Advantages of 3D Methods for Face Recognition Research in Humans

Chang Hong Liu[1] and James Ward[2]

[1] Department of Psychology, University of Hull,
Hull, HU6 7RX, United Kingdom
C.H.Liu@hull.ac.uk
[2] Department of Computer Science, University of Hull,
Hull, HU6 7RX, United Kingdom
J.W.Ward@dcs.hull.ac.uk

Abstract. Research on face recognition in humans has mainly relied on 2D images. This approach has certain limitations. First, observers become relatively passive in face encoding, although in reality they may be more spontaneous in exploring different views of a 3D face. Moreover, the volumetric information of a face is often confined to pictorial depth cues, making it difficult to assess the role of 3D shape processing. This paper demonstrates that 1) actively exploring different views of 3D face models produces more robust recognition memory than passively viewing playback of the same moving stimuli, 2) face matching across 2D and 3D representations typically incurs a cost, which alludes to depth-cue dependent processes in face recognition, and 3) combining multiple depth cues such as stereopsis and perspective can facilitate recognition performance even though a single depth cue alone rarely produces measurable benefits.

1 Introduction

Research on face recognition in humans has mainly relied on 2D pictures. An implicit assumption of this approach is that face recognition in pictures reveals the truth about face recognition in general. While the resemblance between a real face and its photographic representation is hardly in question, face recognition in the real world may employ additional information that is not available in a 2D representation. The most obvious difference between a real face and a photograph is their depth information. Whereas a photograph contains only monocular depth cues, a face in reality contains both monocular and binocular depth cues. This rich depth information may allow more precise encoding of facial surface geometry.

Furthermore, unlike photographic images, faces and observers in reality are rarely stationary. Although increasing numbers of researchers use dynamic face stimuli [21], a crucial difference between perception of a face in motion pictures and a face in reality remains. Perception of a real face is an interactive process in the sense that the relative motion of the face and its observer determines resulting viewpoint of the face. The observer may from time to time actively engage in this interaction to deliberately register certain desired facial features. This means that face perception in reality in-

W. Zhao, S. Gong, and X. Tang (Eds.): AMFG 2005, LNCS 3723, pp. 244 – 254, 2005.
© Springer-Verlag Berlin Heidelberg 2005

volves the observer’s head movements and locomotion. Research methods based on 2D images do not allow researchers to assess the role of this active process.

In this paper, we will review how these issues are tackled by some recent studies that employed 3D face models and visualization methods. Before that, we will briefly review some prior research that has pioneered the use of 3D methods in psychology.

1.1 Using 3D Face Models

Despite their limitations, photographic materials are still considered as better alternatives to life models in behavioral research. The main reason that life models have never become a popular choice is that they are susceptible to extraneous variables. Photographs allow much better control of variables such as pose, lighting, and facial expression, although at expense of poorer applicability of the research to real-world face recognition.

The use of 3D models offers a promising solution to the problems facing 2D images and life models. When 3D laser-scanned models are shown in virtual environments, for example, they not only can mimic reality more closely, but also allow a

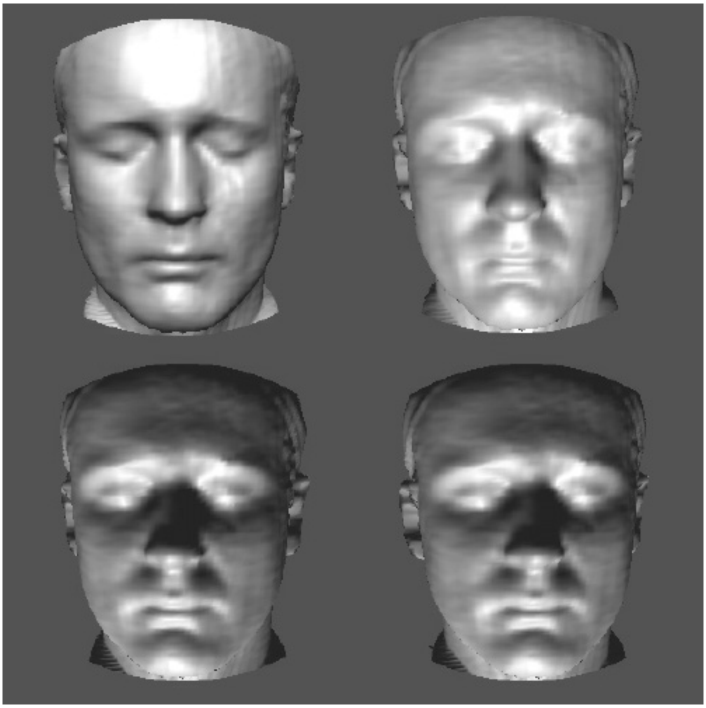

Fig. 1. Face shape defined by shading. The first row shows a top-lit and a bottom-lit face. The second row shows a bottom-lit face in a stereo pair. Bottom-lit faces are more difficult to recognize. Although this is sometimes attributed to disruption of 3D shape reconstruction from shading, stereopsis does not rectify this difficulty [20].

similar level of control over experimental variables. In the early days when the 3D method was first used in behavioral research, however, it was the potential to manipulate certain variables in face recognition that attracted most attention.

Laser-scanned faces were used in numerous studies since the early 1990s. Unlike photographs, laser-scanned models allow separation of 3D surface information from color or pigmentation information. Using this advantage, Bruce and her colleagues studied contributions of shading and surface information in face perception without presence of color or facial textures [3, 4, 5, 11, 12, 25]. Fig. 1 shows images of an example face defined by shading information alone. Although shape-from-shading is an important source of information for human face recognition, the effectiveness of this information is limited to top-lit faces [17]. The disadvantage of bottom-lit faces cannot be corrected by stereo cues, showing that shape-from-shading overrides shape-from-stereo in face processing [18].

The relative contributions of face shape and texture can be assessed by using 3D morphing techniques, where face shape and texture are selectively normalized onto an average [23, 24]. Similar techniques also allow separation of motion from shape or texture gradients from shading, making it possible to measure the contribution of non-rigid motion and texture gradients [13, 14, 19, 27].

The studies described so far have mainly attempted to separate out contributions from various sources of information in face perception. Their use of 3D methods involved reduction of information such that the information of interest could be studied in isolation. This line of approach, however, needs to be complemented by a synthetic approach, whereby different sources of information are combined rather than subtracted. Clearly, 3D models can be comfortably adapted to research in either direction. It is the combination of information that makes 3D models more credible substitutes for live models. It allows researchers to assess whether there is any difference between face recognition in pictures and face recognition in reality. At present, relatively fewer attempts have been made in this direction. In this paper, we will present some of our recent studies that were aimed to explore the roles of active exploration and 3D information commonly found in real-world face perception.

1.2 Methods of Stimulus Presentation

To examine the role of active exploration in face recognition, it is necessary to present 3D face models in real time for observer-face interaction. Most research, however, pre-renders 3D faces under the desired image conditions. For example, a study designed to measure the effect of pose change on face recognition would render 3D faces in various views. The rendered scenes are captured and saved as 2D images for later use in planned experiments. However, this method is not suitable to study the role of active exploration, because the views of a face in such a study have to be dynamically affected by the observer's exploratory actions.

We developed a software tool recently to allow the observer to explore views of a 3D face with a joystick [26]. The software is a user interface between the MATLAB programming environment and graphics libraries that operates directly on 3D face models and creates the desired image conditions in real time. Instead of retrieving pre-rendered images, joystick feedback or simple MATLAB commands are used to control simulated rigid motion of a 3D face model.

Unlike the pre-rendering method, the interactive feature reduces the difference between real faces and face images used in laboratory settings.

2 Active Exploration Improves Recognition Performance

Very little is known about whether spontaneous and active exploration of stimuli plays a role in visual cognition. The issue has only been investigated in computer generated novel objects [10]. Observers in the active condition explored the objects with a track ball during the training session, whereas those in the passive condition simply viewed the playbacks of the rotated objects generated by the active observers. It was found that active observers recognized trained objects more quickly than passive observers. However, there was no difference between the accuracy scores of the two conditions. The study leaves two unanswered questions. First, it is not clear whether recognition of a novel class of objects is readily applied to that of a familiar class of objects. Even if the answer is positive, it is still unclear whether recognition for a class of birds, for example, would be similar to recognition of faces. Unlike any other class of objects, faces are discriminated at individual level without deliberate training. Second, and perhaps more importantly, it is not clear whether recognition accuracy could be affected by active exploration. Accuracy is arguably a more important measure of identification in forensic and social settings. It is certainly also a more critical measure of face recognition in engineering.

We conducted a series of experiments aiming to answer these questions. We used a standard recognition task in a yoked design. The task involved remembering eight faces presented once at the learning session, and later identifying them at the test session where the trained faces were mixed with eight new ones. Observers in the active condition explored views of faces via a joystick, whereas observers in the passive condition simply viewed the replay of the same sequence of face stimuli generated by the active observers. The range of possible views for exploration is shown in Fig. 2. The initial view of each face was determined randomly from this range. Three experiments were conducted in which the condition of active exploration was varied.

2.1 Active Exploration at Both Training and Test Sessions

In the first experiment, active observers were allowed to explore the faces during both training and test sessions of the task. Active observers also decided how long they wished to explore each face at each trial before moving onto the next trial. Passive observers viewed the same views for the same length of time as the active observers. A total of 101 undergraduate students were randomly assigned to the two conditions.

It was found that the active condition produced better recognition accuracy ($M = 83.5\%$, $SD = 16.4$) than the passive condition ($M = 73.5\%$, $SD = 19.7$), $F(1, 99) = 10.59$, $p < .002$. The result shows that human observers acquire more robust recognition memory of faces when they are able to interactively control the views of faces.

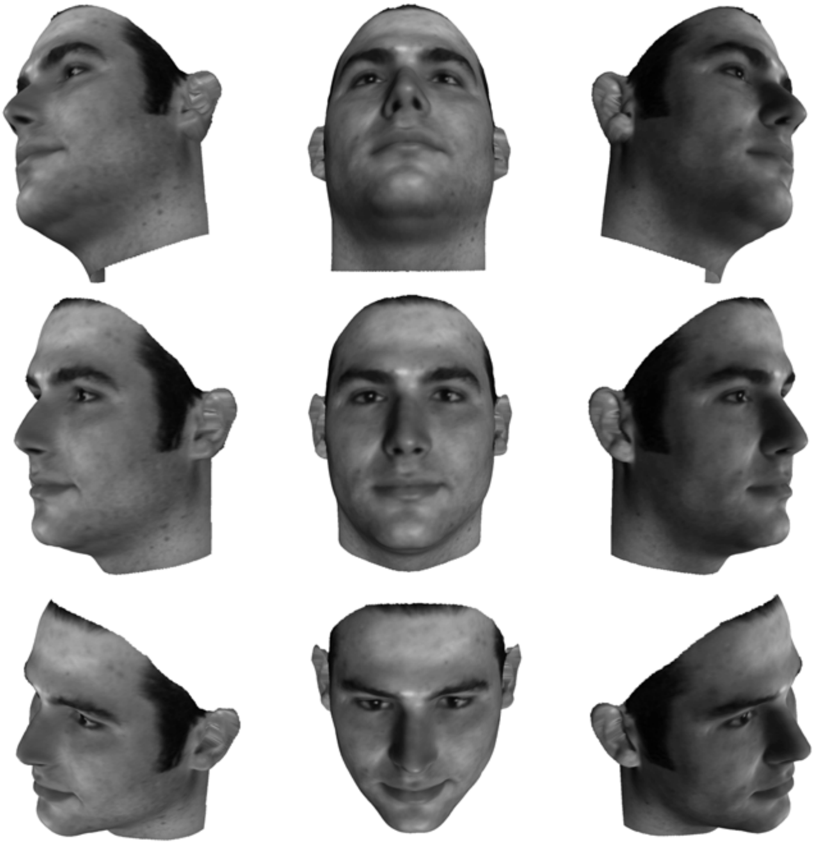

Fig. 2. Limits of rotation (x = ±55°, y = ±30°) from the frontal view in the center. The laser-scanned face database used in the present study was obtained from University of South Florida.

2.2 Active Exploration Only at Training Session

The advantage of active exploration found in the first experiment could either be due to spontaneous control of training views or correlations between the explored views at training and test. In order to determine whether active exploration at training could by itself produce better recognition performance, we ran the second experiment where active participants could explore views of faces only during the training session. A total of 80 undergraduate students participated. Other aspects of the experiment were identical to the first experiment.

The results showed again that the active condition scored higher on recognition accuracy (M = 79.1%, SD = 16.0) than the passive condition (M = 72.3%, SD = 18.6), F (1, 78) = 4.66, $p < .03$, although the mean difference between the two conditions was somewhat lower than that of the first experiment. The result shows that the advantage of the active condition cannot be simply due to correlations of the face views explored at training and test.

2.3 Active Exploration Only at Training Session Within a Fixed Duration

Both experiments described so far allowed active observers to decide the duration of face stimuli on the display. Because preferred inspection time may well differ from one observer to another, the inability to decide the length of inspection time in the passive condition may have been a disadvantage that contributed to the different performance between the two conditions. To investigate this possibility, we ran another experiment where the time to explore each face was fixed at 10 s such that observers in both conditions inspected the faces for the same duration. This specific duration was based on the data from the first experiment where observers on average spent 10.6 s (*SD* = 5.7) learning each face. Again, observers in the active condition only explored the views of faces during the training session. A total of 60 undergraduate students were tested. Other aspects of the experiment were the same as before.

Results again favored the active condition ($M = 81.5\%$, $SD = 8.9$) over the passive condition ($M = 75.6\%$, $SD = 12.8$), $F\ (1, 58) = 4.22$, $p < .05$. This experiment thus rules out the possibility that the advantage of the active condition in the previous experiments was merely due to the free control of preferred learning duration in that condition.

These experiments present the first converging evidence that actively explored face views during learning can improve the accuracy of face recognition memory.

3 Relevance of 3D Shape Information in Face Processing

Both image-based and model-based approaches to face recognition have been developed in engineering over the past decades (see [8] and [28], for a review). A key difference between these two approaches is whether they involve face processing based on volumetric information. The psychological status of these approaches remains largely unknown. Although shading and motion related effects reported in psychology are often attributed to 3D shape processing, studies to date have not yet found convincing evidence that the brain encodes 3D volumetric information of faces for recognition.

The obvious way to test the hypothesis that 3D information is used in face processing is to assess to what extent 3D cues assist face recognition. Research to date has failed to find any use of some key 3D information, such as stereopsis [18, 20] and linear perspective [15, 16]. Although stereo is often considered a more reliable cue for 3D shape than shading and other cues [6], it is shading information that prior research has found to be far more important in face recognition. For example, faces in line drawings that are devoid of shading information are more difficult to recognize than photographs [3, 7]. Stereopsis, on the other hand, hardly produces any measurable effect on face recognition. If face recognition does rely on reconstruction of 3D surface, such a result would appear rather surprising. Intuitively, the 3D shape of a face should be more easily perceived with stereo information. For example, the height of a nose from a frontal view of a face can only be accurately estimated when this information is available. The diminished importance of stereo information has prompted the conclusion that the visual system mainly uses 2D information for face processing [18]. Shading is merely treated as a 2D pattern rather than a cue to 3D shape.

However, two questions remain. First, if 2D information plays a key role in face processing, is a 3D face simply encoded as a collection of 2D images just like the way a face in photographs would be encoded? Second, can certain combinations of 3D cues facilitate face recognition even though each of these cues alone fails to do so? We addressed the first question by looking at how well faces can be identified across their 2D and 3D representations and the second question by measuring how the combination of stereo with linear perspective affects recognition performance.

3.1 Transfer Between 2D and 3D Representations

Matching a 2D face image to a 3D face requires ignoring unmatched depth cues. If a 3D face is simply encoded as 2D images, there should be no cost for such matching. To determine whether face recognition is affected by differences between 2D and 3D representations, we conducted two experiments where the dimensionality of the images used at training and test were either congruent or incongruent [20]. We used two congruent conditions. In the congruent stereo condition, both images of a face used at training and test were presented with stereo information. In the congruent mono condition, both of these images were presented without stereo information. In the incongruent condition, one of the images was presented with stereo, and the other without, and vice versa.

Face images at training and test were always presented in two different views (full-face and 3/4) to avoid matching based on trivial image similarity alone. Both stereo and mono images were observed through shutter glasses. One experiment employed a standard recognition task with a between-subject design. The task was to decide whether the faces presented at the test session had been shown at the training session. Another experiment was a matching task using a within-subject design. The task was to decide whether two sequentially displayed images were of the same person.

Both experiments found significant main effects, $F\ (2, 187) = 4.16$, $p < .02$, and $F\ (2, 112) = 6.85$, $p < .001$, for the recognition and matching experiments respectively. Bonferroni post-hoc tests showed that the congruent stereo condition produced a significantly higher accuracy ($M = 74.3\%$, $SD = 10.0$) than the incongruent condition ($M = 68.8\%$, $SD = 10.7$) in the recognition experiment. In the matching experiment, the accuracy scores in both congruent conditions (*M*s = 86.0 and 85.7%, *SD*s = 8.3 and 7.7, respectively, for the congruent stereo and congruent mono conditions) were significantly higher than that in the incongruent condition ($M = 82.3\%$, $SD = 9.4$). No difference was found between results of the congruent stereo and congruent mono conditions in either experiment.

3.2 Combination of Depth Cues May Facilitate Face Recognition

Evidence to date appears to deny any usefulness of 3D information in face processing. Nevertheless, it remains possible that the usefulness of this information depends on certain combinations of cues. We recently tested this hypothesis in a recognition task where face stimuli with several levels of perspective transformation were either presented in stereo or without stereo. Faces were trained and tested at different distances simulated on a computer screen. As Fig. 3 shows, the 2D projections of facial features and the configuration of a face can be quite different from two camera distances. In a relatively large perspective transformation such as in Fig. 3A, the projection from the near camera not only results in a larger image overall, but also produces quite considerable differences in 2D shape. It has visibly larger internal facial features such as the nose and eyes, and smaller peripheral features such as the neck and fading ears than the image projected from the far camera distance. In order to perceive the correspondence between the two projections despite their image differences, the observer needs to compensate for the differences caused by perspective transformation.

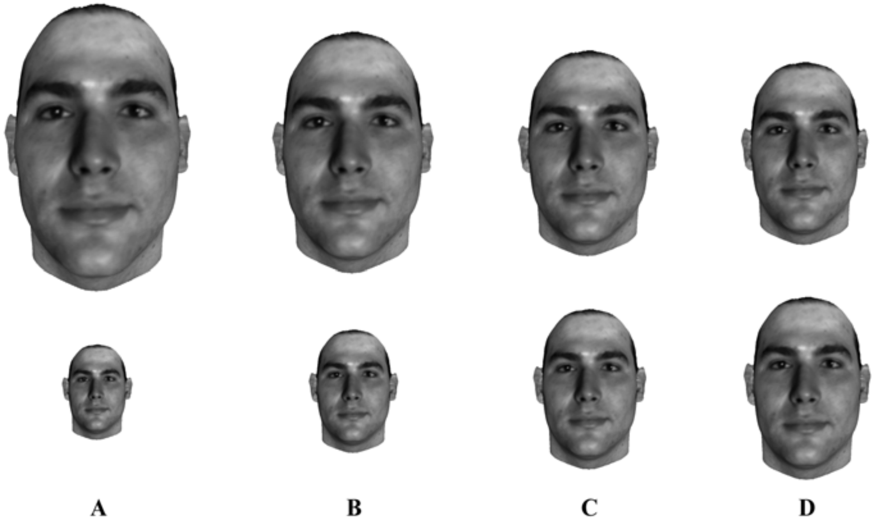

Fig. 3. Four levels of perspective transformation between training and test. A. Large transformation. B. Medium transformation. C. Small transformation. D. No transformation.

We used four levels of perspective transformation which ranged from large to no transformation. The mean visual angles subtended from the vertical extent of the faces (defined by the length from the top edge of the forehead to the tip of the chin) to the observer at these distances are shown in the second and third columns of Table 1. Each face was trained at a far distance and later tested at a near distance or vice versa. Each observer performed the task twice using two different sets of faces, once with stereo and once without. The order of the two conditions was counterbalanced.

Results for the mono and stereo condition under perspective transformation are shown in the last two columns of Table 1. We found a significant main effect of perspective transformation, F (3, 185) = 27.54, $p < .001$. Namely, the greater the distance between the training and test face locations, the more difficult it is to identify the two images as the same person. However, when stereo information was available, recognition performance was less affected by perspective transformation, $F(3, 185) = 13.72$, $p < .001$. Overall, recognition in the stereo condition was 5% better than the mono condition. Even at a small perspective transformation, the stereo condition produced better performance. The stereo advantage diminished when perspective transformation was not present.

Table 1. Percentage accuracy as a function of perspective transformation and presence of stereopsis

Perspective Transformation	Face Size		Mean Accuracy (%)	
	Near	Far	Mono	Stereo
Large	31.7°	9.9°	63.8	70.9
Medium	25.8°	12.9°	66.7	74.4
Small	22.1°	16.7°	80.2	84.3
None	19.5°	19.5°	87.2	87.5

The results show that stereo information can play a role in face processing but only when it is combined with certain other 3D cues. Clearly, not all combinations of depth cues facilitate recognition performance. For example, recognition using shading information alone is similar to recognition using both shading and stereo information as demonstrated in this and prior studies [18].

4 Conclusions

The benefits of 3D methods in behavioral research of face perception were recognized since the early infancy of the technology. However, it is not till quite recently that these methods have been applied to more realistic settings. Using 3D methods, our studies have revealed a number of previously unknown facts about face recognition in humans. First, we found that actively exploring views of a face can lead to improved learning and more robust recognition memory of that face. Second, our results showed that although faces trained and tested in 2D images produce equivalent recognition performance to faces trained and tested in 3D images, transfer from 2D to 3D or vice versa results in reduced recognition performance. Last, stereopsis can enhance face recognition when combined with linear perspective.

These findings show that face recognition may rely on more resources in reality than most research has suggested so far. These resources have not been fully investigated due to the limitations of 2D face images used in laboratory research. However, these unexplored territories can now be more systematically examined using 3D methods.

Our results show that using 3D faces in virtual environments leads to better identification performance. This may have promising implications for forensic applications. Currently, 2D mug shots and video parade systems are still dominant in this field. Simulated 3D environments offer an eyewitness the opportunity to explore any arbitrary views of a suspect in a police lineup parade, along with depth cues that are more compatible with real environments. Because recognition of unfamiliar faces is highly view or image dependent [9], being able to explore a face in various pose conditions should improve the chance of successful identification.

Understanding the benefits of 3D information in human face recognition should help engineers to determine what kind of face recognition software is useful to human users. If 3D models facilitate face recognition, it would be sensible to develop software capable of 3D face synthesis from video or photographic images for use in eyewitness identification.

Apart from practical implications, our research on the role of 3D information in face recognition was aimed to shed some light on the psychological basis of the image-based and model-based theories. Consistent with some prior findings [15, 16, 18], our recent research suggests that 3D information plays a relatively minor role in human face perception. There may be few processes in the brain for reconstruction of 3D structures, or at least such reconstruction is not necessary. However, it has also become clear that the human visual system does employ 3D information for face processing and this can at times improve or optimize recognition performance. Indeed, the utility of the 3D information may have been underestimated given that systematic research on the effects of combination of 3D cues has just started. Whether other combinations of depth cues such as motion parallax and stereopsis affect face recognition will become issues for future research.

Acknowledgement

This research was supported by the British Academy and the Nuffield Foundation.

References

1. Braje, W.L., Kersten, D., Tarr, M.J., Troje, N.F.: Illumination effects in face recognition. Psychobiology, Vol. 26, (1998) 371–380
2. Bruce, V., Burton, A.M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., Linney, A.: Sex discrimination: How do we tell the difference between male and female faces? Perception, Vol. 22, (1993) 131–152
3. Bruce, V., Hanna, E., Dench, N., Healey, P., Burton A.M.: The importance of “mass” in line drawings of faces. Applied Cognitive Psychology, Vol.6, (1992) 619–628
4. Bruce, V., Healey, P., Burton, M., Doyle, T., Coombes, A., Linney, A.: Recognising facial surfaces. Perception, Vol. 20, (1991) 755–769
5. Bruce, V., Langton, S.: The use of pigmentation and shading information in recognising the sex and identities of faces Perception, Vol. 23, (1994) 803–822
6. Bülthoff, H.H.: Shape from X: Psychophysics and computation. In: Landy, M.S., Movshon, J.A. (Eds.): Computational Models of Visual Processing. The MIT Press Cambridge MA (1991) 305–330
7. Davies, G., Ellis, H. D., Shepherd, J.: Face recognition accuracy as a function of mode of representation. Journal of Applied Psychology, 63, (1978) 180–187
8. Hallinan, P.L., Gordon, G.G., Yuille, A.L., Giblin, P., Mumford, D.: Two- and Three-Dimensional Patterns of the Face. A. K. Peters Limited Natick MA (1999)
9. Hancock, P.J.B., Bruce, V., Burton, A.M.: Recognition of unfamiliar faces. Trends in Cognitive Sciences, Vol. 4, (2000) 330–337

10. Harman, K.L., Humphrey, G.K., Goodale, M.A.: Active manual control of object views facilitates visual recognition. Current Biology, Vol. 9, (1999) 1315–1318
11. Hill, H., Bruce, V.: The effects of lighting on the perception of facial surfaces. Journal of Experimental Psychology: Human Perception and Performance, Vol. 22, (1996) 986–1004
12. Hill, H., Schyns, P.G., Akamatsu, S.: Information and viewpoint dependence in face recognition Cognition, Vol. 62, (1997) 201–222
13. Hill, H., Jinno, Y., Johnston, A.: Comparing solid body with point-light animations. Perception, Vol. 32, (2003) 561–566
14. Knappmeyer, B., Thornton, I.M., Bülthoff, H.H.: The use of facial motion and facial form during the processing of identity. Vision Research, Vol.43, (2003) 1921–1936
15. Liu, C.H.: Is face recognition in pictures affected by the center of projection? IEEE International Workshop on Analysis and Modeling of Faces and Gestures, (2003) 53–59
16. Liu, C.H., Chaudhuri, A.: Face recognition with perspective transformation. Vision Research, Vol. 43, (2003) 2393–2402
17. Liu, C.H., Collin, C.A., Burton, A.M., Chaudhuri, A.: Lighting direction affects recognition of untextured faces in photographic positive and negative. Vision Research, Vol. 39, (1999) 4003–4009
18. Liu, C.H., Collin, C.A., Chaudhuri, A.: Does face recognition rely on encoding of 3-D surface? Examining the role of shape-from-shading and shape-from-stereo. Perception, Vol. 29, (2000) 729–743
19. Liu, C.H., Collin, C.A., Farivar, R., Chaudhuri, A.: Recognizing faces defined by texture gradients. Perception & Psychophysics, Vol. 67, (2005) 158–167
20. Liu, C.H., Ward, J., & Young, A.W.: Transfer between 2D and 3D representations of faces. Visual Cognition. (in press)
21. O'Toole, A.J., Roark, D., Abdi, H.: Recognizing moving faces: A psychological and neural synthesis. Trends in Cognitive Sciences, Vol. 6, (2002) 261–266.
22. O'Toole, A.J., Edelman, S., Bülthoff, H.H.: Stimulus-specific effects in face recognition over changes in view-point. Vision Research, Vol. 38 (1998) 2351–2363
23. O'Toole, A.J., Price, T., Vetter, T., Bartlett, J.C., Blanz, V.: Three-dimensional shape and two-dimensional surface textures of human faces: The role of "averages" in attractiveness and age. Image and Vision Computing Journal, Vol. 18, (1999) 9–19
24. O'Toole, A.J., Vetter, T., Blanz, V.: Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: An application of three-dimensional morphing. Vision Research, Vol. 39, (1999) 3145–3155
25. Troje, N.F., Bülthoff, H.H.: Face recognition under varying poses: The role of texture and shape. Vision Research, Vol. 36, (1996) 1761–1771
26. Ward, J., Liu, C.H.: VRVision: A new tool for the display of 3D images in behavioural research. Behavior Research Methods, Instruments, & Computers. (in press)
27. Watson, T.L., Johnston, A., Hill, H.C.H., Troje, N.: Motion as a cue for viewpoint invariance. Visual Cognition. (in press)
28. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, J. Face recognition: A literature survey. Technical Report, CS-TR4167R, University of Maryland (2002)