Camera Distance from Face Images

Arturo Flores, Eric Christiansen, David Kriegman, and Serge Belongie

University of California, San Diego
9500 Gilman Drive, La Jolla, CA, USA
{aflores,echristiansen,kriegman,sjb}@cs.ucsd.edu

Abstract. We present a method for estimating the distance between a camera and a human head in 2D images from a calibrated camera. Leading head pose estimation algorithms focus mainly on head orientation (yaw, pitch, and roll) and translations perpendicular to the camera principal axis. Our contribution is a system that can estimate head pose under large translations parallel to the camera's principal axis. Our method uses a set of exemplar 3D human heads to estimate the distance between a camera and a previously unseen head. The distance is estimated by solving for the camera pose using Effective Perspective n-Point (EPnP). We present promising experimental results using the Texas 3D Face Recognition Database.

1 Introduction

When photographing a human head, the subject's appearance can vary dramatically depending on the camera's distance from the subject. This variation is caused by perspective distortion, and for 3D objects cannot be undone by simply adjusting focal length; see Figure 1 for an illustration using a synthetic head[1].

This distortion presents a problem for automatic cross-condition face recognition, e.g. webcam-based recognition from social media images. Even humans find such recognition difficult [1,2]. It is also a source of information, allowing camera pose estimation in cases where the subject is known [3]. This information could potentially be used to undistort the images and improve recognition results.

In this paper, we show camera distance estimation from 2D images is possible even when the subject is previously unseen. Our technique replaces the known-subject assumption with knowledge of the general distribution of fiducials across people. This distribution turns out to be sufficiently tight to allow surprisingly accurate distance estimation using only a small training set.

The paper is organized as follows. Section 2 covers related work from psychology and computer vision. Section 3 explains our method. In Section 4, we validate our method on the Texas 3D Face Recognition Database. Section 5 is the discussion and conclusion.

[1] Generated using FaceGen (http://www.facegen.com).

G. Bebis et al. (Eds.): ISVC 2013, Part II, LNCS 8034, pp. 513–522, 2013.

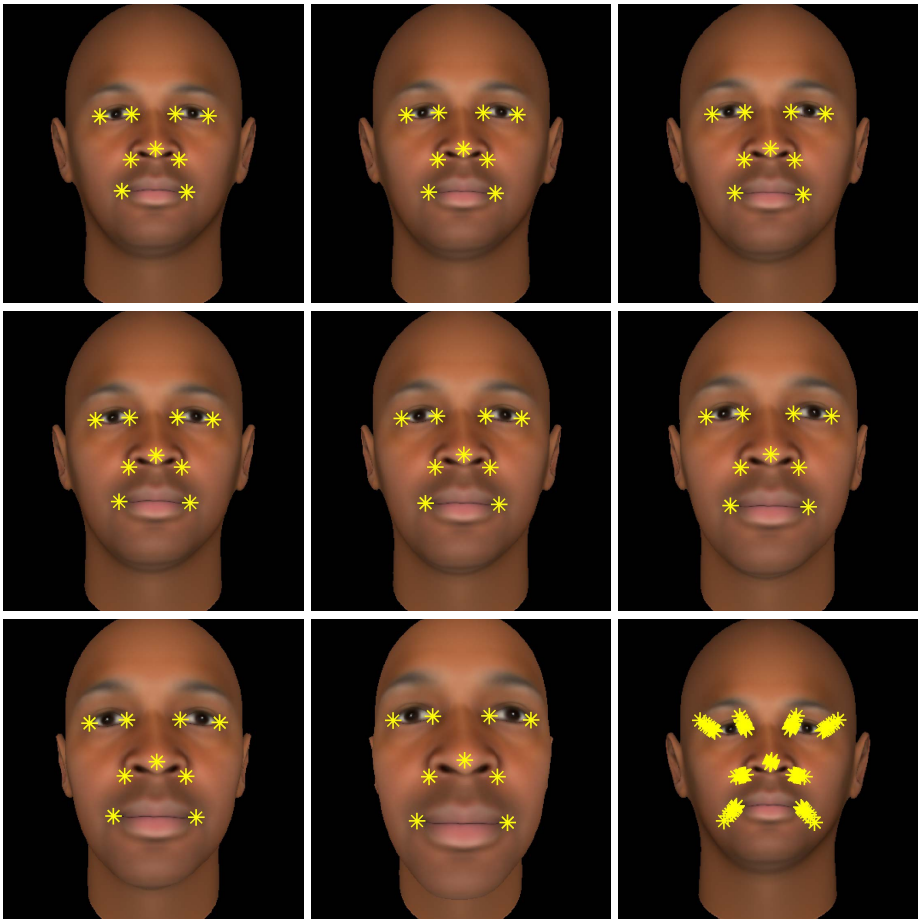

Fig. 1. A synthetic head viewed from different camera distances, illustrating projective distortion. Camera distance decreases in the first eight images, indexed in row-major order. Here, the focal length (zoom) is adjusted to keep the figure at a constant size. Fiducials are shown as red dots. In the first image, the camera is far away, resulting in near orthographic projection. In the eighth image (bottom row, middle column) the camera is very close to the human head. The last image is the same as the first, but with fiducial markers from all images. This illustrates the migration of fiducials as a function of camera distance and focal length.

2 Related Works

There is previous work on head pose estimation from 2D images; see [4] for a recent survey. However, most methods attempt to recover a subset of the yaw,

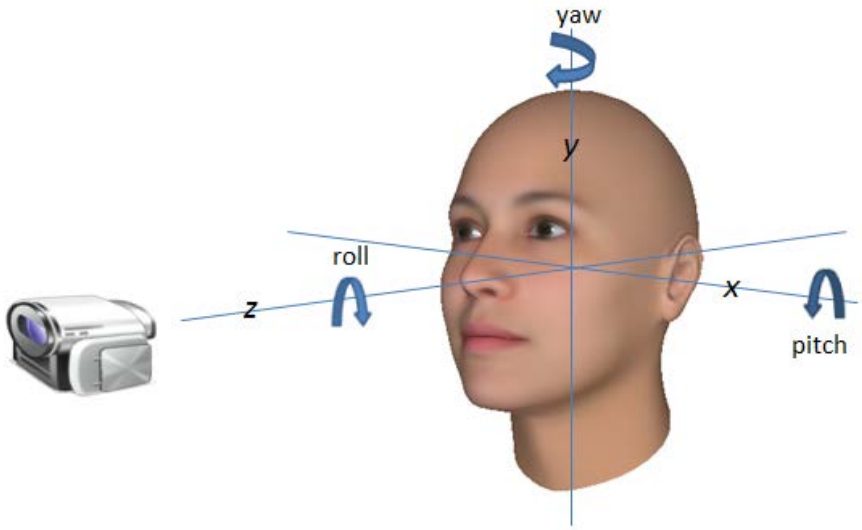

Fig. 2. An illustration of the six degrees of freedom governing head pose relative to a camera. Prior work has focused on estimation of yaw, pitch, and roll [4]. In this paper, we assume these parameters are known and estimate camera distance from the subject, shown here as distance along the z axis.

pitch, and roll of the head with respect to the camera. To our knowledge, these methods do not attempt to estimate the distance between the camera and head; Figure 2 illustrates the difference. In this section we discuss works most similar to our own.

In [1,2], Liu et al. study the effect of face recognition by humans when viewing faces at different perspective convergence angles (effectively focal length or field of view). The study involved a training phase in which face images were displayed to a human test subject. In a later recognition phase the subject was shown images of faces and asked to determine whether each face had been previously displayed. The field of view was changed to see if this had an effect on recognition. Their results show that even humans have a hard time recognizing a face when viewed under different levels of perspective distortion. This is a motivating factor since if humans have trouble with this task, a computer vision algorithm will likely also have the same troubles. Predicting the distance between camera and face is a first step in mitigating the effects of perspective distortion.

A similar psychology based studied is presented in [5,6]. Here, Perona et al. investigate the effects of perspective distortion as visual cue for social judgement of faces. Human subjects were asked to judge an image of a face in terms of

trustworthiness, attractiveness, and competence. Their results show that for social judgements, pictures taken up close are generally rated lower, while pictures taken far away have higher ratings.

In automatic camera calibration, camera parameters are recovered using prior information about the imaged scene. In [7], Deutscher et al. recover camera parameters under the assumption the scene satisfies a Manhattan world criterion. This is similar to our technque, which assumes human fiducial locations are approximately distributed according to an estimated distribution. In [8], Lv et al. use a human subject for calibration, but unlike this work requires multiple frames of video. In [9], Krahnstoever and Mendonca use a full human body for calibration from a single image, where this work uses just the head.

In [3], Ohayon and Rivlin present head tracking as a camera pose estimation problem. Prior to head tracking, 3D points are acquired from the head. During tracking, correspondences between the 3D points and their imaged 2D points are used to estimate head pose by solving the inverse problem, namely camera pose. They use the Perspective n-Point (PnP) formulation to solve for the camera extrinsic parameters (rotation R and translation T). The PnP method is also known as the Location Determination Problem and was first coined in [10]. In effect, the human head is used as a calibration rig. In this paper, the authors show that head pose can be accurately estimated and tracked under varying yaw, pitch, and roll and translations about x and y axes. However, they assume knowledge of the ground-truth fiducial locations and do not address dramatic changes in camera distance.

3 Camera Pose Estimation from Face Images Using EPnP

Our work is based on [3], but we focus primarily on translation along the z-axis, which affects the level of induced perspective distortion. We present a method for estimating the pose of a previously unseen human head using a dataset of exemplar human heads.

Efficient Perspective n-Point (EPnP): The method uses EPnP, a fast, non-iterative, solution to the PnP problem [11]. We use code provided by the authors[2]. As stated earlier, the PnP problem is to estimate the pose of a calibrated camera from n 3D-to-2D correspondences. In particular, EPnP enables us to estimate camera pose based on a set of 2D fiducial locations and their corresponding 3D locations.

Exemplar 3D heads: Note the geometric configuration of fiducial features varies from face to face, but in general fiducial locations tend to form clusters, as illustrated in Figure 3. This means the fiducial locations of a new person are

[2] `http://cvlab.epfl.ch/software/EPnP`

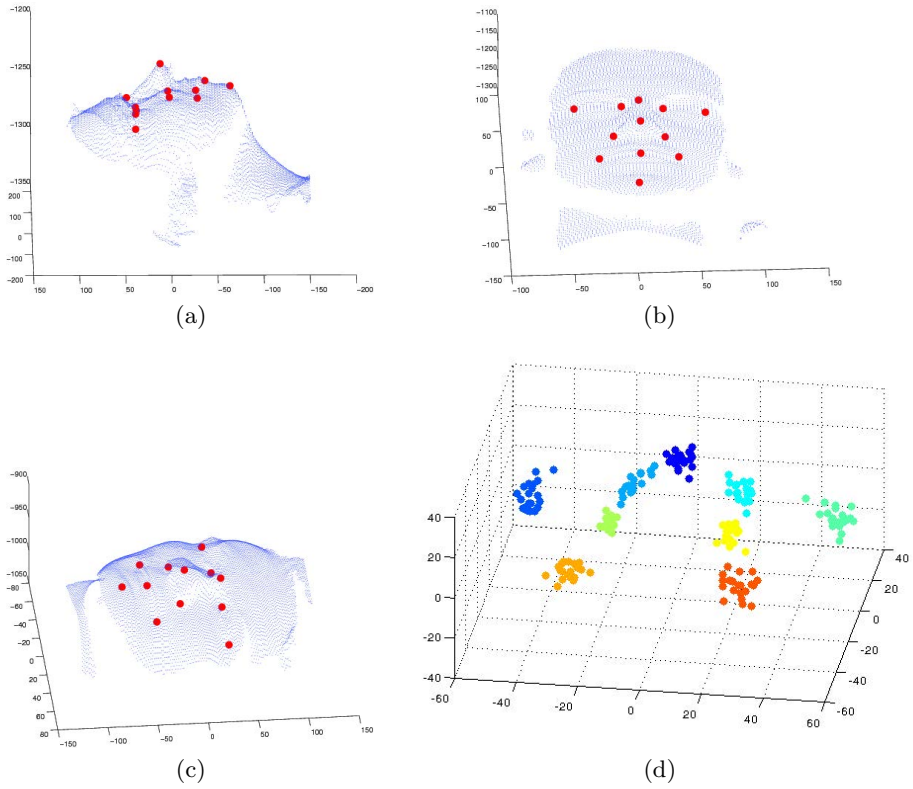

Fig. 3. (a-c) show point clouds of three different faces taken from the GavaBDB [12] dataset. Manually clicked fiducial locations are marked with red dots. (d) shows fiducials for 20 different faces, aligned by subtracting their respective means. Note the fiducials form tight clusters, meaning fiducials do not change dramatically across individuals. This phenomenon enables robust depth estimation even for previously unseen images.

likely to be similar to those in an exemplar set. We take advantage of this by using a set of exemplar 3D heads to estimate the camera pose of a novel head.

The method: The method is based on simple averaging, leveraging the observation from the previous paragraph. Suppose we get an image I of a previously unobserved head. For each exemplar 3D head E, the camera pose is estimated via EPnP under the assumption the fiducials of I match the fiducials of E. This assumption is incorrect, but as mentioned in the last paragraph, it is not far off. The estimated camera distance for I is just the average of the camera distances across all the exemplars.

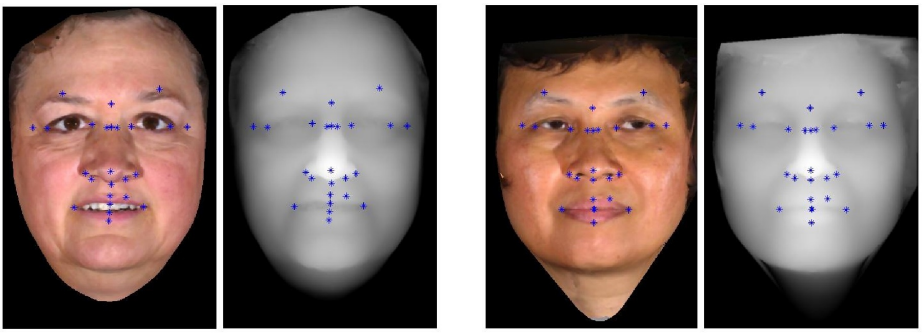

Fig. 4. Example images from the Texas 3DFRD, a database of 3D scans of human faces. These images show the 25 fiducials labeled for each face, as well as example depth maps.

4 Experimental Setup

In this section, we show the surprising effectiveness of our method for estimating camera distance in images of previously unseen subjects.

Our experiments are run against the Texas 3D Face Recognition Database (Texas 3DFRD) [13]; example images are shown in Figure 4. This database is a collection of 1149 pairs of frontal face color and depth images of 105 adult human subjects. Each face also has 25 manually labeled anthropometric facial fiducial points. Color and depth images were captured simultaneously and are perfectly coregistered. This provides ground truth 3D locations of the fiducial locations.

Our first experiment consists of the following:

- Project fiducials for a test subject onto an image plane.
- Use 3D fiducial locations from a set of reference individuals (not including the test individual) as exemplars.
- Estimate the camera distance for the test subject using the method described in Section 3.
- Repeat while simulating a dolly-zoom (or Hitchcock zoom) camera movement, in which the camera moves away as focal length is increased to keep the figure size constant.

We perform this experiment for the frontal and 3/4 profile view of the face. For the frontal case, we assume all fiducials are visible to the camera. For the 3/4 we assume only a subset of the fiducials are visible. We also assume we have a calibrated camera and all intrinsic camera parameters are known. Simulated camera distances range from approximately 10cm to 3m. Note if we did not

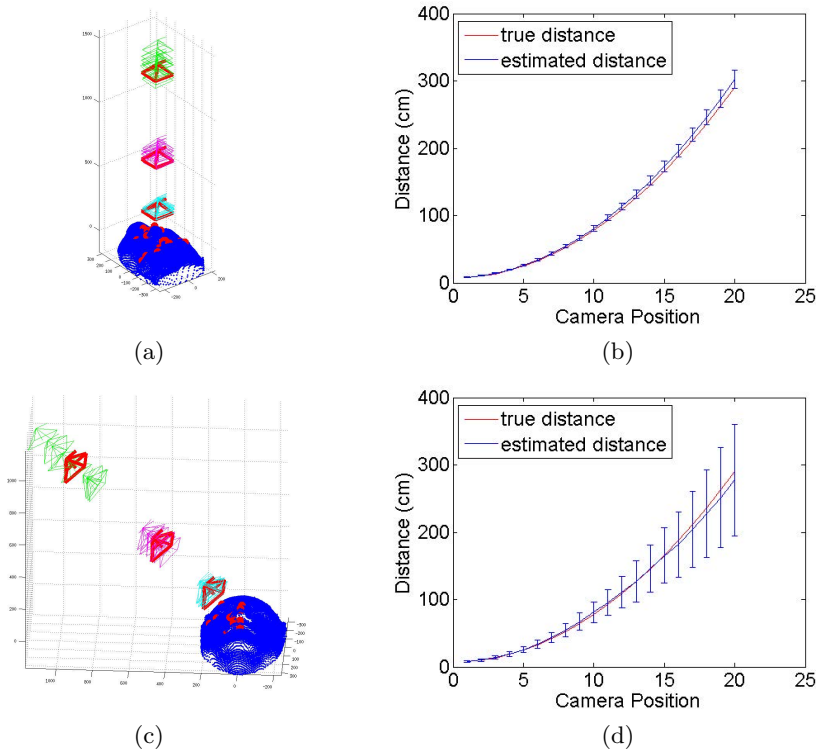

Fig. 5. Results of camera pose and distance estimation for the frontal and 3/4 profile views. (a) and (c) show clusters of estimated camera poses for three distances for the frontal and 3/4 profile views. The clusters of poses for each of the three distances are marked using cyan, magenta, and green camera frames, with the ground truth pose marked in red. (b) and (d) show true and estimated camera distances as a function of camera position, where distance is measured from the camera center to the tip of the nose. We test 20 camera positions. The estimated distances are shown with error bars to represent the variation across test subjects. Note the final distance estimate closely follows the true camera distance, though the error bars get larger as the distance increases. This is especially true for the 3/4 profile view, presumably due to a fewer number of visible fiducials.

adjust the focal length, the distance from the camera to the face could be trivially estimated by the size of the imaged face. For this reason, the focal length is adjusted to keep the outermost fiducials at a near constant distance, which also keeps the imaged face silhouette at a constant size. Results are shown in Figure 5. Our method nearly perfectly recovers camera distance.

Figure 6 shows the same experiment where the number of fiducials is varied. For this experiment, we manually selected the fiducial subsets to be evenly

distributed about the face. This figure shows the distance can be reliably estimated with as few as five fiducials. In the case of five fiducials, the fiducials used were outer corners of eyes and mouth, and center of top lip. When using only four fiducials, the center of top lip was removed, resulting in 4 nearly co-planar points. At this point, the distance estimate becomes unreliable.

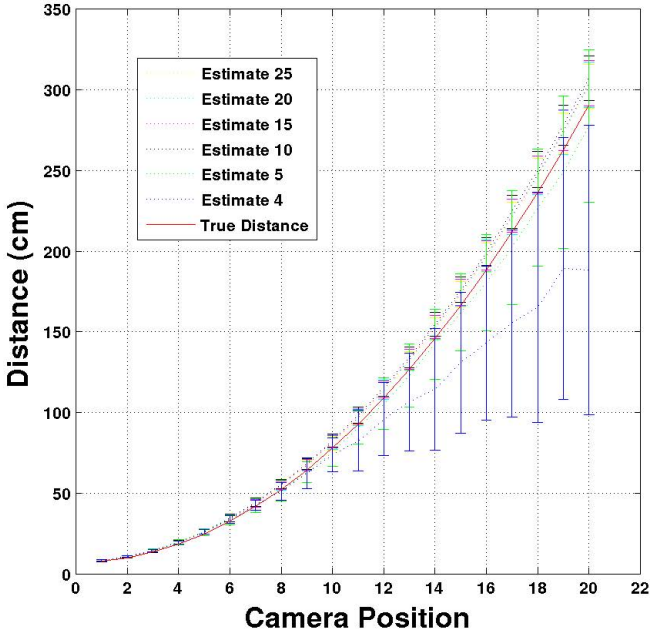

Fig. 6. Results of camera distance estimation for the frontal view for varying numbers of fiducials. As in Figure 5, error bars represent the variation in estimated distance across test subjects. This shows distance can be reliably estimated with as few as five fiducials.

For qualitative analysis, we use a heuristic to select a closest exemplar for each subject and camera distance. Let p_i be the 2D imaged fiducial using the true camera position. Let p'_{ij} be the imaged fiducial using the estimated camera position from the j-th exemplar. We use $\text{argmin}_j \sum_i \|p_i - p'_{ij}\|$ as a heuristic to select the best exemplar. See Figure 7 for some illustrative examples. These examples show fiducial configuration is more than just a means to undistort images; it is a source of biometric information in its own right.

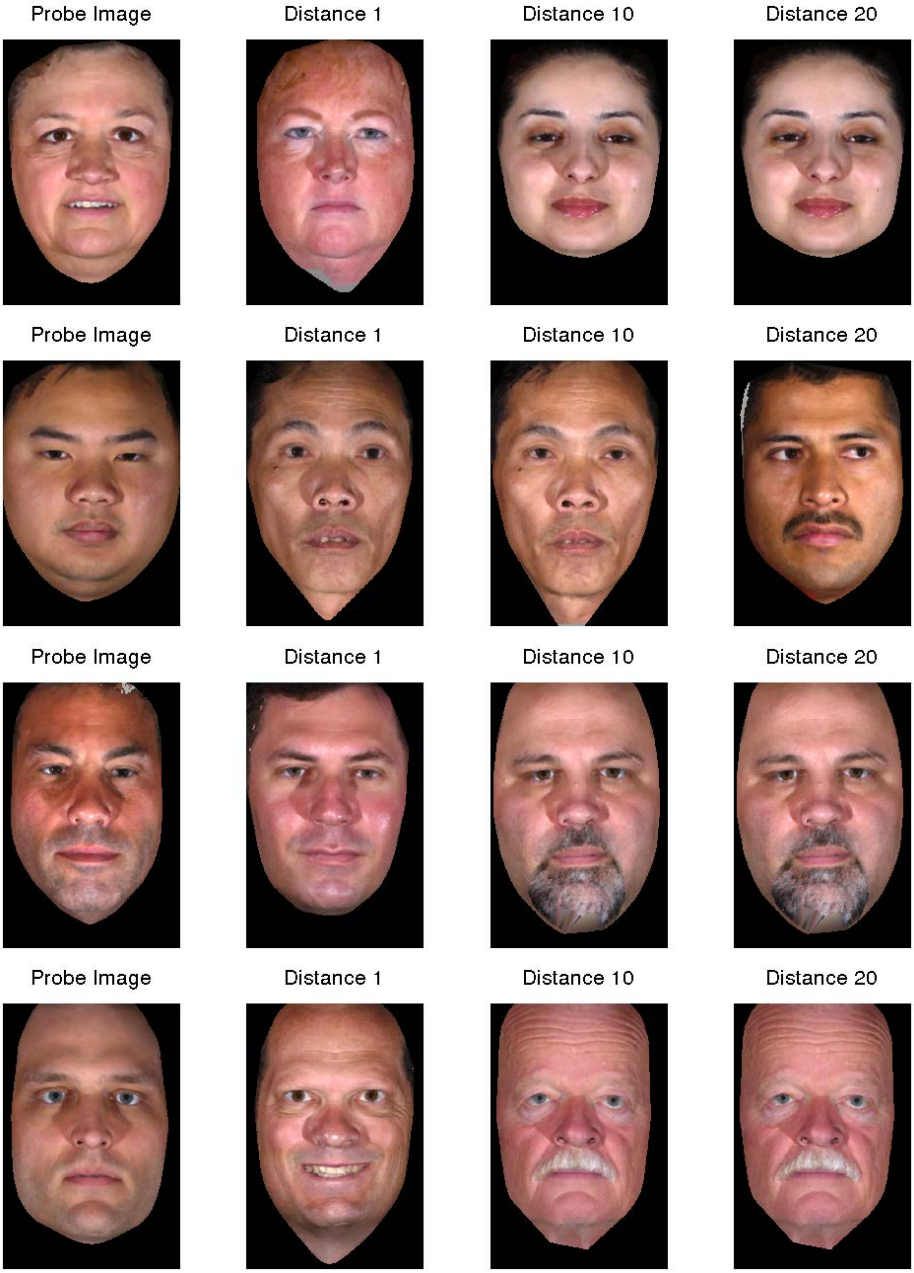

Fig. 7. The closest exemplar as a function of camera distance for four subjects (probes). The probes are shown in the first column. The best exemplars for 10cm, 75cm, and 300cm are shown in the second, third, and fourth columns, respectively. Note the probes tend to match exemplars with similar attributes, e.g. race and gender.

5 Conclusion and Future Work

We presented a method for estimating camera distance to a previously unseen face. The method uses correspondences of 2D image fiducials to 3D fiducial locations on exemplar heads. Though the method is simple, it is accurate for distances ranging from 10cm to 300cm for the frontal and 3/4 profile views. This estimate could be used to mitigate the effects of perspective distortion for face recognition. The method could also be used as a direct source of biometric information, by leveraging the attributes of the closest matching exemplars.

Future work may replicate these results on real images. We may also investigate automatic fiducial localization [14], or naturally ocurring features such as in [3]. Finally, we may estimate more general poses, such as the full extrinsic camera parameters.

This work was supported by ONR MURI Grant #N00014-08-1-0638.

References

1. Liu, C.H., Chaudhuri, A.: Face recognition with perspective transformation. Vision Research 43, 2393–2402 (2003)
2. Liu, C.H., Ward, J.: Face recognition in pictures is affected by perspective transformation but not by the centre of projection. Perception 35, 1637 (2006)
3. Ohayon, S., Rivlin, E.: Robust 3d head tracking using camera pose estimation. In: 18th International Conference on Pattern Recognition, ICPR 2006, vol. 1, pp. 1063–1066. IEEE (2006)
4. Murphy-Chutorian, E., Trivedi, M.M.: Head pose estimation in computer vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 607–626 (2009)
5. Perona, P.: A new perspective on portraiture. Journal of Vision 7, 992–992 (2007)
6. Bryan, R., Perona, P., Adolphs, R.: Perspective distortion from interpersonal distance is an implicit visual cue for social judgments of faces. PloS One 7, e45301 (2012)
7. Deutscher, J., Isard, M., MacCormick, J.: Automatic camera calibration from a single manhattan image. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 175–188. Springer, Heidelberg (2002)
8. Lv, F., Zhao, T., Nevatia, R.: Camera calibration from video of a walking human. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 1513–1518 (2006)
9. Krahnstoever, N., Mendonca, P.R.: Bayesian autocalibration for surveillance. In: Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 2, pp. 1858–1865. IEEE (2005)
10. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 381–395 (1981)
11. Lepetit, V., Moreno-Noguer, F., Fua, P.: Epnp: An accurate o (n) solution to the pnp problem. International Journal of Computer Vision 81, 155–166 (2009)
12. Moreno, A.B., Sanchez, A.: Gavabdb: a 3d face database. In: Proc. 2nd COST 275 Workshop on Biometrics on the Internet, Vigo, Spain, pp. 75–80 (2004)
13. Gupta, S., Castleman, K.R., Markey, M.K., Bovik, A.C.: Texas 3d face recognition database. In: 2010 IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI), pp. 97–100. IEEE (2010)
14. Belhumeur, P.N., Jacobs, D.W., Kriegman, D.J., Kumar, N.: Localizing parts of faces using a consensus of exemplars. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 545–552. IEEE (2011)