Response to Comment on “100% Accuracy in Automatic Face Recognition”

R. Jenkins* and A. M. Burton

Contrary to the suggestion of Deng et al., image registration reduced face-recognition accuracy when divorced from the averaging procedure. Average-to-photo mapping generalizes beyond specific photographs, and averaging either gallery images or probe images can improve the match. The alternative protocol suggested by the authors is unsuitable because it evaluates face-matching algorithms, not face representations, and relies on standard image sets.

We reported that the process of image averaging can dramatically boost automatic face recognition (1). Deng et al. (2) suggest that image registration alone might improve face-recognition performance, and we tested this suggestion. Because the MyHeritage database (3) is constantly expanding, we first re-submitted the photographs and average images used in (1) to establish a current baseline. Forty-eight of the 500 probe images were identical to images in the online gallery, compared with 41 in (1). This increase is consistent with gallery expansion. Of the remaining 452 photographs, 52% were correctly identified, down from 54% in (1). The hit rate for the average images was 100%, as before. Five of the average images matched different photos of the correct person, confirming that the average-to-photo mapping generalizes beyond particular snapshots. To address Deng et al.’s concern, we next submitted manually registered versions of the source photographs. As Deng et al. describe, these were aligned in a standard frontal and upright posture and enclosed by a uniform background. The hit rate for the registered images was 30%. Apparently, registration alone offers the worst of both worlds: It disrupts any informative correspondence in shape between gallery and probe images but does not otherwise stabilize image variability. Registration of the probe images might be less harmful when the gallery images are also registered. In a previous study using a principal components analysis–based image match (4), we carried out exactly this transformation. Performance was poor but was nonetheless improved by averaging.

Deng et al. (2) also express concern that our average images were presented as probes rather than being gallery items. This was a consequence of our chosen methodology. To ensure a stringent test of our averaging technique, we relinquished control over several key aspects of the image match. We used someone else’s gallery photographs together with someone else’s matching algorithm. Our probe images were collected from the Internet. This approach meant that we were not able to add images to the gallery, but we could still submit images as probes. Because face recognition can be reduced to matching pairs of images, the order of each pair was not our main interest, and we treated matching A to B as equivalent to matching B to A. In previous studies, we have shown that averaging also helps when applied to the gallery images (4). Whether identity checks would be better served by an average image stored in an identification document or an average probe generated from the live face is an interesting empirical question. However, it is worth pointing out that averaging probe images specifically finds practical application in forensic face recognition (5).

Deng et al. point out that an average probe need only match one gallery photo of the target to score a hit. The same is true for the photographic probes, yet these performed comparatively poorly. In practice, an average probe can match very different photos of the target, as our new data confirm. This underscores the major benefit of averaging. Matching pairs of photos is extremely difficult, because both items contain information that is not diagnostic of identity. Matching a photo to an average is helpful because it eliminates nondiagnostic information from one item in the pair. There is no doubt that difficulties can still arise in this situation, but this is partly because the pair still includes a photograph. Our response is therefore not to retreat to matching pairs of photos but rather to investigate ways to eliminate photos from the match altogether. Matching pairs of average images is an obvious route to explore, and we are testing this possibility.

Deng et al. recommend the Face Recognition Vendor Test (FRVT) (6) as a methodological template. This is unsuitable for several reasons. First, the FRVT evaluations compare performance of different matching algorithms on standard images. Our proposal concerns the representation of the face and is independent of the matching algorithm. Second, the standard databases consist of posed photographs, which grossly underrepresent the variability of ambient face images. Third, reliance on any standard database carries the risk of solving “database recognition” without tackling face recognition. The real world presents different crowds on different days, and systems aspiring to real-world application cannot ignore this inconvenience.

Finally, we agree with Deng et al. that early processing and automatic feature extraction are interesting problems, but they are clearly separate from the problem of face recognition. To convince yourself of this, note that it is easy to locate landmarks on a face you cannot recognize and that doing so does not trigger identification.

References

21 April 2008; accepted 16 July 2008
10.1126/science.1158428

*To whom correspondence should be addressed. E-mail: rob@psy.gla.ac.uk
Response to Comment on "100% Accuracy in Automatic Face Recognition"

R. Jenkins and A. M. Burton

Science 321 (5891), 912.
DOI: 10.1126/science.1158428