Three-dimensional information in face recognition: An eye-tracking study

Olga Chelnokova  
Department of Psychology, University of Oslo, Norway

Bruno Laeng  
Department of Psychology, University of Oslo, Norway

One unresolved question about face perception is: what is the role of three-dimensional information in face recognition? In this study, recognition performance was compared across changes in viewpoint in different depth conditions: a 2D condition without stereo information and a 3D condition where stereo information was present (by viewing the same face images as anaglyphs through 3D glasses). Subjects’ eye movements were recorded during both 3D and 2D sessions. The findings revealed that participants were more accurate in the 3D condition. Moreover, individual differences in interpupillary distance predicted recognition performance in the 3D but not in the 2D condition. A “region of interest” analysis of gaze data showed that rich volumetric properties provided by certain facial features (e.g., the nose and the cheeks) were attended more in the 3D condition compared to the 2D condition. Taken together, these findings support the conclusion that face recognition across viewpoint transformation is facilitated by the addition of stereoscopic depth cues.

Keywords: face recognition, eye movements, depth perception, stereo vision, anaglyphs, 3D, interpupillary distance


Introduction

Humans show a high level of expertise at recognizing faces of conspecifics, even though faces of different people show relatively little variation in their underlying structure, compared to many other objects of the external world, and their appearances can vary dramatically by changes in pose, expression, and lighting. There is a long-standing debate about the nature of face and object representations in psychological literature. One of the early competing views (Biederman & Gerhardstein, 1993; but see Biederman & Kalocsai, 1997) suggests that the human brain creates a three-dimensional viewpoint-invariant representation of each particular face from the two-dimensional images that it receives from the retina. According to the opposing theory, two-dimensional representations of multiple views of faces are used, and recognition occurs by interpolation to the closest of previously seen views (Poggio & Edelman, 1990).

One of the components of this puzzle—view dependency—has been tackled both by psychophysical and neurophysiological studies. It has been shown that recognition of unfamiliar faces is view-dependent, but its view dependency decreases with the increase of face familiarity (Hancock, Bruce, & Burton, 2000). However, the fact that recognition of relatively unfamiliar faces appears to be view-dependent does not in itself indicate that view-dependent representations are necessarily two-dimensional and the fact that view dependency decreases with the increase of face familiarity also does not prove that the brain constructs three-dimensional representations after exposure to multiple views. Therefore, the question of whether three-dimensional information is relevant for face recognition remains an open and under-investigated question.

In the course of childhood development, as well as in normal communication situations, face learning occurs in a situation of contact over short distances when 3D cues are available for the observer. This creates the possibility that 3D information is indeed used and plays a role in forming internal face representations. Current data from haptic face recognition studies (e.g., Kilgour & Lederman, 2002) and tasks involving mental rotation of faces (Schwaninger & Yang, 2011) provide evidence that human face recognition may employ 3D representation mechanisms.

A recent study by Lee and Saunders (2011) demonstrated that the addition of stereo depth information improves recognition of random 3D objects across the change in viewpoint, even when rich monocular 3D cues are present. A number of studies that used facial stimuli have been able to show that the presence of volumetric information does help to construct internal face representations that can be invariant to changes such as viewpoint or illumination (Burke, Taubert, & Higman, 2007; Jiang, Blanz, & O’Toole, 2009); volumetric information has also been shown, in a computational study, to support better sex classification than the gray-level information (O’Toole, Vetter, Troje, & Bülthoff, 1997). Nevertheless, a few studies suggest the opposite conclusion that building internal face representations does not involve a reconstruction of the 3D structure information (Liu, Collin, & Chaudhuri, 2000; Liu, Ward, & Young, 2006).
The aim of the current study is to reassess the question of whether the addition of three-dimensional information facilitates face recognition from different viewpoints. In addition to behavioral measurements, we employ the eye movement recording technique, since this method allows us to observe whether 2D and 3D images of faces are treated differently in terms of overt attention. Specifically for this study, we expect eye fixations to reveal whether the presence or absence of real three-dimensional depth cues affects the control of gaze and whether viewpoint modulates such an effect.

Findings from previous eye-tracking studies indicate that subjects’ eye movements can be sensitive to 3D visual cues. Vishwanath and Kowler (2004) recorded eye movements of participants asked to look at novel ellipsoid geometric shapes. Although no stereo images were actually used in their experiment, the target shapes could be made to appear as either two-dimensional or three-dimensional by removing or adding shading information to the images. They found that for the 2D shapes, saccades landed near the 2D center of gravity of the shape, whereas for the 3D shapes, saccadic landing positions fell at either the 2D or 3D center of gravity. Thus, for shapes perceived to have depth properties, saccades were shifted toward the parts of the shape projected to be located at a greater distance from the viewer.

Previous research on eye fixations for 2D faces indicates that most of the normal viewers’ fixation time is received by the eyes and the nose (Bindemann, Scheepers, & Burton, 2009; Henderson, Williams, & Falk, 2005; Walker-Smith, Gale, & Findlay, 1977), followed by the mouth (particularly when emotional faces are presented, e.g., Dalton et al., 2005; Pelphrey et al., 2002) or the cheeks (Sæther, Van Belle, Laeng, Brennen, & Øvervoll, 2009). Hsiao and Cottrell (2008) showed that during recognition of full-frontal face views, the first two eye fixations land around the center of the nose, with fixations on the eyes occurring at a later stage. Some face regions, such as the nose and the cheeks, seem to provide more information about the volumetric structure of the face than other regions, such as the eyes. Considering the findings of Vishwanath and Kowler (2004), longer fixation times can be expected for the volume-rich face parts when viewing 3D face images.

**Present study**

In the current study, we used a face recognition task where each face was seen in several facial views. The task was designed after that of the Cambridge Face Memory Test (Duchaine & Nakayama, 2006). In such a task, participants see a target face and then are tested with forced choice items that consist of several faces, four in our specific case, only one of which is a target face. Importantly, in each trial, the test items are shown in views that are always different from the view seen for the initial, sample face. Thus, the task requires a transfer or generalization from the remembered facial information to the novel view. The relevant addition to the paradigm was to introduce a change of dimensionality in our face recognition task. In some blocks of the task, participants view both target and test faces in 2D, while in the rest of the blocks both target and test items are presented in 3D. In order to obtain a realistic 3D effect, we created facial “anaglyph” images by superimposing two equal 2D gray-level images of chromatically opposite red and cyan colors, which were then viewed through red–cyan stereo glasses. Anaglyph images in red–cyan are currently the most common 3D graphics in use for entertainment (e.g., comic books and photos posted on the Internet) or for the display of scientific data sets (e.g., to illustrate mathematical functions). Viewing anaglyphs through appropriately colored glasses causes each eye to see a slightly different picture. In a red–cyan anaglyph, the eye viewing through the red filter sees the red parts of the image as “white” and the cyan parts as “black,” while the eye viewing through the cyan filter perceives the opposite effect. True white or true black areas are perceived the same by each eye; hence, there is no change in tone quality when viewing gray-level images. The brain fuses together the image it receives from each eye and interprets the differences as being the result of a difference in distances, which, in turn, yields a normal stereograph image.

If three-dimensional information facilitates face recognition and helps create a transformation-invariant representation of a face, then performing a task that involves a transformation across viewpoints should be easier in the 3D condition. If, on the other hand, the presence of 3D depth cues has little influence on face recognition, then the participants’ performance for the 3D trials should not be different from or could even be worse than that of the 2D condition, given the unfamiliar situation of viewing 3D images through red–cyan stereo glasses on a computer screen.

We recorded eye-tracking data from 2D and 3D trials in order to find out whether there were any differences in visual sampling of the images for both of the conditions. Region of interest analyses were employed for the eye movement data, as we expected different facial parts, based on their volumetric features, to receive most attention in the 3D than in the 2D condition. Specifically, we predicted that volume-rich facial parts (such as the nose and the cheeks) would be inspected longer in the 3D condition compared to the 2D condition.

Finally, given that the distance between pupils can influence the quality of the perceived depth or stereo effect, we measured each participant’s interpupillary distance (IPD) and used the difference between each individual’s IPD and the one used for creating 3D images as a variable to predict the face recognition performance. Specifically, we expected that (a) IPD would play a significant role only for the 3D condition and (b) the highest performance...
would be achieved by those individuals whose IPDs were closest to the distance used to generate the anaglyphs, as this should yield an optimal stereo effect.

**Experiment 1**

**Methods**

**Subjects**

Eighteen female and twelve male students (mean age = 25.00; SD = 4.62) from the University Oslo, all with normal or corrected-to-normal vision, volunteered to participate in a study of face perception.

**Stimuli and apparatus**

Grayscale photographs of four young Caucasian women and four young Caucasian men were used as stimuli. All models volunteered to have their faces photographed for a psychological experiment. Four photographs were taken for each model from a distance of 1 m: Two pictures of a full-frontal view and two pictures of a left-sided intermediate 22.5° view. The 22.5° angle was selected on the basis of the results of previous studies (Blanz, Tarr, & Bulthoff, 1999; Laeng & Rouw, 2001). For each photograph, the models were asked to assume a neutral, emotionless, facial expression. Photographs were taken with a Canon digital SLR camera in a room illuminated with artificial fluorescent ceiling light. The camera’s built-in frontal flash was used to give sufficient lighting to the faces of the models. Subsequently, each image was edited with Adobe Photoshop software so that models’ head size was normalized, facial blemishes were touched over, and hair and the background were attenuated. Three-dimensional anaglyph images were then created with the use of RedGreen V3.0.1 software by merging two equal two-dimensional pictures taken at a 6.5-cm distance. A 6.5-cm distance was used, as it is a reasonable estimate of the mean interpupillary distance of the adult human population (e.g., Dodgson, 2004). To create a monochrome anaglyph image, two images of chromatically opposite colors (red and cyan in this case) are superimposed but offset with respect to each other. Viewing these “anaglyphs” through glasses of corresponding colors (in this case, cyan for the left lens and red for the right lens) results in each eye seeing a slightly different picture (see Dubois, 2001; Symanzik, 1993).

Each final sample image was 13 × 15.5 cm in size. Match image items were created by putting together four smaller (8.5 × 9.5 cm) images of the models of the same sex, all in one of the two views. (Figure 1 shows an example of a female face in two two-dimensional views used as sample images and a match image item with three-dimensional full-frontal images of four female models). The order of the images was randomized. Under each image, the number 1, 2, 3, or 4 was printed with 36-point Arial font.

The sample images were presented centered on a computer screen with a resolution of 1280 × 960 pixels, subtending 8.26 × 9.84 degrees of visual angle, equivalent to the size of a real face at a viewing distance of 1 m (i.e., roughly corresponding to the distance between two persons during a normal conversation; Henderson et al., 2005; Hsiao & Cottrell, 2008). Four keys on the response box were marked as 1, 2, 3, and 4. E-Prime 2.0 software

![Figure 1](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932792/)
(Psychology Software Tools, Pittsburgh, PA, USA) was used to present the stimuli and collect response times and accuracy rates. For viewing three-dimensional images, subjects were asked to wear red–cyan anaglyph paper glasses. Subjects’ left eye saw through the cyan filter, while the right eye saw through the red filter. Subjects’ eye movements during face recognition task were recorded for the left eye with the use of the Remote Eye Tracking Device (R.E.D.; SensoMotoric Instruments, Teltow, Germany) in the “eye lab” at the Department of Psychology, University of Oslo. This equipment allows eye tracking also when the participant wears contact lenses and prescription glasses; hence, it was not difficult for the apparatus to also track the eye through the color lenses of the stereo glasses. A built-in chin rest or “column” was used for supporting the head of a participant during the eye movement recording phase, as well as for keeping the distance from the computer screen the same for all subjects (90 cm) so as to ensure a stable 3D effect.

**Procedure**

The experiment started with a set of practice trials that familiarized the participants with the procedures by presenting cartoon faces in the same fashion as the actual test images. Figure 2 illustrates the sequence of events in a single trial. After a presentation of a fixation cross that appeared randomly in one of the four corners of the screen for 1500 ms (this ensured that the central fixation dot at the beginning of each trial did not coincide with any of the critical face regions; i.e., the first fixation on the face would always follow a saccade from the central fixation point), a sample image of a face was presented on the computer screen for 2000 ms. Then, after a 50-ms presentation of a blank screen, a match image item consisting of an image of the same model alongside with three other models’ faces was presented until participants responded by pressing the key on the response box. Observers were requested to decide which face, among the individuals shown in the match image item, was the same one they had seen earlier as a sample image. They were also asked to make their decisions as quickly and as accurately as possible. The angle of view of a face in a sample image was always different from the view of faces in the following match image item, so that if the sample face was presented first in the full-frontal view, then the faces of the models in the following match image item appeared in the 22.5° view and vice versa.

The total number of trials was 64. They were organized in 4 blocks according to an ABBA order: two-dimensional trials, three-dimensional trials, three-dimensional trials, and two-dimensional trials. Sixteen trials in each block were ordered so that the combinations of all conditions (model, view, correct response) were pseudorandomized.
Results

Behavioral data

All behavioral data analyses were carried out using the StatView statistical package. A preliminary analysis had shown neither significant main effects nor interactive effects with any of the other factors for both “Sex of the Model” and “Sex of the Participant.” Therefore, these two factors were excluded from the following analyses.

A repeated measures ANOVA of the percentage accuracy data was performed with depth (2D, 3D) and sample’s view (full frontal, intermediate) as within-subject factors. The analysis revealed a highly significant effect of depth ($F(1, 29) = 17.08, p < 0.001$). An effect of sample’s view ($F(1, 29) = 3.74, p = 0.06$) approached the 0.05 cutoff of statistical significance (Figure 3). The interaction effect of depth and sample’s view factors was not significant.

The same repeated measures ANOVA was performed for the response time (RT) data with depth (2D, 3D) and sample’s view (full frontal, intermediate) as within-subject factors. Prior to the analysis, error trials were excluded, and outliers defined as 2.5 SD from subject’s mean RT were trimmed. The only statistically significant effect found was the main effect of depth ($F(1, 29) = 4.54, p < 0.05$), with longer times spent processing the 3D images compared to the 2D ones.

To check for possible speed–accuracy trade-off effects, two simple linear regression analyses were performed on accuracy and RT data of all participants of the 2D and 3D conditions. For both conditions, the regression analyses revealed significant negative correlations (Table 1).

As it can be appreciated from the slope coefficients in Table 1, higher accuracy rates predicted shorter response times in both cases. Thus, in none of the conditions were participants trading speed for accuracy.

Finally, we assessed whether the perceived 3D effect depended on how different each viewer’s IPD was from the standard distance used to generate 3D images. The subjects’ IPDs were measured at the beginning and end of each experiment. By subtracting the camera distance from the average value of these two measurements, we computed the new variable that we called “Deviation from Optimality.” This variable was used as the regressor with either the percentage accuracy or RT data of the 2D and 3D conditions as the dependent variables (Table 2). As expected, no significant results were found for the 2D condition either for accuracy or RT data. However, in the 3D condition, a nearly significant trend was observed. As illustrated by Figure 4B, IPDs larger than the camera distance of 6.5 cm corresponded to lower percentage accuracy rates, with the highest accuracy rates corresponding to the IPDs that were slightly smaller than the camera distance. No significant results were observed for the RT data.

Eye-tracking data

Eye-tracking data from 11 of the 18 female participants and 9 of the 12 male participants (mean age = 25.75,
were analyzed using BeGaze 2.3 software (SensoMotoric Instruments, Teltow, Germany). Data from 10 participants had to be excluded due to recording artifacts. Again, the StatView statistical package was used for performing ANOVAs.

For analyzing eye-tracking data, the following regions of interest (ROIs) were selected for each of the faces: eyes (left and right eyes), nose, cheeks (left cheek and right cheek), mouth, forehead, and chin (see Figure 5 for an illustration of the ROI selected for one of the faces).

For each image and region of interest, the percentage of total fixation times was computed. Percentage values were selected as they provide a measure that is independent of the absolute length of recorded data samples. This makes it possible to minimize the effect of data loss that occurred in the 3D condition as a result of using the colored glasses. A repeated measures ANOVA was performed on fixation time data with depth (2D, 3D), sample’s view (full frontal, intermediate), and facial part (eyes, nose, mouth, cheeks, chin, forehead) as within-subject factors. Sex of the Model and Sex of the Participant were not considered because a preliminary ANOVA showed no effects or interactions with either of these two factors. Results showed significantly longer fixation times for the 2D condition compared to the 3D condition ($F(1, 19) = 8.37, p < 0.001$). An interaction effect of depth and view factors ($F(1, 19) = 5.45, p = 0.03$) revealed an increase of fixation time in the 2D condition compared to the 3D condition, which was larger for the intermediate than the frontal view.

A highly significant main effect of facial part ($F(5, 95) = 15.94, p < 0.0001$) showed the following order of facial parts that received the longest fixations: eyes, nose, mouth, cheeks, forehead, and chin. As indicated by a highly significant interaction effect of depth and facial part factors ($F(5, 95) = 5.62, p = 0.0001$), this order was, however, significantly different in the 3D condition, where participants’ fixation times for the eyes were lower than those for the nose and the fixation times for the cheeks surpassed the time spent fixating on the mouth (Figure 6).

No significant results were obtained for the view and facial part interaction, as well as for the interaction of the depth, view, and facial part factors.

To further illustrate the differences in overall gaze data for facial parts in the two-dimensional condition compared to the three-dimensional condition, “attention maps” were generated by use of the BeGaze software (also developed by SMI). In these images, cumulative patterns of fixations are visualized by altering the brightness of the stimulus display. In other words, the luminance of a specific area in these images reflects the amount of overt attention that was received by the area. Four attention maps were created using gaze data from all experimental trials of all participants in each of the following four conditions: 2D full-frontal sample’s view trials, 2D intermediate sample’s view trials, 3D full-frontal sample’s view trials, and 3D intermediate sample’s view trials.

Figure 5. ROIs selected for one of the stimulus faces.
3D intermediate sample’s view trials. Two images were used for both the 2D and 3D attention maps: a morphed image of all male faces in frontal view and a morphed image of all female faces in the intermediate view (Figure 7).

These attention maps provide a clear illustration of the fact that in the 2D condition gaze was more evenly distributed among eyes, nose, and mouth, forming a T-shaped pattern. This happened both for the full-frontal and intermediate views. In contrast, in the 3D condition, participants tended to look less at the eyes of the models, and overall, gaze appears shifted toward the nose and cheeks.

Discussion

We found that, after viewing a face in the intermediate view, participants were more accurate in recognizing the same face in full-frontal view than they were in a reverse condition. This finding indicates a more efficient face recognition performance for the intermediate view compared to the full-frontal view, which is in line with a number of previous behavioral findings (Bruce, Valentine, & Baddeley, 1987; Laeng & Rouw, 2001; Troje & Bülthoff, 1996). Contrary to previous findings (Bruce et al., 1987; Laeng & Rouw, 2001), no significant differences between views were found in response time, although there was a tendency to perform the recognition task faster after seeing a face in the intermediate view.

Importantly, the results of the present study revealed a highly significant advantage of “stereo depth” in recognition accuracy. Participants recognized more accurately faces seen in 3D compared to the 2D condition. In contrast with accuracy, analysis of RTs revealed longer response times for 3D condition compared to 2D condition. With the absence of speed–accuracy trade-off effects, this finding might seem surprising and even counterintuitive. We believe that the added time simply reflects the initial period that is needed for the eyes to fuse the left- and right-eye images in order to view the stereo image. Thus, it should have taken some extra time to select a matching face in the 3D condition compared to the 2D condition.

The observed differences in fixation time lengths between 2D and 3D conditions can be explained as reflecting variations in processing difficulty (Rayner, 1998; Underwood, Jebbett, & Roberts, 2004), as participants fixated more on the 2D face images that were harder to recognize based on accuracy rates. However,
any direct comparison of fixation durations in 2D and 3D conditions should be taken with caution, since using the glasses with colored lenses can result in some data loss when recording eye movements, compared to a condition in which lenses are not used.

Region of interest analysis of gaze data revealed similar distribution of fixation time over the face parts as indicated in previous research (Dalton et al., 2005; Pelphrey et al., 2002; Walker-Smith et al., 1977). More interestingly, we found differences in the order of ROIs that received most fixations between the 2D and 3D conditions. In the 3D condition, participants prioritized the nose and the cheeks, while they looked less at the eyes than they did in the 2D condition. Because volumetric properties become more salient in stereoscopic view, it may not be surprising that these facial features attracted more attention in the 3D condition than the less structurally informative eyes. The attention maps (see Figure 7) allow us to qualitatively visualize the location of eye fixations separately for the 2D and 3D conditions. It is clear from these maps that 2D and 3D images are viewed differently, as the distribution of gaze (i.e., lighting in the images) differs between the 2D and 3D conditions. This can be interpreted as further evidence indicating that the participants’ eye movements are sensitive to the presence of stereoscopic cues.

There are, however, several caveats regarding the methods used in Experiment 1. First, in the 3D condition, participants viewed anaglyph images through the colored glasses, while in the 2D condition standard grayscale images were presented without the glasses. This could have introduced a large low-level difference between the two conditions. Second, using the glasses with colored lenses resulted in a greater data loss when recording eye movements, compared to the condition in which the lenses were not used. Thus, to address these potential confounds, a second experiment was performed, in which face recognition performance was measured over all four combinations (i.e., by counterbalancing) of image type and glasses variables (standard grayscale images viewed without the 3D glasses, standard images viewed through the glasses, anaglyph images viewed without the glasses, and anaglyph images viewed through the glasses). To increase the complexity of our recognition task, we added the 90° angular rotation to the set of views used in Experiment 1. Finally, in Experiment 1, the two images used to generate the stereometric images were obtained by shifting the camera from one position to another position, while the model remained immobile. This method is necessarily imprecise. Hence, for Experiment 2, we acquired a digital 3D camera, with two built-in lenses, which allows taking both disparity images simultaneously.

We predicted that face recognition performance would be the highest in the stereoscopic condition, i.e., where anaglyph images are viewed through the 3D glasses. For the gaze data, we expected similar increase of fixation times for the volumetric face regions like the nose and the cheeks that we observed in Experiment 1 when analyzing eye fixation data for both 2D and 3D conditions in which the colored glasses were used.

### Experiment 2

#### Methods

##### Subjects

Nineteen female and thirteen male students (mean age = 27.47; $SD = 6.52$) of the University of Oslo, all with normal or corrected-to-normal vision, participated in the face recognition study.

##### Stimuli and apparatus

A new, larger set of stimuli was created for Experiment 2. Grayscale photographs of ten young Caucasian women and ten young Caucasian men were used as stimuli. Six photographs were taken for each model from a distance of 1 m: two pictures of a full-frontal view, two of a left-sided intermediate 22.5° view, and two of a left-sided profile view. For each pair of photographs, the models assumed a neutral, emotionless, facial expression and were asked to wear a black cap to cover the hair. Photographs were taken with a Fujifilm digital 3D camera in a room illuminated with controlled artificial lighting. The camera has two lenses separated by a 7.5-cm distance and, thus, can capture two pictures simultaneously from the two viewpoints. Subsequently, each image was edited with Adobe Photoshop software so that models’ head size was normalized, facial blemishes were touched over, and background was attenuated. Three-dimensional anaglyph images were then created with the use of RedGreen V3.0.1 software following the same procedure as described in Experiment 1.

Each final sample image was 11 × 14 cm in size. Match image items were created by putting together four smaller (8.5 × 11 cm) images of the models of the same sex, all in one of the three views. (Figure 8 shows an example of a male face in 3 two-dimensional views used as sample images and a match image item with three-dimensional full-frontal images of four male models.) The order of the images was randomized. Under each image, the number 1, 2, 3, or 4 was printed with 48-point Arial font.

The sample images were presented centered on a computer screen with a resolution of 1680 × 1050 pixels, subtending 9.04 × 11.40 degrees of visual angle. Four keys on the computer keyboard were marked as 1, 2, 3, and 4. E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA, USA) was used to present the stimuli and collect response times and accuracy rates. For viewing three-dimensional images, subjects were asked to wear red–cyan anaglyph paper glasses. Different from Experiment 1, in...
this experiment, all participants saw through the red filter with their left eye, while the right eye saw through the cyan filter. Eye movements during face recognition task were recorded with the use of a binocular Remote Eye Tracking Device (R.E.D.; SensoMotoric Instruments, Teltow, Germany) in the “eye lab” at the Department of Psychology, University of Oslo.

Procedure

The procedure was similar to that of Experiment 1. After a presentation of a fixation cross that appeared randomly in one of the four corners of the screen for 1500 ms, a sample image of a face was presented on the computer screen for 3000 ms. Then, after a 1000 ms presentation of a blank screen, a match image item consisting of an image of the same model alongside with three other models’ faces was presented for 5000 ms. Participants were requested to decide as quickly and as accurately as possible which face, among the individuals shown in the match image item, was the same one they had seen earlier as a sample image. The angle of view of a face in a sample image was always different from the view of faces in the following match image item.

The total number of trials was 240. They were organized in 4 blocks based on the combinations of viewing conditions: a block of standard images viewed without the anaglyph glasses, a block of standard images viewed through the glasses, a block of anaglyph images viewed without the glasses, and a block of anaglyph images viewed through the anaglyph glasses (note that the 3D effect was only present when anaglyph images were viewed through the anaglyph glasses). The order in which the blocks were presented was counterbalanced across the participants. Sixty trials in each block were ordered so that the combinations of all conditions (model, view, correct response) were pseudorandomized (i.e., more than three consecutive trials repeating the same condition were not allowed).

Results and discussion

Behavioral data

All behavioral data analyses were carried out using the StatView statistical package. A preliminary analysis had shown neither significant main effect nor interactive effects with any of the other factors for “Sex of the Participant.” Therefore, this factor was excluded from the following analyses.

A repeated measures ANOVA of the percentage accuracy data was performed with image type (standard, anaglyph), glasses (no, yes), sample’s view (full frontal, intermediate, profile), and model’s sex (male, female) as within-subject factors. The analysis revealed a highly significant effect of sample’s view (F(2, 62) = 13.26, p < 0.0001), with participants being significantly less accurate after seeing a face in profile view (Figure 9). This finding supports the conclusion from previous studies that show lower recognition performance for the faces in profile view (e.g., Bruce et al., 1987; Troje & Bülthoff, 1996).
In line with our predictions, there was a significant interaction effect of image type and glasses factors \(F(1, 31) = 4.31, p < 0.05\), as illustrated in Figure 10. The highest accuracy rate occurred in the condition where anaglyph images were viewed through the 3D glasses (i.e., where stereo information was present). Paired \(t\)-tests showed that in this condition, participants were significantly more accurate than in any of the other three conditions \((p < 0.05\) for each of the three comparisons). This finding supports the results from Experiment 1 by showing a significant improvement of face recognition performance in the stereoscopic condition over the other three 2D viewing conditions.

The same repeated measures ANOVA was performed for the response time (RT) data with image type (standard, anaglyph), glasses (no, yes), sample’s view (full frontal, intermediate, profile), and model’s sex (male, female) as within-subject factors. Prior to the analysis, error trials were excluded, and outliers defined as 2.5 \(SD\) from subject’s mean RT were trimmed. The only significant main effect was the effect of sample’s view \(F(2, 62) = 17.87, p < 0.0001\); see Figure 11). This advantage of the 22.5° intermediate view for face recognition resulting in the faster processing rates replicates the finding from Laeng and Rouw (2001) and supports the conclusion from Experiment 1 that seeing a face in the intermediate view leads to more efficient recognition performance.

Several interaction effects also reached significance level. The interaction effect of image type and sample’s view factors \(F(2, 62) = 5.17, p < 0.01\) yielded that for the anaglyph images, responses were faster for the full-frontal views and slower for the profiles, compared to the standard images. Scheffé’s test confirmed that the difference was significant for the profile view only. In the absence of a significant three-way interaction between image type, glasses, and sample’s view factors, this effect is rather difficult to interpret. We propose the following explanation for this finding: When anaglyphs were viewed without the 3D glasses, a “ghosting” effect of the superimposed red and cyan images might have required longer time for the observers to see the contours of the profiles. Alternatively, when anaglyph images were viewed through the 3D glasses, more information about volumetric face properties became available to the observers in addition to the profile contours, which in turn required more time to inspect the images. Taken together, these two explanations could provide a possible interpretation of the observed finding.

The interaction effect of model’s sex and sample’s view factors \(F(2, 62) = 24.05, p < 0.0001\) revealed faster recognition for female faces in the intermediate view and slower recognition in the full-frontal view, compared to
male faces. Results of the Scheffé’s test confirmed significant differences in response times between male and female faces presented at the full-frontal ($M = 2000.05$ and $M = 2122.52$, respectively) and intermediate views ($M = 2047.95$ and $M = 1885.36$, respectively).

Again, we assessed whether the perceived 3D effect depended on the deviation of the viewers’ IPDs from the distance between 3D camera lenses. The “Deviation from Optimality” scores were computed and used as the regressor with either the percentage accuracy or RT data of the stereo condition (anaglyph images viewed through the 3D glasses) as the dependent variables (Table 3). As it can be appreciated from the regression slope (Figure 4A), bigger deviations from the camera’s interlens distance predict lower accuracy rates. Similar to the results in Experiment 1, the highest accuracy scores correspond to IPDs that are slightly smaller than the distance between 3D camera lenses (Figure 4A). Again, no significant results were observed for the RT data.

### Eye-tracking data

Eye-tracking data from 16 of the 19 female participants and 9 of the 13 male participants (mean age = 26.2, $SD = 5.83$) were analyzed using BeGaze 2.3 software (Senso-Motoric Instruments, Teltow, Germany). Data from 7 participants had to be excluded due to recording artifacts. Again, the StatView statistical package was used for performing the ANOVAs.

To avoid the potential problem of quality differences in the data recorded with and without the colored glasses, data from only two conditions with the glasses—standard image with 3D glasses and anaglyph image with 3D glasses (i.e., the actual stereo viewing condition)—were used for the analysis.

The same regions of interest (ROIs) as in Experiment 1 were selected for each of the faces: eyes (left and right eyes), nose, mouth, cheeks (left cheek and right cheek), forehead, and chin. For each image and region of interest, the mean fixation time (in ms) was computed. A repeated measures ANOVA was performed on fixation time data with image type (standard, anaglyph), sample’s view (full frontal, intermediate, profile), facial part (eyes, nose, mouth, cheeks, chin, forehead), and model’s sex as within-subject factors. Sex of the Participant was not considered because a preliminary ANOVA showed no effects or interactions with this factor. Analysis revealed a highly significant effect of sample’s view ($F(2, 48) = 14.21, p < 0.001$) with shorter times spent fixating on faces in the intermediate and profile views compared to the full-frontal view. No significant differences in the fixation times were found between the standard and anaglyph image types. However, an interaction effect of image type and sample’s view factors ($F(2, 48) = 7.00, p < 0.01$) showed a significant increase of fixation time in the anaglyph image condition for the profile view.

A highly significant main effect of facial part ($F(5, 120) = 36.67, p < 0.001$) showed the following order of facial parts that received the longest fixations: eyes, nose, cheeks, mouth, forehead, and chin (Figure 12A). Comparable to the results from Experiment 1, the eyes and the nose received the longest fixation times, which is in line with a number of previously reported results (e.g., Henderson et al., 2005; Hsiao & Cottrell, 2008). Analogous to the results of Experiment 1, a significant interaction effect of image type and facial part factors ($F(5, 120) = 2.63, p < 0.05$) revealed decreased fixation times for the eyes and increased fixation times for the nose and the cheeks when participants viewed anaglyph images compared to when they looked at the standard images (Figure 12B). Again, facial regions with stronger volumetric

### Table 3

Regression analysis results for interdeviation from optimality, accuracy, and RT data for the stereo viewing condition.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
<th>$R^2$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.34</td>
<td>1.99</td>
<td>0.06</td>
<td>0.12</td>
<td>3.95</td>
</tr>
<tr>
<td>RTs</td>
<td>-0.27</td>
<td>-1.56</td>
<td>0.13</td>
<td>0.08</td>
<td>2.43</td>
</tr>
</tbody>
</table>

Figure 12. (A) Mean fixation times as a function of facial part. The error bars represent standard errors. (B) Mean fixation times as a function of facial part and image type. The error bars represent standard errors.
properties become more salient in stereoscopic view and, thus, attract more attention in the anaglyph viewing condition.

As indicated by the highly significant interaction effect of the sample’s view and facial part factors ($F(10, 240) = 10.19$, $p < 0.001$), similar fixation patterns were observed for the full-frontal and intermediate views. In the profile view, however, participants spent less time fixating on the eyes, the nose, and the mouth, whereas the cheeks received more fixation time. Comparable to the results from Experiment 1, we did not find any significant differences in gaze fixations for facial parts between the frontal and intermediate views, and the order of saliency for different facial features remained the same across the change between the full-frontal and intermediate viewpoints. This is consistent with the findings of Stephan and Caine (2007), who investigated contribution of information from different facial features for recognition across viewpoint change and concluded that with view transformations from the full-frontal to the intermediate view, the same set of facial features (eye, nose, mouth) remains equally visible and/or relevant for face recognition. Compared to Stephan and Caine, our intermediate view (22.5°) had a smaller angular distance from the full-frontal view; thus, the necessary amount of information could have possibly been extracted in similar ways from either of the views. In contrast, significantly different fixation pattern was observed for the profile view, which is in line with a number of previous findings (Bindemann et al., 2009; Sæther et al., 2009) that show the highest saliency of the cheek region for this particular viewpoint.

Finally, the significant interaction effect of the image type, sample’s view, and facial part factors ($F(10, 240) = 2.43$, $p < 0.01$) once again revealed decreased fixation times on the eyes and increased fixation times on the nose and the cheeks in the anaglyph image viewing condition compared to the standard image condition in all three views, with the largest increase observed for the cheeks in the profile view (see Figure 13). This latter finding illustrates that, in the profile view, the highly salient cheeks attract even more attention in the stereo viewing condition, where more volumetric properties become available.

**General discussion**

We compared face recognition performance across changes in viewpoint in two-dimensional and three-dimensional conditions and made the following predictions: (a) performance in the task that involves a transformation across views would be facilitated by the addition of 3D depth cues, (b) volume-rich properties of certain facial features would be attended more in 3D condition, and (c) interpupillary distance should affect the strength of the depth effect in 3D images and, consequently, the face recognition performance during 3D viewing and in this condition only.

The results of both experiments yielded a significant advantage of the stereoscopic viewing condition for recognition accuracy. By conducting Experiment 2 with a new, larger stimulus set, we confirmed that neither the anaglyph image nor the glasses alone (i.e., the low-level differences between viewing conditions) but the combination of both resulting in a “stereo depth” effect leads to higher accuracy rates. This finding is in accordance with the results of Burke et al. (2007) and Hill, Schyns, and Akamatsu (1997), since it shows that face recognition across viewpoint change becomes easier with the addition of stereoscopic depth information. However, the present results differ from the findings of Liu et al. (2006), who did not find a difference in performance between stereo and mono conditions. We believe that an explanation for these contrasting results may lie in the nature of the tasks used in different studies. In the second study of Liu et al. (2006) participants performed a sequential matching task, in which
they were asked to decide whether a face seen at learning and a test face presented afterward belonged to the same individual. As indicated by the reported sensitivity scores (about $d^\prime = 3.0$ for both stereo and mono conditions), this task might have been very easy since the participants obtained a high performance level, both when relying solely on 2D information and on 3D information. Therefore, additional stereoscopic information might not have provided any significant benefit when performing a very easy task. Likewise, in a similar sequential same–different task employed by Burke et al. (2007) the overall face recognition performance did not differ significantly between stereo and mono conditions. What differed, however, was the ability to recognize faces across viewpoint transformations, which improved with the addition of stereoscopic information.

Based on the above considerations, we surmise that the effect of depth may be revealed only when the recognition task is perceptually challenging (e.g., requiring generalization across viewpoints). In particular, selecting the correct match for a stimulus face out of multiple choices might have added the needed level of complexity to the task to reveal differences due to depth information. Note also that in our task, the test items were always presented in a different view than the previously seen stimulus image, which forced the observers to perform across-viewpoint generalization. Therefore, based on the results of the current study, we agree with Jiang et al. (2009) that learning overtime the three-dimensional structure of faces could yield better viewpoint transformation invariance than that based on two-dimensional information alone.

Another goal of the present study was to shed light on whether faces are scrutinized by gaze differently when three-dimensional cues are present compared to the flat 2D images. We confirmed the findings from previous research that when viewing 2D faces, participants spend most of the fixation time on the eyes and the nose, with increased duration of fixations for the cheeks in the profile view (Bindemann et al., 2009; Pelphrey et al., 2002; Walker-Smith et al., 1977). More importantly, our region of interest analysis revealed increased viewing time for the nose and the cheeks in stereo condition compared to the two-dimensional condition. Both the nose and the cheeks would seem to provide more volumetric information about the facial structure than the eye region (Sæther et al., 2009). Taken together with the observed better face recognition performance in the 3D condition, this finding supports our conclusion that the availability of three-dimensional structure information enhances face recognition.

Since we hypothesized that three-dimensional information facilitated face recognition in our task, we were also led to ask whether there was any between-subject variation in the subjective strength of the experienced 3D effect. Although it seems clear that IPD would play a crucial role in producing a stereo effect, it is not quite clear whether having smaller or larger IPD will significantly affect perception of stereo images. This question has been addressed in the optical engineering literature as an attempt to evaluate the accuracy of stereoscopic visual display systems used by military jet pilots for navigation. It has been shown that the most accurate performance was achieved in the condition where pilots viewed stereo with the normal IPD between the viewpoints provided to the two eyes (Merritt, Cuglock-Knopp, Kregel, Smoot, & Monaco, 2005).

Simple regression analysis results revealed that larger deviations from the camera distance used to generate 3D images predicted lower accuracy rates in the 3D condition only. In Experiment 1, participants with IPDs larger than the camera distance performed worse than those with smaller IPDs. In both experiments, the best performance was observed for those participants whose IPD was slightly smaller (i.e., about 0.5–1 cm) than the camera distance. Hence, it appears that viewing 3D images with a reduced stereo effect resulted in lower performance rates, while experiencing an either optimal or a slightly enhanced stereo effect led to better performance. To conclude, perceiving either the correct or slightly exaggerated binocular cues optimized perception of the three-dimensional face structure, resulting in a higher recognition performance.

**Conclusion**

In the current study, we were able to show that three-dimensional information facilitates face recognition performance across the viewpoint transformation. With the help of a “region of interest” analysis of gaze data, we showed that the rich volumetric properties provided by certain facial features are attended more in the 3D condition compared to the 2D condition. Taken together with our behavioral results, these findings support our general conclusion that stereoscopic depth information plays a significant role for the task of face recognition. Finally, we were able to show that the interpupillary distance affects perception of three-dimensional images and, as a consequence, subjects’ performance in the face recognition task. A simple regression analysis revealed that IPDs equal to or slightly smaller than the one used for creating the stimuli predicted better accuracy rates.

**Acknowledgments**

The authors are grateful to Jørn Lang for assistance with photography and to Berit Serina Fuglestad for help with data collection.

Commercial relationships: none.
Corresponding author: Bruno Laeng.
Email: bruno.laeng@psykologi.uio.no.
Address: Department of Psychology, University of Oslo, Box 1094, Blindern, 0317 Oslo, Norway.
References


