Object Similarity by Humans and Machines

Cong Yang and Marcin Grzegorzek
Research Group for Pattern Recognition
University of Siegen, Germany

Introduction

In this paper, we briefly address a research regarding how to objectively evaluate machine-based object similarity measures by human-based estimation. Based on a novel approach for similarity measure of 3-D objects (Feinen et al. 2013), we create a ground truth of 3-D objects and their similarities estimated by humans. The automatic similarity results achieved by (Feinen et al. 2013) are evaluated against this ground truth in terms of precision and recall in an object retrieval scenario. To further illustrate the reciprocity properties between machine and human perception, we compare the similarities achieved by both on testing data and show how it can be used to address other problems and formulations.

Object Similarity by Machines

In order to compare two objects based on their skeletons, utilising the proposed features of (Feinen et al. 2013), we compute the distance between their feature vectors.

$$sim_{\cos}(\mathbf{f}_n, \mathbf{f}_m) = \frac{\langle \mathbf{f}_n, \mathbf{f}_m \rangle}{\|\mathbf{f}_n\| * \|\mathbf{f}_m\|} \quad (1)$$

In addition to this, we rate the quality of a query with well-known indicators from the field of information retrieval, namely *recall* (completeness) and *precision* (accurateness).

Object Similarity by Humans

In order to collect the Ground Truth (GT) by human perception, we employed a group of volunteers (consisting of 15 students from different research disciplines) and asked them to rate the similarity of each object pair within our database according to their understanding of similarity. Since the number of objects in each class is not equal and the manner of human perception is influenced by the diagnosicity hypothesis, the optimal hitting set in this experiment is estimated in a heuristic way as follows:

1. For each object O_h all remaining GT objects are arranged in a descending order within a list according to their corresponding GT similarity values ($s(O_h)$).
2. Afterwards, we compute all differences of similarity values as shown in Eq. 2 where d_{O_v} indicates the similarity value of an object according to its position (v).

$$\Delta(d_{O_v}, d_{O_{v+1}}) \quad (2)$$

3. Finally, we detect the position of the fourth largest delta value (Δ) and select all objects above this row as part of the optimal hitting set (H_G).

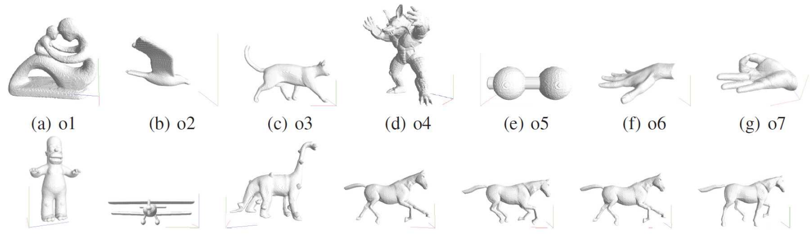
An example of the procedure is shown in Table 1. The decision to use the fourth largest difference as a threshold is based on empirical observations. Besides this, the actual rating of similarities, which has been performed by our volunteers, was unrestricted. Every test person was free to choose continuously different perspectives on the objects in order to rate their similarity.

list	$s(O_h)$	$\Delta(d_{O_v}, d_{O_{v+1}})$	Δ-Position	$O_v \in H_G$
O_i	0.9	-	-	√
O_j	0.65	0.25	1	√
O_k	0.64	0.01	7	√
O_l	0.45	0.2	3	√
O_m	0.3	0.15	4	x
O_n	0.06	0.24	2	x
O_o	0.02	0.04	5	x
O_p	0.0	0.02	6	x

Table 1: Proposed heuristic to select whose objects that build an optimal hitting set H_G for an arbitrary query object O_h.

Evaluation

The objects of our data set are shown in Fig. 1. Naturally, the human skill to detect simililarities is based on fuzzy degrees and thus is not binary. However, our tests discovered two major issues. (i) Expressing this vagueness in numbers constitutes a challenging job for humans and (ii) this makes it even harder to arrange these values subsequently in a consistent and uniform distributed way. This arrangement proficiency is also affected by the individual's perception of fuzziness and how the test person assesses and handles this perception gap. To simplify the expression of vagueness we limited the information to two categories: *similar* and *not*



(a) o1 (b) o2 (c) o3 (d) o4 (e) o5 (f) o6 (g) o7

(h) o8 (i) o9 (j) o10 (k) o11 (l) o12 (m) o13 (n) o14

Figure 1: As shown, most of the objects are articulating. Articulated objects are highly suitable to work with skeletons.

Q	Human	Machine	R	P
O_1	$O_2, O_3, O_4, O_5, O_6, O_8, O_{10}, O_{11}, O_{12}, O_{13}, O_{14}$	$O_3, O_4, O_8, O_9, O_{10}, O_{11}, O_{12}, O_{13}, O_{14}$	0.73	0.89
O_2	O_3, O_5, O_6, O_7, O_9	$O_1, O_3, O_4, O_5, O_6, O_7, O_9, O_{10}, O_{11}, O_{12}, O_{13}, O_{14}$	1.00	0.42
O_3	O_{11}, O_{13}	O_7, O_9, O_{11}	0.50	0.33
O_4	$O_1, O_3, O_8, O_{10}, O_{11}$	$O_1, O_8, O_9, O_{10}, O_{11}, O_{12}, O_{13}, O_{14}$	0.80	0.50
O_5	O_6, O_7	$O_1, O_2, O_3, O_4, O_7, O_8, O_9, O_{10}, O_{11}, O_{12}, O_{13}, O_{14}$	0.50	0.08
O_6	$O_1, O_2, O_3, O_5, O_7, O_8, O_{10}$	O_3, O_7, O_{11}	0.29	0.67
O_7	$O_1, O_2, O_5, O_6, O_8, O_9, O_{10}$	O_3, O_6	0.14	0.50
O_8	$O_1, O_3, O_4, O_6, O_{10}, O_{11}$	$O_1, O_4, O_9, O_{10}, O_{11}, O_{12}, O_{13}, O_{14}$	0.67	0.50
O_9	$O_2, O_5, O_7, O_8, O_{11}$	$O_1, O_4, O_8, O_{10}, O_{11}, O_{12}, O_{13}, O_{14}$	0.40	0.25
O_{10}	O_{11}	$O_1, O_4, O_8, O_9, O_{11}, O_{12}, O_{13}, O_{14}$	1.00	0.13
O_{11}	O_{13}	$O_1, O_3, O_4, O_9, O_{11}, O_{12}, O_{13}, O_{14}$	1.00	0.13
O_{12}	O_{14}	$O_1, O_4, O_8, O_9, O_{10}, O_{11}, O_{13}, O_{14}$	1.00	0.13
O_{13}	O_{11}	$O_1, O_4, O_8, O_9, O_{10}, O_{11}, O_{12}, O_{14}$	1.00	0.13
O_{14}	O_{12}	$O_1, O_3, O_4, O_8, O_9, O_10, O_{11}, O_{12}, O_{13}$	1.00	0.11

Table 2: Retrieval Results on objects in Fig. 1 (Q: Query, R: Recall, P: Precision).

similar. Consequently, we lost the possibility to employ a rational differentiation between these two quantities. As a result of this, we were forced to compare between two scales: (i) an ordinal scaled range obtained by humans and (ii) a continuous and metric interval-scaled range derived from our similarity measure.

Additionally, we have to consider a very interesting and crucial phenomena which we call ”unconscious background-knowledge”. This means, humans refer to unconscious relationships between objects during the similarity rating. For example, it is not surprising that the object group consisting of Fig. 1(o11) to Fig. 1(o14) obtains a high similarity value. But in the case of the objects of Fig.1(o5), 1(o6) and 1(o7), the results are unexpected. However, considering the human unconscious background-knowledge, the results are replicable and thus the results are factual connections based on the human mind.

The evaluation results can be found in Table. 2. Altogether, most of the results are quite promising and it can been assumed that system rates similarity of 3-D objects according to a certain degree of human perception.

Discussion

As shown in Section 4, the object similarity achieved by humans can be used as ground truth for machines. However, in many cases, human conception cannot be used directly. This is due to objects in an image being considered to be different by human and machine. For example, in Fig. 2, the black area is the input object for a machine-based similarity measure. However, with human conception, the graph background would be a meaningful object, in which you can find four letters LIFT. This phenomenon shows that Cognitive Psychology can also be applied for machine-based similarity measures. Since our contact with the world is through our senses, the question which arises is whether we see reality or whether what we see is guided by expectation. We normally treat a result retrieved by machine as wrong because the retrieved object is not what we expected. However, studies on visual illusions make it clear that we often make mistakes when viewing our environment (Fulcher 2003). For a machine-based object retrieval system, we can generate some meaningful features with an image background rather than drop it.

Figure 2: Example of an illusionary graph. Word **LIFT** is hidden in this graph.

References

Feinen, C.; Barnowsky, D.; Paulus, D.; and Grzegorzek, M. 2013. Robust 3-d object skeletonisation for the similarity measure. In *ICPRAM*, 167–175.

Fulcher, E. 2003. *Cognitive Psychology*. Crucial.

Tversky, A. 1977. Features of similarity. *Psychological Review* 84:327–352.