Image Categorization

Ina Klein Teeselink, Frans Blommaert and Huib de Ridder

IPO, Center for User-System Interaction, Eindhoven, The Netherlands

A study was conducted to investigate whether images of natural scenes can be categorized with respect to information content and whether a relation exists with perceived foreground-background separation. In an experiment, one group of subjects carried out a 'free categorization' task, (subjects were free to choose similarity criteria), while another group of subjects sorted images with respect to 'foreground–background' separation. A diverse set of photographic pictures was used in this experiment. The created clusters were analyzed by applying nonmetric multidimensional scaling and hierarchical cluster analysis. The results indicate that subjects are able to classify images in a consistent manner. Another outcome of this study is that observers classify images into categories that reflect scene content. The most striking image feature based on scene content is the presence of people or not. The second important image feature is whether these images show landscapes or objects/buildings. The results of the 'free categorization' task compared to the 'foreground–background categorization' task show some interesting correlations. Pictures of landscapes nearly always present only background information. A portrait demonstrates almost only foreground information. Pictures of objects or buildings, on the other hand, show a combination of fore- and background information.

Journal of Imaging Science and Technology 44: 552–559 (2000)

Introduction

We look differently at different images. Sometimes we can even look quite differently at the same image. This fact has to do with the diversity of the task of vision, which is to provide the biological system with visual information that is needed to perform a variety of information processing activities successfully. Visual information is quite diverse: it ranges from material and spatial information about the outside world to information that is more abstract such as in texts. The usefulness of information is also task dependent: we look different at the same scenery depending on whether we are driving a car or just enjoying nature.

Task and image content guide our viewing strategy: the task determines which aspect of the image we are interested in while image content determines which information is available. Although much is known about visual information processing, we are still far away from understanding how task and image content guide eye fixation and duration in dynamic viewing.

In the technological world of image processing and display, the question of viewing strategy plays a special role. The reason for this is that it takes a lot of data to represent visual information. This might be done more economically if we knew beforehand which part of an image is most useful for a viewer, and which part serves only as background information. The obvious technique is to take advantage of the fact that we fixate the objects of interests with our fovea, which possess far more resolving power than the more peripheral parts. Therefore, background information can be represented more sparsely. The ideas have led, among others, to the introduction of "importance maps"[1] in image display techniques. Another issue in image reproduction research is the fact that perceived display and coding quality depends on image content.[2,3] Although this may be partly attributed to signal properties, such as the statistic of edges, it is probably also related to the perceived category of information.

The usual method of studying the perceptual weighting of image content is by recording and analyzing eye movements when observers are looking at displayed images.[4] The technique we use in this study looks more like the method applied by Rao and Lohse[5] in their study to identify relevant high level features in texture perception. In our study, we let viewers categorize images and try to find out (1) if there are well defined perceived categories in image content, (2) if there is a clear notion of foreground-background separation, and (3) if there is a generalized relation between perceived categories and perceived foreground–background separations. The rationale behind this technique is the notion that the class of images that is photographed or filmed is a selection of all possible images constrained by the fact that there are specific objects of interest in at least most of the images. This must be, since the photographer or camera person decided the image or image sequence worthwhile to be recorded at all. Not only was the scene worthwhile to be recorded but it was also recorded in a specific format (e.g., constrained with respect to size and position of objects of interest). Furthermore, the observer's task in viewing photographs or displayed images is probably also constrained. The usual task here is to reconstruct and recognize the displayed objects of interest.

Original manuscript received February 23, 2000

One of the goals of this research is that it should eventually lead to knowledge about which parts of which images provide the more useful information to an observer and hence which image parts must be represented less sparsely. Another goal is to find out what the influence of information category is on perceived image quality in order to be able to adapt display and coding properties to image category.

The approach we have adopted is bottom-up: We let observers categorize a set of photographic images into categories of (1) their own choice, and (2) foreground–background separation. Results are then analyzed by using standard techniques such as cluster analysis and multi-dimensional scaling.

Experiment

Method

Subjects. Eighteen subjects recruited from students and staff of the Institute for Perception Research (IPO) participated in this experiment. All subjects were volunteers. Ten subjects served in the 'free categorization' task, eight served in the 'foreground–background categorization' task. None of the subjects participated in both tasks. An approximately equal number of males and females took part in each assignment.

Stimulus Materials. The stimulus set consisted of 37 color photos of natural scenes, taken from a Kodak Photo CD.[6] We selected these photographs because the set appeared to be sufficiently representative. In a comparable pilot experiment we used an extended set of 91 photographic images; the results of categorization tasks were quite identical. All pictures were printed on a Kodak XL 770 color printer. Small copies of the 37 Kodak images are shown in Fig. 1.

Procedure

Subjects were told that they took part in an experiment concerning categorization of images. We randomly arranged 37 photos on a table in front of the subjects. Each subject was tested individually. All pictures were shuffled before they were shown to the next subject.

'Free Categorization' Task. Subjects were instructed to categorize the 37 photos into groups of similar images. Subjects were free to choose their own similarity criteria. First each subject was asked to arrange the images into two groups such that the images within each group appear to be as similar as possible and the two groups appear to be as different as possible. After subjects finished this grouping task, we asked them to describe the created clusters and to point out the most and least representative picture of each group. These remarks helped us to understand the created clusters. After the task of two groups was completed, subjects were instructed to form three, then four, and then five groups. There was no time limit to complete the task, but the assignment took about 30 min per subject.

'Foreground–Background Categorization' Task. Here, the subject's task was to sort the pictures into groups with relation to foreground–background distances of the images. Subject were told to do this in such a manner that one extremity of an imaginary line contained a group of photos presenting only background information, for instance an image of a landscape. The other extremity consisted of pictures demonstrating only foreground information, like a passport photo. In between these extremities subjects had to construct groups of images that presented more or less background, in respective to foreground information. Just as in the 'free categorization' task, subjects were asked to construct first two, then three, four and five groups. Again there was no time limit. After each grouping task, subjects gave a description of the categories and indicated the most and the least representative photo per group.

Results and Discussion

For each subject and for each grouping task, a 37×37 matrix was constructed with both the horizontal and the vertical axis, the names of the 37 photos. The matrix contained only zeros and ones, based on pairwise similarities among the 37 pictures. A one in the matrix indicated that the subject put two images in the same category. A zero meant that two pictures were assigned to different categories.

First, we used these matrices to see if there were subjects that did something completely different from the other subjects. We therefore compared the scores of each pair of subjects within both categorization tasks. As a measure of similarity, we computed matching coefficients and Jaccard's coefficients between each pair of individuals. The first coefficient is the ratio of the total number of cell entries on which two individuals match (i.e., 1–1, 0–0), to the total number of cell entries. The second is the corresponding ratio when 'negative' matches (i.e., 0–0) are ignored.[7] Two matrices were constructed; one based on the matching coefficients and the second based on the Jaccard's coefficient. These two matrices were created for each grouping task and for each categorization task. We applied complete linkage cluster analysis on all matrices (see next section). The Jaccard's cluster analysis as well as the matching analysis showed no obvious outliers among the subjects.

After we examined the subjects, the matrices of all subjects serving the 'free categorization' task and the matrices of all subjects who participated in the 'foreground–background categorization' task were added for each grouping assignment. This resulted into eight joint similarity matrices (two, three, four and five groupings for both the 'free' and the 'foreground–background categorization' task). The resulting entries in the four matrices obtained from the 'free categorization' task ranged from zero to ten. The entries in the other matrices ranged from zero to eight. An eight in the 'free categorization' matrix for example, scored on photo 6 and photo 36, means that eight of the ten subjects assigned this image to the same category. All eight joint matrices were analyzed using a clustering method and a multidimensional scaling (MDS) method.

Hierarchical Cluster Analysis

Cluster analysis is a commonly used statistical method for classification tasks. It is a method that attempts to group objects or individuals together that are generally more similar or have some specific similar characteristics.[8] In this study we applied a hierarchical nonmetric clustering method, named *complete linkage* or *furthest neighbor* technique. In this method, similarities are converted to distances. An observation is joined to a cluster if it has a certain level of similarity (small distance) with all members of that cluster. This procedure constructs a dendrogram or three diagram. Version 6.09 of the SAS computer program was used for the complete linkage hierarchical cluster analysis.

Figure 1. The 37 Kodak photographs used in this study.

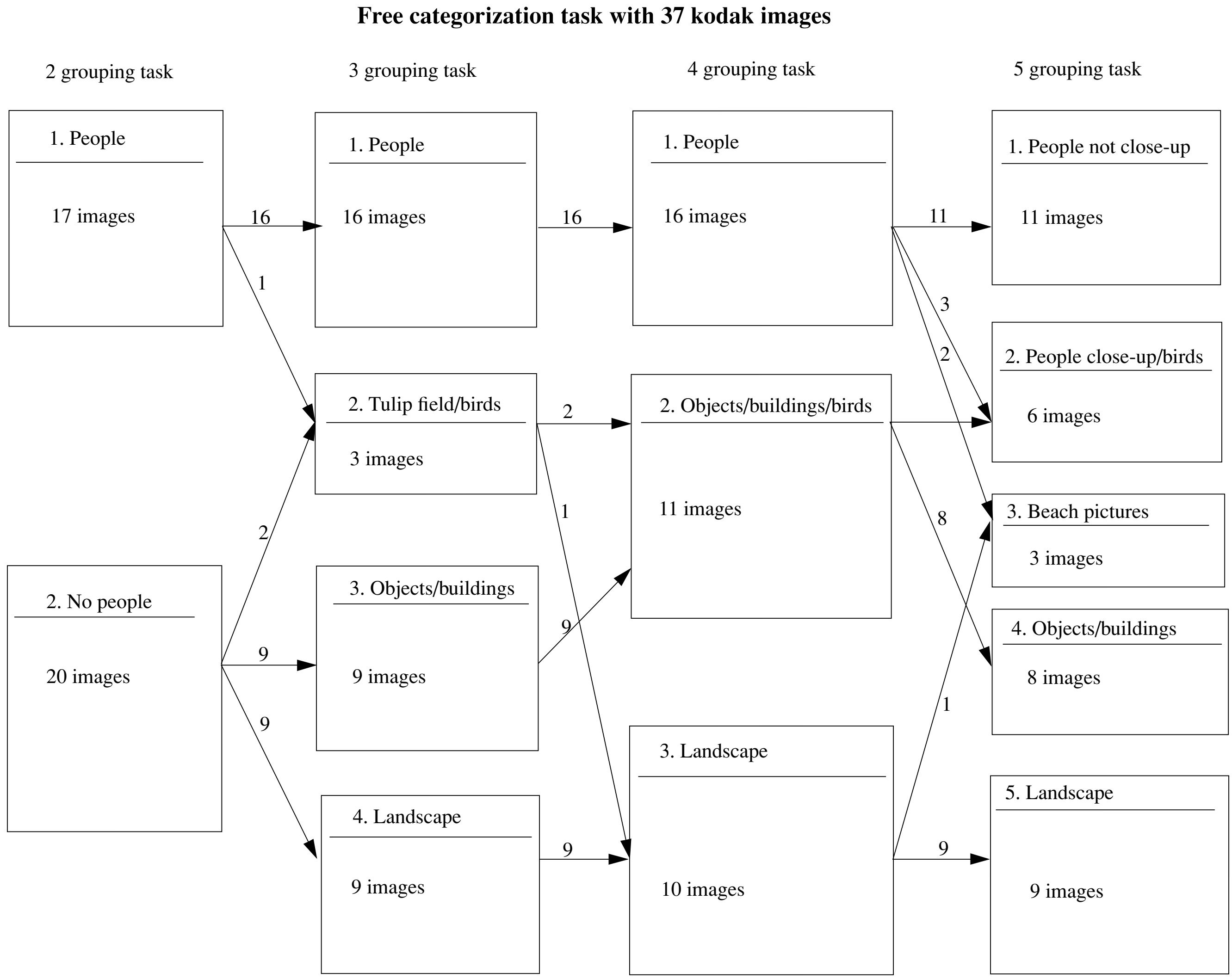

Figure 2. Schematic representation of the cluster analysis results for the 'free categorization' task.

Free Categorization Task. We applied complete linkage clustering to all four joint similarity matrices obtained from the 'free categorization' task. A short reproduction of the results is shown in Fig. 2. The first column of squares in Fig. 2 present the categories provided by cluster analysis applied to the matrix obtained from the initial task of two groups. The second, third and the fourth column show the categories obtained from respectively the three, four and five grouping task.

Our subjects' first task was to create two groups. The results show two main clusters called 'People' and 'No people'. These categories are represented by the first column at the left side of Fig. 2. Most subjects created those two clusters, however, some subjects constructed clusters called 'Living creatures' and 'Dead things'. These two clusters only differ from the 'People', and 'No people' cluster by the two photos of a bird (see Fig. 1: image p18 and p23). An eagle is a 'living creature', but doesn't belong to the 'People' cluster.

After the two grouping task, subjects constructed three groups of images. Although they were asked to create three categories, hierarchical cluster analysis resulted into four main categories (see the second column in Fig. 2). Cluster 'People' created during the two grouping tasks, was basically kept intact. Cluster 'No people' was split up into one category named 'Landscape', and another category called 'Objects/buildings'. The remaining cluster 'Tulip field/birds' is less easy to describe. The image 'Tulip field' (see Fig. 1) shows a tulip field and a woman in the distance. This picture is sometimes added to the cluster containing photos of people, because of the woman. In contrary, other subjects assigned this picture to the cluster called 'Landscape' because of the tulip field. In a same way, the two images of birds are sometimes added to the 'Living creatures' and other times to the group named 'Landscape'.

Cluster analysis applied to the matrix resulted from the four grouping task (third column) showed three main categories. These clusters appeared to be almost the same as the clusters obtained from the three grouping task. Only cluster 'Tulip field/birds' provided by the three grouping task no longer exists.

The fourth column in Fig. 2 presents the five main clusters obtained from the five grouping task. Cluster 'People' is now split up in one cluster named 'People not Close-Up' and another cluster containing images of 'Close-Ups of People/birds'. Subjects didn't agree about the fifth group. They constructed groups named 'Boats', 'Vehicles', 'Birds', 'Sport', 'Sun and See' and 'Art'. Therefore, it's difficult to give a description of this cluster 'Beach pictures'). The large diversity of these groups may be an explanation for the results.

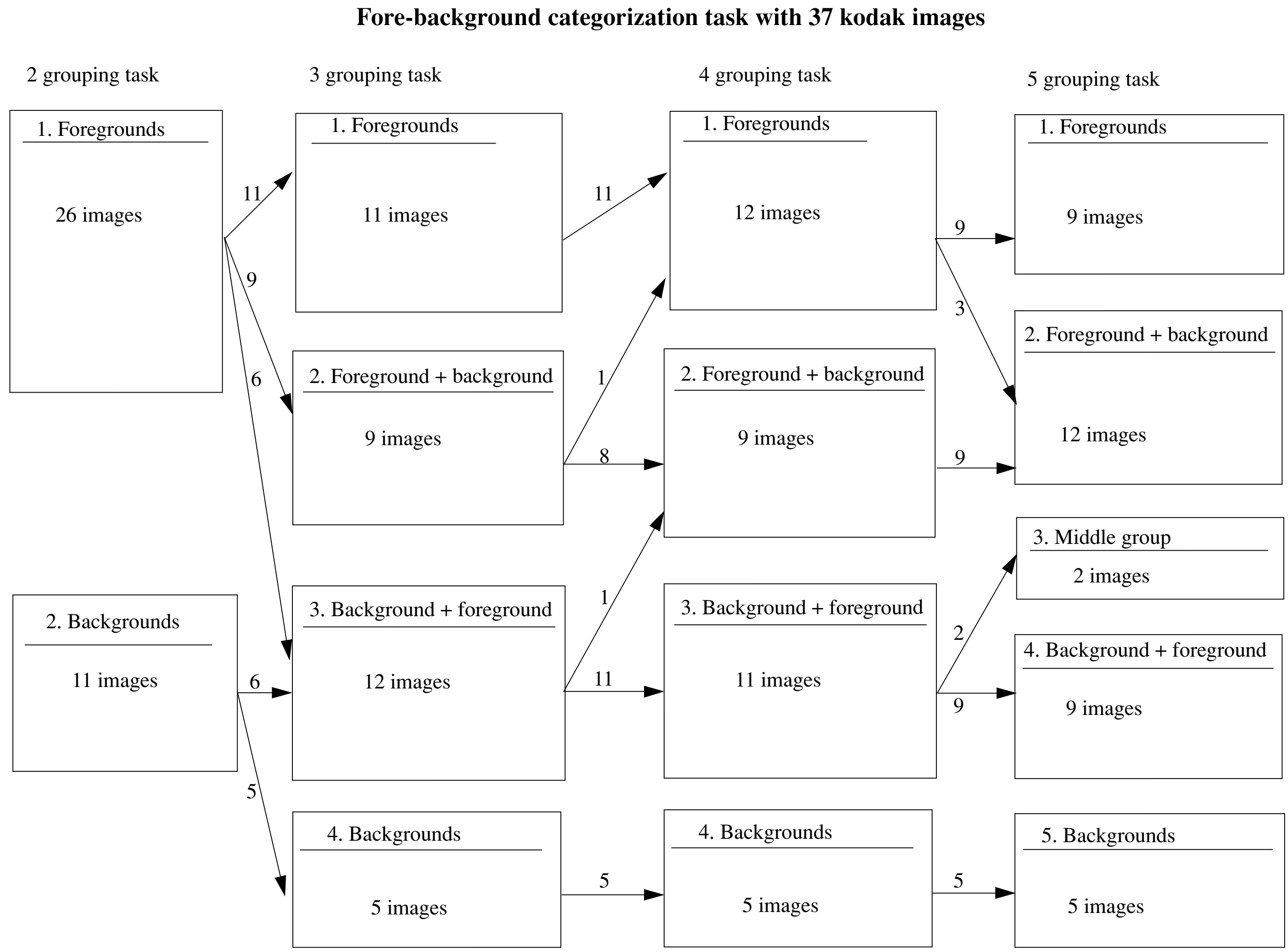

Figure 3. Schematic representation of the cluster analysis results for the foreground-background categorization task.

'Foreground–Background Categorization' Task. Just like the 'free categorization' task, we applied complete linkage clustering to all four joint similarity matrices provided by the 'foreground–background categorization' task. Figure 3 shows a brief presentation of the output from cluster analysis. The categories resulted from the two grouping task are presented by the first column at the left side of Fig. 3. The second, third and fourth columns show the clusters obtained from respectively the three, four and five grouping task.

First, subjects constructed one group of images containing 'foreground' pictures, such as a passport image, and a second group containing 'background' images, such as a photo of a mountain view. Cluster analysis was then applied to the matrix provided by this two grouping task and yielded two main clusters called 'Foreground' and 'Background' (see Fig. 3).

Next, subjects created not only a 'Foreground' and 'Background' group, but also a middle group showing both 'foreground' and 'background' information. For example image 'beach' (see image p11 of Fig. 1). This picture shows an obvious point of interest, namely two people. However, not only the 'foreground' (two people) the 'background' (the beach) also seems to be important.

Although we asked subjects to create three groups, the output from cluster analysis reveals four main groups. Figure 3 (second column) shows that the middle group is split up into two parts. Subjects didn't agree about the third group. They moved different pictures from the 'Foreground' or 'Background' cluster to the middle cluster. Hence, the resulting plot shows four main groups. Pictures in the 'Background+foreground' cluster show more background than foreground information, like image 'Beach people' described above. Almost all the images in 'Background+foreground' category were taken from the 'Background' group. Pictures belonging to the 'Foreground+background' cluster were all taken from the 'Foreground' cluster. The images belonging to this cluster show more foreground than background information.

Cluster analysis was also applied to the data provided by the four grouping task (see third column in Fig. 3). The clusters derived from the three groupings remained basically the same for the four grouping task. Only two images moved from one category to another. The image presenting a 'Painted house' moved from the 'Foreground-ground+background' group to the Foreground category. The image named 'Half timbered houses' was taken from the 'Background+foreground' cluster and moved to the 'Foreground+background' cluster.

The fourth column in Fig. 3 shows results of the five grouping task; 'Foreground', 'Foreground+background', 'Middle group', 'Background+foreground' and 'Background'. Figure 3 demonstrates that the 'Background' category remained identical for the three, four and five grouping assignment. The 'Middle group' consists of two

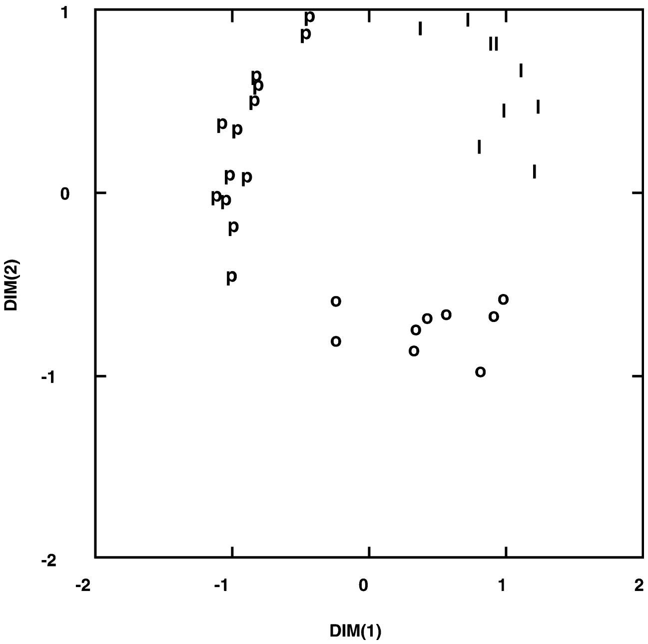

Figure 4. The two-dimensional MDS solution for four groupings in the 'free categorization' task. The resulted groups are: p = people, l = landscape, and o = objects/buildings.

images taken from the 'Background+foreground' cluster obtained from the four groupings. The image 'Boy in grass', 'Child+bridge' and 'Painted house' moved from the 'Foreground' to the 'Foreground+background' cluster. The rest of the 'Foreground+background' and 'Fore–ground' cluster remained the same.

Multidimensional Scaling

We took a second independent approach to analyze the data by using the multidimensional scaling method (MDS). MDS is based on an assumed relationship between the psychological concept of similarity or dissimilarity, and the mathematical concept of distance.[9] The goal of MDS is to represent a set of n stimuli geometrically by n points, so that the inter-point distances corresponds to the experimental similarities (dissimilarities) between points. A great benefit of the multidimensional scaling technique, is that it can help systematize data in areas where organizing concepts and underlying dimensions are not well developed. MDS can uncover the 'hidden structure' of a data set.[10] A disadvantage of the MDS method is that it gives no information about the meaning of the dimensions. Sometimes dimensions cannot be labeled, but a geometric representation can still be useful.

The monotonic MDS method was applied to both the four joint similarity matrices of the 'free categorization' task and the four joint matrices of the 'foreground–background categorization' task. Version 5.1 of the SYSTAT computer program was used for multidimensional scaling. In all cases, MDS plotted a Shepard diagram in which the distances between points are plotted against the observed similarities. The functions in this plot are smooth (without large steps). This should be considered good in monotonic scaling and suggests no degenerated solution.[10]

'Free Categorization' Task. We applied the MDS method to the four joint similarity matrices that resulted from the 'free categorization' task. For each of the four grouping assignments we found solutions for the one, two, three, four and five dimensions. In order to decide how many dimensions we should accept, we plotted the obtained stress values against the different dimensionalities for each grouping. Based on this result we examined the two dimensional configuration.

Subjects first created two groups. The stress value for this grouping task and the two dimensional solution[11] is 0.044. The results show two obvious distinguishable groups named 'People' and 'No people'. The same two groups were also obtained from the hierarchical cluster analysis. Next, subjects divided the 37 images into three clusters. The stress value for this solution is 0.10. Here again the MDS solution confirms the clusters identified by cluster analysis called 'People', 'Landscape' and 'Objects/buildings'.

The obtained stress value for the two dimensional solution for the four grouping task is 0.11. The resulting plot is presented in Fig. 4. We will only show this plot because this solution provides the best results concerned to a sufficiently low stress value and well distinguishable obtained clusters. In the figure, the groups provided by cluster analysis are presented with different letters.

The output from cluster analysis revealed identical categories. Just like the three grouping assignments, most (eight of the ten) subjects created again three clusters called 'Landscape', 'Objects/buildings' and 'People' (see Fig. 4). To construct a fourth category, some subjects split up the last described cluster into one group showing portraits of people/birds, and a second group presenting people but not close-ups. Other subjects created a fourth cluster named 'Vehicles'. We compared the results of the three grouping task with the four grouping task, and saw that the clusters obtained by the last task were a bit more separated from each other.

The obtained stress value for the two dimensional solution and the five grouping task is 0.11. The same five categories obtained from cluster analysis can be distinguished in here. Compared to results of the four grouping task, the distance between the elements within each cluster is larger. The clusters are less separated from each other.

The lowest stress value is obtained for the two grouping task and the three dimensional solution. The best results were not provided by this solution in our opinion. It is not so difficult to create two groups and find an acceptable two- or even three-dimensional solution. A sufficiently low stress value is obtained for the four grouping assignment and two dimensions.

'Foreground–Background Categorization' Task. We applied the MDS method to the four joint similarity matrices resulted from the 'foreground–background categorization' task. For all groupings, the two-dimensional solution provided the optimal results. We therefore examined the two-dimensional configuration. The results obtained from MDS verify the clusters obtained from cluster analysis for all grouping tasks.

Subjects started to construct one group of images containing foreground pictures and a second group containing background pictures. For the two-dimensional MDS solution the stress value is 0.038.

Figure 5 presents the plot resulted from the two-dimensional solution provided by the three grouping assignment. The obtained stress value for this solution is 0.047. The letters in the plot are based on the results obtained by complete linkage cluster analysis.

In the next step subjects created four groups (stress value is 0.06). We compared the results of the three

TABLE I. The Two-Dimensional Frequency Matrix of the 37 Kodak Pictures, Based on the Results of the Four Grouping and Three Grouping Procedure for the 'Free' and 'Foreground–Background Categorization' Task, Respectively

Free task/Foreground–background task	Foreground	Foreground–background	Background–foreground	Background	Total
People	7	4	5	0	16
Objects/buildings	4	5	2	0	11
Landscape	0	0	5	5	10
Total	11	9	12	5	37

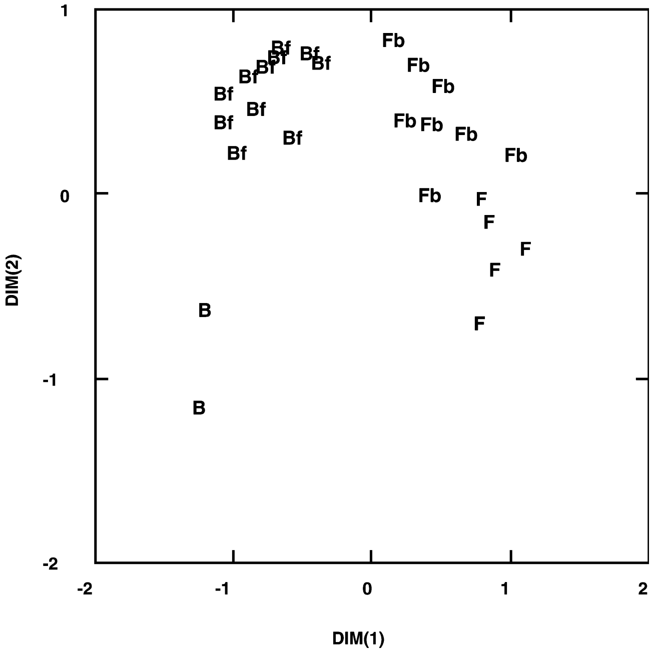

Figure 5. The two-dimensional MDS solution for three groupings in the foreground-background categorization task. The resulted groups are F= foreground, Fb= foreground with a bit background, Bf= background with a bit foreground and B= background.

grouping task with the four grouping task, and saw a large agreement. Only three images moved from one category to another. Although the obtained groups appear to be almost the same, the clusters created by each individual subject differ more in comparison with the three grouping task. The agreement among subjects was less for the four than for the three grouping assignment. The stress value increased from 0.047 (three groups) to 0.06 (four groups). This value decreased slightly (to 0.058) when students were asked to construct five groups. Comparing the results of the five with the four grouping task showed us that in the first task the distance between images presented in the plots within each cluster is larger due to less agreement among subjects for the five grouping task. Examining the categorization process from two through five groupings, seems as if there can be distinguished a horse shoe shape (see Fig. 5). Kruskal and Wish[10] describe this phenomenon as a nearly one-dimensional configuration bent around a horse shoe shape. This means that only one curvilinear dimension is sufficient to give a reasonable description of the data.

In our opinion, the optimum situation is obtained for the two-dimensional solution and the three grouping task. Subjects seem to be more homogeneous in making three rather than four groups. The lowest stress value is obtained for the two groupings, but this may be a trivial solution. It is relatively easy for subjects to divide the total set of images into two groups.

'Free Categorization' Task Compared to the 'Foreground–Background Categorization' Task. We compared the 'free' to 'foreground–background sorting' task. It seems as if subjects behave more homogeneous while serving the 'foreground–background categorization' task than in serving the other task. This probably reflects the fact that the first task is more constrained. For all the grouping tasks, the obtained stress values are lower for the 'foreground–background sorting' assignment than for the 'free sorting' task.

The clusters obtained from the two grouping task are mainly the same for both the 'free' and 'foreground–background' assignment. Comparing the clusters resulted from the three grouping task shows that half of the images appeared to be assigned to the same groups. The agreement between the four cluster task was a bit less than for the three groupings. For the five groupings only the extreme groups, like landscapes and portraits appeared to be the same. Pictures of landscapes nearly always present only background information. A picture of a portrait demonstrates mainly foreground information. Pictures of objects or houses in contrast can show only fore or background information or both types of information.

These results suggest an interesting correlation between the 'free' and 'foreground–background categorization' task. We therefore produced a two-dimensional frequency matrix of the 37 Kodak images. This matrix was based on the results of the three and four grouping assignment for respectively the 'foreground–background' and the 'free categorization' task (see Table I). As a measure of association we calculated the Goodman lambda coefficient.[12] Based on the data in Table I this coefficient is 0.24. This indicates that there are moderate relations between clusters created in the 'free categorization' task and cluster constructed in the 'foreground–background categorization' task. To determine if this relation is significant we calculated the log likelihood ratio (G^2).[13] Here $G^2 = 28.21$ (df = 6; $p < 0.001$), this means that the relation is significant.

Discussion

One of the goals of this study was to investigate whether subjects are able to classify images with respect to their perceptual significance. The results show that subjects are able to categorize images in a consistent manner when they are free to decide the similarity criteria, and that the similarity criteria are based on the material information content of the images. The results also demonstrate that the more groups subjects are asked to create, the less agreement there exists among them. It is also specified what the most important categories of image content are.

The subjects first created two clusters; one cluster contained images of people and the second cluster consisted of images not showing people. This means that the appearance of people in the images is the most important image feature in the opinion of the subjects. After the first grouping task, subjects maintained the ‘People’ cluster and divided the ‘No people’ cluster into a group called ‘Landscape’ and a group called ‘Objects/buildings’. In our opinion, subjects filtered the ‘Landscape’ images out of the ‘No people’ cluster and attempted to name the remaining images. Therefore, the second relevant image feature is whether an image demonstrates a landscape or not.

In the next sorting assignment half of subjects split up the ‘People’ cluster into two parts; one containing portraits, the other showing people or living creatures but not portraits. Hence, a third but less convincing image feature may be whether an image shows a portrait or not.

In a pilot test, we found quite similar results. The study was based on 91 images including the 37 Kodak Photo-CD images. In contrast with the experiment described in this article, subjects were instructed to create as many categories as they wished. Cluster analysis and MDS revealed categories called ‘Landscape’, ‘People’, ‘Objects’ and ‘Buildings’ also obtained by the experiment in this article. Additionally, subjects made clusters named ‘Child and toys’, ‘Animals’, ‘Vehicles’, ‘Sport’ and ‘City life’. These lead to the question of how generalized the results of this categorization are, or how strongly the results depend on the stimulus set. From comparison of the different experimental results it can be tentatively concluded that if the stimulus set is sufficiently diverse and large, the category set will at least contain a grouping in terms of ‘Landscape’, ‘People’, ‘Objects/Buildings’.

The results of the foreground–background experiment show that subjects are able to split up the set of 37 images into 2 groups in a consistent way, but the more clusters they constructed the less the agreement among them. Subjects make similar groups of images providing only foreground or background information is present. The classes in between these two extremities differ from person to person.

If we compare the ‘free categorization’ task with the ‘foreground–background categorization’ task, we see a correlation that is promising from a practical point of view. Pictures of landscapes nearly always present only background information. A picture of a portrait demonstrates almost only foreground information. Pictures of objects or houses, however, can show a combination of foreground and background information or just one of these types of information.

References

1. A. Maeder, J. Diederich and E. Niebur, Limiting human perception for images sequences, *Proc. SPIE* **2567,** 330–337 (1996).
2. B. Nystrom and M. Fairchild, Perceived image quality of 16:9 and 4:3 aspect ratio video displays. *J. Electronic Imaging* **1(1),** 99–103 (1992).
3. W. Zou, K. Ellsworth, J. Kutzner, and P. Hearty, Subjective testing of broadcast-quality compressed video, *SMPTE J.* 789–794 (1994).
4. G. Deffner, M. Yuasa, M. McKeon, and D. Arndt, Evaluation of display image quality: eperts vs non-experts, *SID94Digest* 475–478 (1994).
5. A. Rao and G. Lohse, Identifying high level features of texture perception, *GVGIP: Graphical Models and Image Processing* **55 (3),** 218–233 (1993).
6. *Kodak Photo-CD Photo Sampler*. Distributed by Apple Computer for Eastman Kodak Company (1991).
7. B. Everitt, *Cluster analysis*, Edward Arnold, London (1993).
8. P. Dunn-Rankin, *Scaling methods*, Lawrence Erlbaum Assc., Hillsday, New Jersey (1983).
9. J. Ramsay *Multiscale manual*, Internal report (1991).
10. J. Kruskal and M. Wish, *Multidimensional scaling,* Sage Publications, Beverly Hills, 1983.
11. J. Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrical*. **29,** 1–27 (1964).
12. D. Hinkle, W. Wiersma and S. Jurs, *Applied statistics for the behavioral sciences*, Houghton Mifflin Company, Boston (1994).
13. W. Hays, *Statistics*, Holt, Orlando (1988).