Multimedia retrieval is increasingly important with the skyrocketing multimedia volumes produced every day. Therefore many image and video retrieval tools are being developed utilising visual similarity modelling algorithms for similar image retrieval or various visualisations. As such, the quality of the similarity modelling is crucial for these tools. This thesis explores diverse similarity models, their agreement with human perception of similarity and possible improvements of these models. The examined similarity models consisted of colour-based, SIFT-based, and DNN-based models. For the purpose of model evaluation, a user study was conducted to create a dataset of relative image similarity comprising both generic images as well as two compact domains. In this study, the participants were asked to state which of the candidate images was more similar to the query image. The collected data showed the superiority of DNN-based models compared to other evaluated variants. Nonetheless, all similarity models performed significantly better than a random guess. In order to further enhance the performance of the similarity models, we fine-tuned the best-performing model (W2VV++) with the collected dataset and achieved significant improvement in some areas.