Face Recognition by humans: Nineteen results all computer vision researchers should know

Akarsha Sehwag
2015010
IIIT Delhi

The author starts with how the current algorithms only deal with constrained environments with all the attendant variability in imaging parameters such as sensor noise, illumination etc and ends with how human visual systems recognize faces and vaguely describes why it's likely to be useful for computer vision researchers. They brought together several diverse results as an endeavour to provide the user a comprehensive understanding of face recognition in human vision. The author discusses nineteen results which broadly classified into five themes.

A. Recognition as a function of available spatial resolution

The first result demonstrates that fine featural details are not necessary to obtain good face recognition performance. Although the diagnosticity resides in the overall configuration of the individual features. And the important aspects of the configuration and how to encode them has been left as an open qustion to the reader. Moving onto the 2nd result, we see how the ability to tolerate degradations increases with familiarity. And thus the benchmark for evaluating machine-based face recognition system is said to be human performance with familiar faces.

3 result suggests that high spatial information, by itself, is not adequate cue for human face recognition processes since the images which exclusively contain contour information are very difficult to recognize. Line drawing appears to be sufficient for recognition because it contains significant photometric cues.

B. Nature of Processing: Piecemeal Versus Holistic

Feature processing is important for facial recognition since the features are sometimes sufficient for recognition. But the pattern of results siggest that configural processing is equally important and the facial recognition is dependent on holistic processes involving an interdependency between featural and configural information.

The next result discusses that out of different facial features, eyebrows are among the most important for recognition. This is because eyebrows appear to be important in conveying emotions and other non-verbal signals and they serve as very stable features.

The 6 result is that important configural relationships appear to be independent across width and height dimensions. It's observed that to a limited extent, rotations in depth around x and y axes approximate two-dimensional compressions.

C. Nature of cues used: Pigmentation, Shape and Motion

Result 7 suggests that face-shape appears to be encoded in slightly caricatured manner. This seems absurd when we think of it intuitively but demonstrations proove that some departures from veridicality are actually beneficial for human face recognition. Even drastic compressions of face don't render them unrecognizable. Thus we consider caricature effects which tend to be strongest in degraded images.

Next we see how prolonged face viewing can lead to high level aftereffects, which suggests prototype based encoding. Seeing any sort of aftereffect following prolonged viewing of a face stimulus provides proof for norm-based contrastive coding of faces. The induced after-effect are suspected to be the results of adaptation at relatively high levels of the visual system.

Result 9 shows that pigmentation cues are at least as important as shape cues. The author argues that the two kinds of cues are used about equally by humans to recognize faces. And this has been investigated and the subjects have performed equally well using either shape or pigmentation cues.

Color cues play a significant role specially when shape cues are degraded. Color may facilitate this task by supplementing the luminance based cues and thereby leading to a better parsing of a degraded face images.

Next result says that contrast polarity inversion dramatically impairs recognition performance, possibly due to compromised ability to use pigmentation cues. When pigmentation cues are unavailable, recognition is significantly worse with negative contrast.

Coming onto result 12, it states that Illumination changes Influence Generalization. This has been proven in a recent study that humans are capable of generalizing representation of face to radically novel illumination conditions. This generalization to this is not perfect though.

Result 13 discusses that the view-generalization appears to be mediated by Temporal Association. Basically, they suggest that the temporal proximity of images is a powerful tool for establishing object representations.

Result 14 suggests that motion of faces appear to facilitate subsequent recognition. Face motion is more than just a sequence of viewpoints to the face recognition system. The dynamic cues from expressive movements provide information about facial feature aspects and thus have multiple viewpoints.

D. Developmental Progression

Result 15 states that visual system starts with a rudimentary preference for face-like patterns. A simple arrangement of three dots within an oval may serve as an appropriate template for detecting faces in the initial stages of face-learning systems in babies.

Next result discussed about the visual system progress from a piecemeal to holistic strategy over first several years of life. Over the course of time, a shift in strategy occurs from infants adopting largely piecemeal, feature-based strategy to a more sophisticated holistic strategy involving configural information.

E. Neural Underpinnings

Moving onto result 17, which states that Human visual system appears to devote specialized neural resources for face perception. It's suggested that rather than being a true face module, the fusiform face area(FFA) may be responsible for either subordinate or expert-level categorization of generic objects. The author could not explain the concept in great detail, but said that the FFA provides a valuable set of constraints for computational systems indicating the extent of selectivity and generality we should expect from face recognition systems.

The next result states that latency of responses to faces in IT cortex is about 120ms, suggesting a largely feedforward computation. While impoverished images likely require iterative processing, cleaner images can be dealt with quick. This also provides a constraint on face recognition algorithm since it indicates the information to be extracted immediately and might not be ”cleaned up” later.

The last result of the author describes how facial identity and expression might be processed by seperate systems. Theoritical models and neural systems propose the above mentioned statement in facial perception pathway, leaving them to act in parallel using distinct representations. Calder and young pointed out that there are certain neurons responding to both factors. But they agree that representations like identity, expression, and identity expression are processed largely independently at the end.

Conclusion

Concluding the paper, author writes about how the understanding of visual system can facilitate and in turn be facilitated by better computational model. They hoped that their results will help the computer vision initiatives to create face recognition systems which can match or even exceed the ability of humans.

References

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=arnumber=4052483