solutions seem to be overly complicated.

Even if only a single camera is available, a method was proposed to separately rectify the face image and the image of the eyes with an affine transformation [25]. Other single camera solutions propose to replace the eyes of the user with direct looking eyes [26,27].

New inexpensive hybrid (depth + color) cameras, such as the Kinect, enable the capture of the 3D model of a scene. With the aid of a face tracker, the 3D model of a face can be isolated and then rotated so that just the gaze-corrected image of the face can be seamlessly merged into the original image in real-time [28]. Yet another recent method [29] is similar but uses just a single standard camera. A generic 3D head mesh is deformed to fit the facial features of a videoconference participant in real-time. The 3D model is then used for gaze correction of the head and the corresponding face region is again merged into the original image.

In this article, we propose a third type of solution, based on psychophysical properties of the human visual system, that also uses a single camera and somehow tricks the human perceptual system to have a feeling of a better eye contact just by changing the geometry of the entire picture which is displaying the face of our videoconferencing partner.

2.3. Psychophysics of gaze direction

To be able to establish eye contact, one must be able to estimate quite accurately the gaze direction of the eyes of our visible and close enough corresponding partner. Gaze direction is a vector pointing from the fovea of the looker through the center of the pupil to the gazed-at spot. The eye gaze direction can be determined by the dark-white configuration or the position of the pupil in the visible part of the eye. But the perception of gaze direction depends not only on the dark-white configuration of the eye but also on the head turn of the looker, as demonstrated by Todorović [30]. According to Todorović, the observer-related gaze direction is an additive combination of the observer-related head orientation and the looker-related gaze direction. With this geometrical basis for gaze perception Todorović explained the Mona Lisa effect and the Wollaston effect, the two well known phenomena of gaze direction.

The Mona Lisa effect, which is named after Leonardo da Vinci's painting, is in fact an instance of the differential rotation effect (DRE). The differential rotation effect manifests itself when pictures are looked at from an oblique direction. However, when pictures are viewed from an oblique direction, it is important to distinguish two different perceptual phenomena [31]: (a) the perception of the 3D layout of objects depicted in the pictures and (b) the direction a depicted object appears to point out of the picture, when that object is reaching out of the picture into the observer's space. It is well known that the perception of spatial layout depicted in a picture remains relatively constant despite the changes in the viewing angle of the observer. In the second phenomenon, the perceived orientation of objects pointing out from the picture plane more or less perpendicularly, undergo large changes when an observer changes the viewing angle. Objects depicted on pictures that point directly out of the picture appear to rotate so that they follow the moving observer and keep pointing at him as in Fig. 3. Objects on pictures that point to the side, however, still follow the observer, but not synchronously, so that they do not maintain a constant direction relative to a moving observer. This is a manifestation of the differential rotation effect (DRE) [32].

The differential rotation effect occurs not only on objects that extend in depth from the picture plane towards the observer but also on the perceived gaze direction. In classical Greece, it was already well known that "when a straight-on face is looking directly at an observer, its eyes will rotate to follow the observer so that they appear to be looking directly at the observer, no matter where he or she is relative to the picture" [32]. However, Goldstein [32] also demonstrated that, in accordance with the DRE, the more the perceived gaze deviates from the central direction to the left or to the right, the less it follows the observer of the picture when he moves around.

The Wollaston effect demonstrates that the perceived gaze direction of a portrait depends not only on the position of the irises but also on the orientation of the head [30]. For example, the gaze direction of a face image, with the face turned leftwards and the irises rightwards, is directed towards the observer. If one makes a mirror copy of that face image, turning the face consequently rightwards, and then pastes the original rightwards looking irises into the mirrored face, in place of the mirrored irises, the resulting gaze also shifts rightwards. The Wollaston effect thereby shows that the perceived gaze direction can be changed without any manipulation of the irises but only by turning the head of the looker.

Our mind is constantly interpreting and giving structure to the raw visual input from our eyes. We prefer an ordered world, familiar shapes and regular patterns. One of such features of human perception is that our brain tends to order visual features in a regular, orderly, symmetric and simple manner, as formulated by the Gestalt school in psychology [33]. Therefore where possible, we see stable rectangular forms although these forms appear most of the time distorted due to the perspective projection and are also constantly changing due to our movement. This principle is called shape constancy [34].

The exact mechanism that supports correct space perception from deformed retinal images is still disputed in the human perception research community. Cutting [35], for example, explains that the visual system corrects the distortions based on the assumption that objects are rigid. Sedgwick gives a theoretical analysis based on the concept of available visual information [36]. More recent research in human perception has shown that the adjustment to oblique viewing is achieved before the contents of the image are interpreted [37]. The adjustment to oblique viewing is based on the local slant of the picture which can be estimated from binocular disparity and the perspective of the picture frame. When viewing at the picture's surface from a very oblique position so that the image slant is larger than 60°, the estimation of the local slant becomes uncertain and the adjustment for oblique viewing diminishes [37]. This explains why the dissociation of the supporting surface and the image is so important for the anamorphic effect to materialize.

Anamorphosis is a human visual perceptual effect which was discovered in art in the late 15th Century [38,39]. When viewing an anamorphic image up-front, it is usually so distorted that it is unrecognizable. However, when we look at an anamorphic image at a radically oblique angle, we can observe the anamorphic image undistorted. Probably the most famous example of anamorphosis in art history is the 1533 painting *The Ambassadors*, by Hans Holbein, where a human scull at the bottom of the picture is anamorphically distorted [40]. Nowadays, anamorphic images are often produced as pavement or sidewalk art drawn in chalk for public amusement [41–43]. Recently, we have introduced the concept of dynamic anamorphosis which can dynamically adapt the shape of the anamorphic image to the changing position of the observer in such a way, that the observer always sees the anamorphic image in its un-deformed form [17]. If there is a frame around the anamorphic image it should also be deformed so that it supports the anamorphic effect.

Fig. 3. The outstretched hand of the man on the placard follows the observer when he changes the observation angle. Note, that the gaze direction of the man on the placard also follows the observer.

Straying away from the right viewpoint of an anamorphic image can quickly deteriorate the anamorphic effect and the image appears strangely distorted. Viewing *normal* pictures from an oblique angle does not result in a distorted picture since human perception can automatically compensate for the distortion using the principle of shape constancy [35,44]. A person viewing a *normal* picture from an arbitrary viewpoint must treat separately two perspective effects on the image that he perceives: the perspective deformation due to oblique viewing *and* the perspective deformation that is due to the content of the picture. Namely, if the pictorial perception would depend only on the geometry of the projected retinal image, the perception of the depicted space would be deformed in comparison to the actual depicted pictorial space [32].

3. Proposal for a better eye contact experience in videoconferencing

When researching anamorphosis [17] we observed that we could manipulate the perceived gaze of a person looking out of an image by applying different deformations to the image. We presented our initial hypothesis, that the eye contact could be improved by rotating the face image of the corresponding partner by rotating it along the *x*-axis in 2011 in a short conference article [16] and later presented it as an application of dynamic anamorphosis, a new concept, which we introduced in [17]. In this article we describe the proposed solution in the wider context of how the problem of missing eye contact in videoconferencing was approached in general, offer some new experiments and a more in-depth analysis and discussion of our proposal.

Trying to solve the problem of establishing eye contact in videoconferencing systems, Microsoft researchers also discussed the possibility that it might be possible to change the eye gaze by rotating the image if the face in the image is initially not looking at the viewer [45]. This report encouraged us to explore our initial, independently formed idea for improving eye contact, further. It should be noted that all methods that try to alleviate the problem of missing eye contact, including the proposed method, do not force videoconferencing partners into eye contact but just try to make it possible when and if the videoconferencing partners want to engage into eye contact.

Since the most common physical configuration of the computer screen and the camera used in videoconferencing applications is achieved by placing the camera centrally above the screen, we formed our proposal around this configuration and also performed all experiments using such a configuration (Fig. 1).

If we rotate the image of our videoconferencing partner in 3D for a moderate angle around *x*-axis, so that the top of the picture moves away from us, we still perceive the partner as before because our human perception estimates the amount of the rotation to correct for the perspective deformation (Fig. 4). We observed, however, that the eye gaze of our video partner which in the original, not rotated image, is directed somewhere below our face (as in Fig. 2(a)) also seems to rotate. When the amount of the rotation of the image plane is appropriate, observers report a better eye contact (Fig. 5). We decided to test this hypothesis with several experiments.

3.1. Experiment A

We decided to test out hypothesis, which is succinctly presented in Fig. 5, with the following experiment that we first described and discussed in [17].

Fig. 4. The image of a videoconference participant, shown in Fig. 2(a), is rotated in 3D space around *x*-axis for 15° with the top edge of the image moving away of the observer. Our hypothesis is that this enables better eye contact between videoconferencing participants.

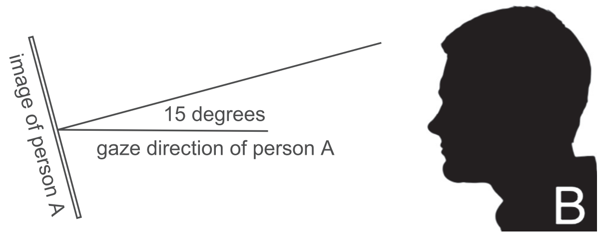


Fig. 5. By rotating the image plane depicting person A, A's perceived eye gaze direction also seems to rotate so that person B has now a subjective better experience of having eye contact with person A.

3.1.1. Method

We have asked four different people to sit in front of a 27-in. computer monitor and to look towards the center of the monitor. We took still pictures of them with a camera, built in the computer monitor centrally above the computer screen, as shown also in Fig. 2(a). These subjects were asked to assume a comfortable position, they were not very strictly guided on how to sit behind the monitor and how to look at it. We made a web application that presented the four pictures in random order as potential partners in a videoconferencing session. In the web application, the pictures could be rotated in 3D, around the *x*-axis in one degree increments using an interactive slider provided by the application below the presented face image.

3.1.2. Participants and the execution of the experiment

A group of 54 mostly undergraduate and graduate students was asked to visit the experiment's web page and to rotate the displayed face picture for the angle, at which they experienced the best eye contact with the person in the image. The time to select the orientation was not limited. The participants could be performing the experiment over the Internet from anywhere and we asked them only to visit the experiment's web page from a large table top computer monitor. We could not control, therefore, neither the size of the computer monitor that they were actually using, nor the position of their body relative to the monitor.

3.1.3. Results

Two students responded that no amount of rotation could improve the eye contact. The results of the other 52 subjects are summarized in Table 1 and Fig. 6.

The histogram of votes in Fig. 6 suggest that an increase of eye contact satisfaction can be achieved by a 3D rotation of the image, depicting the face of our videoconferencing partner, around *x*-axis,

Table 1
Descriptive statistics of experiment A, where the angle of rotation of the image around *x*-axis was modified according to the quality of eye contact experience.

	Votes	Mean [°]	Bias [°]	Std. error [°]	95% Confidence interval Lower [°]	Upper [°]
Face 1	52	13.27	−0.06	2.61	7.60	17.79
Face 2	52	12.40	−0.10	2.35	7.50	16.73
Face 3	52	15.29	−0.08	2.44	10.10	19.71
Face 4	52	12.40	−0.08	2.73	6.44	17.21

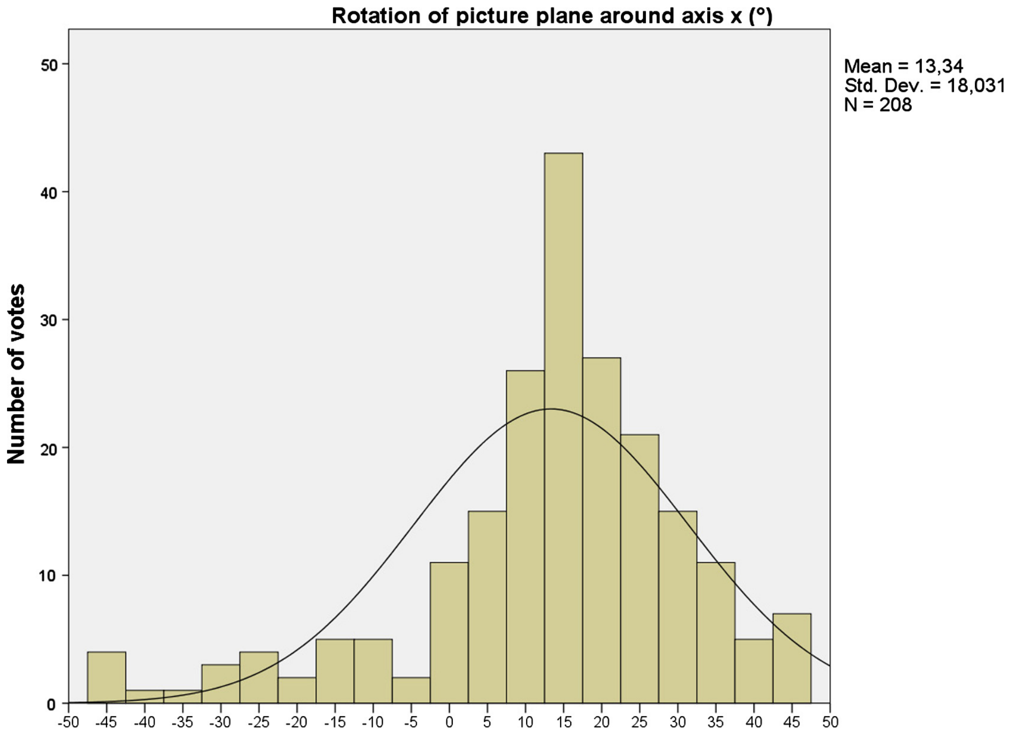


Fig. 6. Results of experiment A: 52 subjects, who tested four images each, are included in the histogram. The subjects had to determine interactively the amount of rotation of each face image, at which they experienced the best eye contact with the depicted person. The peak of the histogram of selected rotational angles for all four images and all 52 participants is at 15°, which corresponds approximately to the angle when someone sits in front of a 27-in. monitor and shifts his gaze from the middle of the monitor to the top of the monitor where the camera is normally located.

so that the top edge of the image moves away from us. Approximately 80% of the votes supported this approach for improving eye contact. The peak of the histogram shows that the most efficient angle of rotation around *x*-axis is at 15°, while 42% of votes selected angles between 10° and 20° and 60% of votes selected angles between 5° and 25° (Fig. 6).

If in the statistical evaluation of experiment A we select as the null hypothesis, H_0: angle of rotation around *x*-axis = 0, and H_1: angle of rotation around *x*-axis $\neq 0$, then the null hypothesis H_0 can be rejected. The 95% confidence interval for the angle of rotation for all four faces used in the experiment A is between 6.44° and 19.71° (Table 1).

Note that when we sit in front of a 27-in. monitor with our eyes about 60 cm away from the screen, our eye gaze direction, directed towards the center of the screen, also changes for about 15° when we raise the gaze towards the camera mounted above the display.

3.2. Experiment B

Since the first experiment A indicated, that the rotation of the image of a potential videoconferencing partner for 15 degrees around *x*-axis was the predominant result which supposedly improved the subjective experience of eye contact, we decided to compare next, unrotated images (Fig. 7(a)) directly with images, pre-rotated for 15 degrees (Fig. 7(c)) [17]. To broaden the experiment we compared in the experiment also images where the potential videoconferencing partner was looking straight into the camera above the computer screen (Fig. 7(b)), as the image which should give the best possible experience of eye contact, and pre-rotated images (Fig. 7(c)), where the perspectively deformed image frame of the rotated image was vertically cropped off (Fig. 7(d)).

3.2.1. Method

Using the same equipment as in the first experiment (27-in. monitor, built-in camera above the screen) we took still images of nine people (5 females and 4 males) who were looking into the center of the screen, where the face of their video conferencing partner would normally appear, as well as a second image, when they were looking directly into the camera above the screen—similar as in Figs. 2(a) and (b), to produce the following set of four images for each person, whose images were used in the second experiment (Fig. 7):

(a) image of a person looking into the center of the screen,
(b) image of a person looking into the camera above the screen,

(a) person looking into the center of the screen

(b) person looking into the camera above the screen

(c) image (a), pre-rotated for 15° around x-axis

(d) image (a), pre-rotated for 15° around x-axis, with vertically cropped off edges

Fig. 7. In the second experiment B the following four types of images of a potential videoconferencing partners were used [17]. Images of nine different potential videoconferencing partners (4 males and 5 females) were used in the experiment and 223 subjects participated in the experiment.

(c) image of a person looking into the center of the screen, pre-rotated for 15° around x-axis, and
(d) image of a person looking into the center of the screen, pre-rotated for 15° around x-axis and with the perspectively deformed left and right edges of the image frame vertically cropped off.

3.2.2. Participants and the execution of the experiment

The participants of this experiment did not overlap with the participants of the first experiment and a large majority of them was also not familiar with our hypothesis on how to improve eye contact. They could perform the experiment from anywhere over the Internet.

By means of a web application we were asking each participant in the test group, to compare pairwise all possible combinations of the four test images and select that image of the potential videoconferencing partner, where the experience of eye contact is the strongest.

Since we compared pairwise four types of images (Fig. 7(a)–(d)), there were six image combinations for each of the nine persons in the test group. Images of persons when they were looking straight into the camera (b) were included as a reference, since these images should obviously offer the best eye contact. Images of a person looking into the center of the screen rotated for 15° around x-axis (c) should, according to our hypothesis and the results of the first experiment, offer better eye contact than images of type (a). Finally, to investigate also the role of the perspectively deformed picture frame in the perception of the rotated pictures, we included in the comparison also images rotated for 15° around x-axis where the left and right perspectively deformed image edges were then vertically cropped off (Fig. 7(d)).

The experiment was conducted over the web in the following way. Each participant of the test was asked to open the test web application on a large table top computer monitor and to switch to a full screen mode during the test. Participants were then guided through 18 image pairs, each pair representing one of the six possible combinations of the four images of the same person (Fig. 7). For each image pair, the test participant was asked to select among the two displayed images the one, which gave him or her a better experience of eye contact. If one could not decide on any of the two images, because the difference in perceived eye contact was so small, one could select a third option—cannot decide. To remove any bias, the sequence of 18 image pairs was generated for each test participant individually in the following way: the sequence of persons appearing on test images was selected randomly and each among the 9 imaged persons appeared in an image pair exactly two times, among the 18 image pairs each type of image pair combination out of six possible combinations appeared exactly three times. The left/right ordering of each image pair was also generated randomly. Therefore, in none of the testing sequences for a single tested subject, the same image pair could appear twice. In the second experiment 229 subjects voted on the experiment's web page.

3.2.3. Results

The results of voting are summarized in Table 2. The first pairwise comparison in Table 2 includes images of people who were looking into the center of the screen, and images where people were looking directly into the camera above the screen. This comparison is actually a demonstration of the problem of missing eye contact on the most common equipment configuration. As expected, the large majority of subjects selected the latter images.

The second pairwise comparison confirms our basic hypothesis that rotating in 3D the image of a potential videoconferencing partner around x-axis improves the experienced eye contact. Although almost half of the subjects (45%) could not decide which image is

Table 2
In the second experiment B, four different image types of the same potential videoconferencing partner (Fig. 7(a)–(d)) were pairwise compared to decide, which image offered a subjectively better feeling of eye contact.

Which image in each image pair gives subjectively a better feeling of eye contact?					
1.		8% Fig. 7 (a)	3% undecided	89% Fig. 7 (b)	
2.		19% Fig. 7 (a)	45% undecided	36% Fig. 7 (c)	
3.		21% Fig. 7 (a)	54% undecided	25% Fig. 7 (d)	
4.		12% Fig. 7 (c)	2% undecided	86% Fig. 7 (b)	
5.		10% Fig. 7 (d)	3% undecided	87% Fig. 7 (b)	
6.		31% Fig. 7 (c)	48% undecided	21% Fig. 7 (d)	

better, in the other half, subjects preferred the rotated image almost twice as often (36%) as the original image (19%). If the perspectively deformed left and right edge of the rotated image is vertically cropped off (comparison 3 in Table 2), its preference drops off to 25% but is still slightly better than the original image (21%). This indicates that the perspectively deformed image frame plays an important part in the discussed effect since it makes the observation that the image plane is rotated in 3D around x-axis much easier.

The pairwise comparison No. 4 in Table 2 clearly shows that the best possible image—the image of a person looking straight into the camera—is definitely better for eye contact than the rotated image of a person looking into the center of the screen. Note, however, that rotating the image does raise slightly the percentage of participants who prefer the rotated image by 4%, in comparison to the non-rotated image (see the first pairwise comparison).

The results of pairwise comparison No. 5 are nearly the same as for comparison No. 1, since the rotated image, with vertically cropped off left and right edges, is difficult to distinguish from the original, non-rotated image, when we compare it to the image where the person is looking into the camera. Pairwise comparison No. 6 indicates that the observed effect is stronger, if the frame of the rotated image is also perspectively deformed. Note that the 10% difference in favor of images with a perspectively deformed frame in this last comparison is nearly the same, as when these two types of images are compared separately to the original images (see comparisons 2 and 3).

Although the results of the second experiment are in accordance with our hypothesis, the results do not manifest the effect of rotating the image on the appearance of eye contact as strongly as in the first experiment. We decided therefore to test our hypothesis not only with static images of faces but using live video connection. In the third experiment we focused only on our basic hypothesis, the comparison No. 2 in Table 2.

3.3. Experiment C

In our third experiment, we decided to test our hypothesis, that rotating the image depicting our videoconferencing partner can help in improving eye contact, with live video images in a mock videoconferencing session between two persons.

3.3.1. Method

We prepared two identical setups consisting of a large computer screen (22-in.) with a web camera on top of it. A pair of subjects was asked to discuss over this videoconferencing link a topic of their own choice for about two minutes. Each subject could individually and independently switch anytime among the standard image of his videoconferencing partner (image A in Fig. 8) and the pre-rotated version (15 degrees around x-axis) of the same image (image B in Fig. 8) to determine, which image among the two offers a subjectively better experience of having eye contact with the other subject (Fig. 8). At the end of the test session each subject had to fill out a paper form, selecting separately for each image A and B one of the values: *very bad*, *bad*, *good*, *very good*, to evaluate their experience of having eye contact during the test.

3.3.2. Participants

The subjects who took part in the experiment were not familiar with our hypothesis. A total of 78 subjects took part in this experiment.

3.3.3. Results

The votes recorded by subjects in experiment C are shown in Table 3. According to the Wilcoxon signed-rank test, the selection of images A and B by the test subjects does not differ significantly ($p = 0,248$). The third experiment was hence not as successful in supporting our hypothesis as even the second one.

Table 3
Results of Experiment C.

	Image A		Image B	
	Votes	Percentage	Votes	Percentage
very bad	2	2.6	1	1.3
bad	14	17.9	12	15.4
good	44	56.4	42	53.8
very good	18	23.1	23	29.5
Total	78	100.0	78	100.0

4. Discussion

Most studies of eye gaze perception experimented with left-right movements of the eyes and not with the up and down movements that we are addressing in this article. According to Sedgwick [36], moving up-down is from a geometrical point of view the same as moving left-right relative to the observed picture. However, due to the elongated shape of the eye, the possible up-down movement of the iris is much smaller than the left-right movement.

Our hypothesis is somewhat surprising in the light of published psychophysical studies on perception of gaze direction and of slanted pictures. According to Todorović [30], correcting the eye gaze just by rotating an image should not be possible: "if the looker's gaze misses me from one vantage point by a particular angle, specified by a particular combination of head and eyes cues, then it will miss me from most any vantage point by a similar angle."

The observer-related gaze direction is an additive combination of the observer-related head orientation and the looker-related gaze direction. A possible underlying mechanism for gaze determination is that the human visual system first extracts independently the iris eccentricity and head orientation information from the looker's head, and then combines these two measures to form the gaze direction judgment [30]. This can be explained by the fact that reaction times for judgements of gaze direction are shorter when the eyes and the head are turned in the same direction than when they are turned in different directions [5]. This means that gaze direction is determined independently and in parallel from the eyes and the head and then they are combined. These two information sources may be either congruent or incongruent, leading to corresponding acceleration or de-acceleration of gaze processing [30]. The Wollaston effect explains that it is possible to manipulate with the perceived gaze direction by changing separately, either the position of the iris, or the head of the looker, or both at the same time, but in different amounts so that the relation between the iris and the head are changed.

4.1. Our hypothesis

The hypothesis that we propose in this paper is as follows: *one can influence the perceived gaze direction of a person portrayed frontally on an image, by rotating the image around the horizontal image axis.*

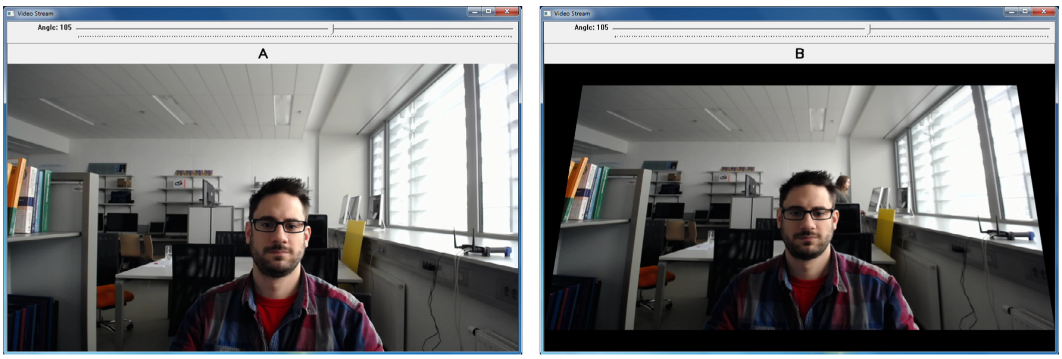


Fig. 8. Image A (left) and image B (right), which is a pre-rotated version of the same image, are frames from a live video stream used in the experiment C.

Based on the above deliberations about gaze perception and human visual perception in general, we put forward the following explanation of the mechanism behind our hypothesis: by rotating the picture, the differential rotation effect steps into action for the perceived gaze direction, which is estimated from the iris eccentricity, and the face/head of the looker. Since the perceived gaze direction "extends" out of the picture more than the head, according to DRE, the gaze seems to turn more than the face itself. This difference in perceived rotation introduces a perceived change in the head/eyes relation, causing in line with the Wollaston effect also a change in the perceived gaze direction of the looker. The perceived gaze direction of the lookers in Figs. 4, 7(c), and 8(b), which was originally determined only on the looker's eyes, therefore rotates according to DRE more than the face of the looker. The final perceived gaze direction is then assembled after the DRE effect takes place, out of the rotated eye based gaze direction and the head based gaze direction. Since in our case, the head of the looker is in frontal orientation, the head orientation does not have a large influence on determining the perceived gaze direction anyway.

Despite a thorough search, we could not find a similar hypothesis in literature.

4.2. Explanation of the hypothesis

An earlier psychophysical experiment can also be interpreted in support of our hypothesis. Goldstein performed an experiment (Experiment 3 in [32]) to study how the viewing angle affects the perceived gaze direction of portraits looking directly at the observer. In the experiment he used six frontally faced drawings of faces that appear due to iris eccentricity to be looking to the left or right of the observer. For the portrait looking directly at the observer, the results are exactly such as predicted by DRE, which is, that for all viewing angles the perceived gaze direction is the same as the viewing angle, meaning the portrait looks directly to the observer. The faces with other gaze directions, however, rotate less for the same change of viewing angle. For example, for the portrait that is looking 15° to the left and is observed frontally (from 0°), the perceived gaze direction is 30° to the left, if the same portrait is observed from 20° to the left, the perceived gaze direction is 45° to the left, if the portrait is observed from 40° to the left, the perceived gaze direction is 60° and if the portrait is observed from 70° to the left, the perceived gaze direction is 75°. Since the perceived gaze direction of the looker rotates slower than the direction of observation, the original offset between the looking direction and the perceived gaze direction at the frontal position of the observer is decreasing when the observer moves in the direction of the perceived gaze. When the direction of observation becomes close enough with the perceived gaze direction, eye contact could be established. The experiment demonstrates that if the observer tries to reach the perceived gaze direction of the looker on a picture by moving or rotating his gaze into the direction of the looker's gaze, the difference between the two gaze directions decreases. This mechanism can explain our hypothesis. Note again that Goldstein performed his experiment for left-right movements of the head and the eyes, while the hypothesis described in this article employs up and down movements. Goldstein's experiment [32] described above may hint that, in this way, the difference in view directions can only be decreased but never completely eliminated.

In all our experiments the camera was placed centrally above the computer screen. The proposed solution should work also with other possible configurations, such as camera at the bottom or on the side of the computer screen which can be clearly deduced from the experiments performed by Goldstein [32] which are described above. However, since these alternative configurations are very uncommon in videoconferencing systems we did not perform experiments using such configurations. Note, however, that Chen [46] observed higher sensitivity of observing a gaze change in all directions except when the looker was looking down. When the looker was looking down, observers noted a loss of eye contact only when the parallax angle reached 10°. This might be an additional explanation why our proposal works.

4.3. Other considerations

There are still many open questions whether this strategy of following the perceived gaze of the videoconferencing partner by rotating his image can be really successful. The first experiment that we conducted was very supportive of our hypothesis, the second one was not as explicitly supportive, while the third one even less so.

Since the first two experiments described in this article were conducted over the Internet, we could not control the size of the computer screen that subjects participating in the experiments were using, nor the position of the subjects relative to the monitor. To make the experimental conditions more uniform for all subjects and to reduce the number of variables in the experiments we decided to use still images in the first two experiments. People whose still images were used in the experiments were also not very strictly guided on how to sit behind the monitor and how to look at it. These inter-personal differences also contributed to the variability of results. In the third experiment C, a pair ob subjects was used in a mock videoconferencing session, using thus live video imagery.

The initial first experiment allowed that the subjects rotated the image plane themselves into a position that they preferred. In the second and third experiment, the subjects were given besides the "standard" videoconferencing image, also a pre-rotated version of the same image for comparison. The observed effect on the subjective feeling of eye contact was in the following two experiments not as strong as in the first experiment. It seems, therefore, that the ability of the subjects to rotate the images themselves, heightens their perception that the image plane is actually rotated in 3D and, consequently, intensifies the effect of the rotation and therefore increases also the subjective feeling of eye contact as a consequence.

The simultaneous perspective deformation of the image frame is therefore also important to perceive this as a rotation of the image plane in 3D. The importance of the perspective deformation of the image frame is also clearly demonstrated in the second experiment B. The pairwise comparison of images of a person looking into the center of the screen with the same image, pre-rotated for 15° around *x*-axis and with vertically cropped off edges (image pair No. 3 in experiment B – Table 2) shows that subjects had difficulty in distinguishing this two images since the number of undecided votes was very high (54%), while the preference votes for both images was practically the same (21% vs. 25%). The pairwise comparison of the same image, with the perspectively deformed image frame and with the vertically cropped off edges (image pair No. 6 in experiment B – Table 2), also indicates that the deformed image frame plays a role in the described effect. Therefore, one can conclude, that any user interface that would employ this effect should preferably offer the possibility that the user himself adjusts/rotates the image of his video-conferencing partner to raise his awareness that he observes an image with its picture plane rotated in 3D.

The perceptual mechanisms that underlay the described observations should be explored more thoroughly by more carefully designed psychophysical experiments in a more closely controlled environment.

5. Conclusions

The problem of missing eye contact in videoconferencing systems is due to the fact that when people want to engage in eye contact with the speaking partner they cannot achieve it due to technical limitations of the videoconferencing systems. The normal way of establishing eye contact is to mutually look into each other's eyes. Since the optical axes of the cameras and viewing directions towards the displayed face images on videoconferencing systems are not aligned, this simple strategy does not work and at least subconsciously causes frustration which hinders a prolonged or often use of videoconferencing. This may be the cause why videoconferencing systems are not as widely used as it is now technically possible. To solve the problem of gaze awareness we formulated a hypothesis on how, by rotating in 3D the image plane of the videoconferencing partner around the x-axis, one can influence the perceived view direction of the depicted person. We performed three different experiments to explore and illuminate our hypothesis. Although the results of all experiments are not conclusive, we believe that we have demonstrated that by rotating in 3D the image plane of a videoconferencing partner around x-axis of the image so that the top edge of the image moves away from us, the subjective experience of eye contact between two videoconferencing partners could be improved. By this simple means one can somehow trick the human visual perceptional system to perceive a change in gaze direction of our videoconferencing partner.

We offer a psychophysical explanation for this hypothesis. A more conclusive demonstration of the practical value of the proposed solution could be achieved by developing a dedicated open-source videoconferencing application, implementing the functionality of rotating the image plane by the user around x-axis, and asking a large and broad user pool to test and use it. The software would automatically record how much and at what angles was the functionality of image plane rotation actually used by a large number of user in a longer time frame. The analysis of data collected in this manner could help to answer not only the question, how this approach to solving the missing eye contact problem in videoconferencing is grounded in psychophysics, but also whether it has actually any practical value. Such a study is unfortunately out of the scope of this article. However, in the future, the problem of missing eye contact in videoconferencing can be most efficiently solved by a new generation of computer displays where the imaging sensors will be integrated into the computer screen, between the display elements of the screen, aligning in this way completely the optical axis of the camera and the viewing direction.

Acknowledgements

We thank Robert Ravnik and Jaka Konda for help with conducting the three experiments and Jože Gosenca for help in the statistical evaluation of the experiments.

References

[1] Sheldon Hochheiser, Before Facetime or Skype, there was the Picturephone, IEEE Spectrum 56 (2015).
[2] Anne Suwita, Martin Böcker, Lothar Mühlbach, Detlef Runde, Overcoming human factors deficiencies of videocommunications systems by means of advanced image technologies, Displays 17 (2) (1997) 75–88.
[3] Alex Colburn, Michael F. Cohen, Steven Drucker, The Role of Eye Gaze in Avatar Mediated Conversational Interfaces, Technical Report MSR-TR-2000-81, Microsoft Research, Microsoft Corporation, 2000.
[4] Leanne S. Bohannon, Andrew M. Herbert, Jeff B. Pelz, Esa M. Rantanen, Eye contact and video-mediated communication: a review, Displays 34 (2) (2013) 177–185.
[5] Stephen R.H. Langton, The mutual influence of gaze and head orientation in the analysis of social attention direction, Quart. J. Exp. Psychol. 53 (3) (2000) 825–845.
[6] Judy C. Nixon, Judy F. West, Intercultural preparation for managers going to Japan, Am. J. Bus. 10 (2) (1995) 57–64.
[7] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, Jack M. Loomis, Equilibrium theory revisited: mutual gaze and personal space in virtual environments, Presence 10 (6) (2001) 583–598.
[8] Ernst Bekkering, J.P. Shim, i2i trust in videoconferencing, Commun. ACM 49 (7) (2006) 103–107.
[9] Gwyneth Doherty-Sneddon, Anne Anderson, Claire O'Malley, Steve Langton, Simon Garrod, Vicki Bruce, Face-to-face and video-mediated communication: a comparison of dialogue structure and task performance, J. Exp. Psychol.: Appl. 3 (2) (1997) 105.
[10] Vicki Bruce, The role of the face in communication: implications for videophone design, Interact. Comput. 8 (2) (1996) 166–176.
[11] Ben Yip, Jesse S. Jin, Pose determination and viewpoint determination of human head in video conferencing based on head movement, in: Yi-Ping Phoebe Chen (Ed.), 10th International Multimedia Modeling Conference, IEEE Computer Society, 2004, pp. 130–135.
[12] Rembert R. Stokes, Human factors and appearance design considerations of the Mod II Picturephone station set, IEEE Trans. Commun. Technol. 17 (2) (1969) 318–323.
[13] Peter J. Werkhoven, Jan Maarten Schraagen, Patrick A.J. Punte, Seeing is believing: communication performance under isotropic teleconferencing conditions, Displays 22 (4) (2001) 137–149.
[14] Jaron Lanier, Virtually there, Sci. Am. 284 (4) (2001) 66–75.
[15] Gregorij Kurillo, Ruzena Bajcsy, 3D teleimmersion for collaboration and interaction of geographically distributed users, Virt. Real. 17 (1) (2013) 29–43.
[16] Franc Solina, Robert Ravnik, Fixing missing eye-contact in video conferencing systems, in: Vesna Luzar-Stiffler, Iva Jarec, Zoran Bekić (Eds.), Information Technology Interfaces (ITI), Proceedings of the ITI 2011 33rd International Conference on, 2011, pp. 233–236.
[17] Robert Ravnik, Borut Batagelj, Bojan Kverh, Franc Solina, Dynamic anamorphosis as a special, computer-generated user interface, Interact. Comput. 26 (1) (2014) 46–62, http://dx.doi.org/10.1093/iwc/iwt027.
[18] Terence J. Nelson, Lanny S. Smoot, Eye Contact Apparatus for Video Conferencing, US Patent 5,117,285, 1992. <https://www.google.com/patents/US5117285>.
[19] Zach Spear, Apple Files May-2016 for Camera Hidden Behind Display, 2009. <http://appleinsider.com/articles/09/01/08/apple_files_patent_for_camera_hidden_behind_display>. (accessed 10-May-2016).
[20] Apple Invents a Wild New Display that Could Conceal a Camera, Strobe Flash and/or Fingerprint Scanner Until Needed, 2013. <http://www.patentlyapple.com/patently-apple/2013/05/apple-invents-a-wild-new-display-that-could-conceal-a-camera-strobe-flash-andor-fingerprint-scanner-until-needed.html> (accessed 10-May-2016).
[21] Jedidiah Becker, Sony Patents Technology to Put Camera and Sensors Behind Smartphone Display, 2012. <http://www.redorbit.com/news/technology/1112501121/sony-patents-technology-to-put-camera-and-sensors-behind-smartphone-display/> (accessed 10-May-2016).
[22] Jim Gemmell, Kentaro Toyama, C. Lawrence Zitnick, Thomas Kang, Steven M. Seitz, Gaze awareness for video-conferencing: a software approach, IEEE MultiMedia 7 (4) (2000) 26–35.
[23] Maximilian Ott, John P. Lewis, Ingemar Cox, Teleconferencing eye contract using a virtual camera, in: INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems, CHI '93, New York, NY, USA, ACM, 1993, pp. 109–110.
[24] Maarten Dumont, Steven Maesen, Sammy Rogmans, Philippe Bekaert, A prototype for practical eye-gaze corrected video chat on graphics hardware, in: Pedro A. Amado Assunção, Sérgio M.M. de Faria (Eds.), SIGMAP 2008, INSTICC Press, 2008, pp. 236–243.
[25] Ben Yip, Face and eye rectification in video conference using affine transform, in: ICIP 2005, 2005, pp. 513–516.
[26] Lior Wolf, Ziv Freund, Shai Avidan, An eye for an eye: a single camera gaze-replacement method, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 817–824.
[27] Yalun Qin, Kuo-Chin Lien, Matthew Turk, Tobias Höllerer, Eye gaze correction with a single webcam based on eye-replacement, in: Advances in Visual Computing, Springer, 2015, pp. 599–609.
[28] Claudia Kuster, Tiberiu Popa, Jean-Charles Bazin, Craig Gotsman, Markus Gross, Gaze correction for home video conferencing, ACM Trans. Graph. (TOG) 31 (6) (2012) 174.
[29] Dominik Giger, Jean-Charles Bazin, Claudia Kuster, Tiberiu Popa, Markus Gross, Gaze correction with a single webcam, in: Proceedings IEEE International Conference on Multimedia & Expo, July 2014, pp. 1–6.
[30] Dejan Todorović, Geometrical basis of perception of gaze direction, Vision. Res. 46 (21) (2006) 3549–3562.
[31] E. Bruce Goldstein, Perceived orientation, spatial layout and the geometry of pictures, in: Stephen R. Ellis, Mary K. Kaiser, Arthur J. Grunwald (Eds.), Pictorial communication in virtual and real environments, second ed., Taylor & Francis, Inc., Bristol, PA, USA, 1993, pp. 480–485.
[32] E. Bruce Goldstein, Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle, J. Exp. Psychol. Hum. Percept. Perform. 13 (2) (1987) 256–266.
[33] Kurt Koffka, Principles of Gestalt Psychology, Harcourt, Brace and World, New York, 1935.
[34] Zygmunt Pizlo, A theory of shape constancy based on perspective invariants, Vision. Res. 34 (12) (1994) 1637–1658.
[35] James E Cutting, Rigidity in cinema seen from the front row, side aisle, J. Exp. Psychol. Hum. Percept. Perform. 13 (3) (1987) 323–334.

[36] Harold A. Sedgwick, The effects of viewpoint on the virtual space of pictures, in: Stephen R. Ellis, Mary K. Kaiser, Arthur J. Grunwald (Eds.), Pictorial communication in virtual and real environments, second ed., Taylor & Francis, Inc.,, Bristol, PA, USA, 1993, pp. 460–479.
[37] Dhanraj Vishwanath, Ahna R. Girshick, Martin S. Banks, Why pictures look right when viewed from the wrong place, Nat. Neurosci. 8 (10) (2005) 1401–1410.
[38] Daniel L. Collins, Anamorphosis and the eccentric observer: inverted perspective and construction of the gaze, Leonardo 25 (1) (1992) 72–82.
[39] Daniel L. Collins, Anamorphosis and the eccentric observer: history, technique, and current practice, Leonardo 25 (2) (1992) 179–187.
[40] David Topper, On anamorphosis: setting some things straight, Leonardo 33 (2) (2000) 115–124.
[41] Julian Beever, Pavement Drawings, 2010. <http://www.julianbeever.net> (accessed 11-May-2016).
[42] Manfred Stader, Master of 3D-street Painting, 2010. <http://www.3d-street-art.com> (accessed 23-July-2010).
[43] Kurt Wenner, Artist and Architect, 2010. <http://www.kurtwenner.com>. (accessed 23-July-2010).
[44] Dejan Todorović, Is pictorial perception robust? The effect of the observer vantage point on the perceived depth structure of linear-perspective images, Perception 37 (1) (2008) 106–125.
[45] C. Lawrence Zitnick, Jim Gemmell, Kentaro Toyama, Manipulation of Video Eye Gaze and Head Orientation for Video Teleconferencing, Technical Report MSR-TR-99-46, Microsoft Research, Microsoft Corporation, 1999.
[46] Milton Chen, Leveraging the asymmetric sensitivity of eye contact for videoconference, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2002, pp. 49–56.