Similarity Perception of Visual objects: A Machine-Learning Approach

Na-Yung Yu[1], Takashi Yamauchi[2], Ricardo Gutierrez-Osuna[3]

[1, 2]Department of Psychology, Texas A&M University, College Station, Texas, USA nayungyu@tamu.edu [1]

[3]Department of Computer Science, Texas A&M University, College Station, Texas, USA

Abstract

Though a traditional assumption in similarity judgment is that people selectively attend to certain features, few studies have explored the actual method of identifying "salient features" in visual stimuli. In this study, we used complex, realistic images to examine whether people selectively process salient features. Stimuli were triads of original and morphed animal face pictures. In the behavioral experiment, participants viewed two original pictures and a morphed composite of the originals and decided which original picture was more similar to the morphed picture. In the computational analysis, we employed Gabor function and wavelets, Gray-Level Co-occurrence Metrics (GLCM) combined with principal component analysis to extract candidate visual features, such as Gabor texture, brightness, size, and contour for the entire face as well as parts of the face. The simulated annealing algorithm was applied to behavioral data to determine possible weight distributions for the candidate features. The analysis suggests that people selectively attend to a few features when comparing visually complex and realistic images.

How do people perceive similarity between visual objects? Similarity research has traditionally assumed that people attend to matching and mismatching features selectively (Sloutsky and Fisher, 2004; Tversky, 1977). Although researchers have demonstrated this idea using semantic concepts and their verbal attributes (Lee & Zeinfuse, 2008), few studies have investigated whether people selectively use a small number of features to perceive similarity of complex visual stimuli. In this article, we collected behavioral data from human participants and applied image-processing and machine-learning techniques to identify the visual features that were consistent with similarity judgments from the behavioral data In brief, we present a computational method to find possible weight distributions for candidate features in visual similarity judgment of animal faces.

In the behavioral experiment, human participants judged similarity of original and morphed animal pictures. We collected ten original animal face photographs and created five animal pairs (i.e., bear-fox, cow-pig, hippo-sheep, koala-rat, and lion-horse pair). For each pair, one original picture (i.e., source) was merged with the other original picture (i.e., target) in Morphman 4.0 (2003). Eighteen morphed pictures were generated for each animal pair. Altogether, 90 morphed pictures (5 animal pairs X 18 morphed pictures) and 10 original pictures were used in the experiment (Figure 1). Participants viewed two original pictures of each pair and one morphed picture of the pair and judged which original picture (left or right) was more similar to their morphed composite (for a similar task, see Sloutsky and Fisher, 2004). The proportion of participants selecting the source picture (e.g., the original hippo face for the hippo-sheep pair) was recorded.

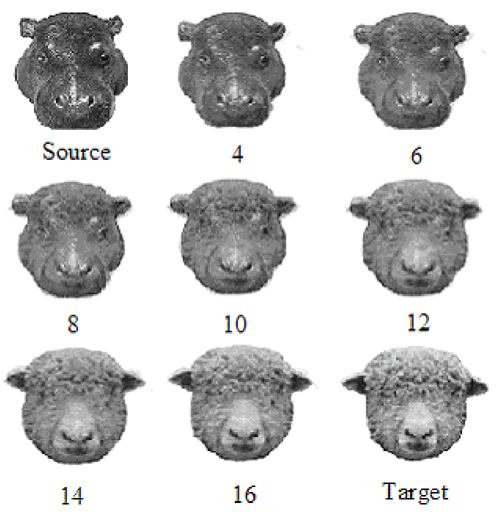

Figure 1: Sample stimuli

To obtain a computational analogue of the behavioral data, we first identified 37 potential visual features that our participants might have used for similarity judgment of animal faces. These candidate features included texture difference, relative brightness and size, and contour of the faces. To obtain textural information, we computed Gabor-based textures

(Manjunath & Ma, 1996) and the rate of co-occurring features (Howarth & Ruger, 2004), which have commonly been used in the field of content-based image retrieval. To measure brightness, we averaged the gray values of the image. We captured the relative size of the faces by computing the ratio of width to height of individual face pictures. Finally, to identify features related to the contour of the faces, we measured the distance from the center of each face to the contour and performed principal component analysis. Because people can extract and attend to features from the entire face as well as from specific parts of the face (e.g., Ullman, 2006), texture, brightness, and size features were extracted from the whole image as well as from nine sub-regions, which roughly corresponded to different parts of the animal face.

To determine the relative salience of the 37 features, we computed a weight for each feature such that our computational measure of similarity between face images paralleled the similarity judgment data obtained in the behavioral study. Our rationale was that salient features should have higher weight values than less salient features (Sloutsky & Fisher, 2004; Tversky, 1977). We first measured the Euclidean distance between input image and the source image $d(X_i, S)$, and input image and the target image $d(X_i, T)$; for the use of Euclidean distance see Nosofsky (1986). X_i denotes the test image, S denotes the source image, and T denotes the target image in the same animal pair. Our behavioral data represented the probability that one input image (X_i) was judged to be more similar to the source (S) than to the target (T). To simulate participants' probability scores, we employed Luce's choice model where the distance was transformed as follows:

$$sim'(X_i, S) = 1 - \frac{d(X_i, S)^P}{d(X_i, S)^P + d(X_i, T)^P} = \frac{d(X_i, T)^P}{d(X_i, S)^P + d(X_i, T)^P} \qquad (1)$$

Parameter P denotes the power function of the measured distances; for the use of the parameter P see Nosofsky, Gluck, Palmeri and McKinley (1994).

To identify salient features people may use in the similarity judgment, the weight distributions for the candidate features were estimated. The pattern of feature weights that minimized the energy E was deemed an optimal allocation of feature weights, and those features that garnered large weights were extracted as salient features (Russell & Norvig, 2002). To obtain the weight vector of the features, we employed Simulated Annealing (SA), a probabilistic algorithm that is known to find reliable approximations of global optima in high dimensional spaces (Kirkpatrick, Gelatt, & Vecchi, 1983).

The results suggest that people focus on a small number of facial features to judge similarity of animal faces. For example, our analysis suggests that, in judging similarity between hippo and sheep faces, the most salient dimension was a holistic feature, the brightness of the entire face. For the bear and fox faces, however, our analysis suggests that local features as well as holistic feature may be used (e.g., brightness of the center of the face, size of the mouth, and contour of the entire face). The results imply that people may process animal faces not only in a holistic manner as often reported in the study of human face recognition (e.g. Bukach, Gauthier, & Tarr, 2006) but also in a piecemeal fashion as reported in the study of visual object recognition (e.g., Ullman, 2006).

References

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *TRENDS in Cognitive Sciences, 10*, 159-166.

Howarth, P., & Ruger, S. (2004). *Evaluation of texture features for content-based image retrieval.* Paper presented at the International Conference on Image and Video Retrieval.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*, 671-680.

Manjunath, B. S., & Ma, W., Y. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 18*, 837-842.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39-57.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., & McKinley, S. C. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition, 22*, 352-369.

Russell, S., & Norvig, P. (2002). *Artificial intelligence: A modern approach (2nd ed.)*: Prentice Hall.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*(2), 166-188.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352.

Ullman, S. (2006). Object recognition and segmentation by a fragment-based hierarchy. *TRENDS in Cognitive Sciences, 11*, 58-64.