Mandatory scale perception promotes flexible scene categorizations

Aude Oliva
Laboratoire de Traitement d'Images et Reconnaissance de Formes
Institut National Polytechnique de Grenoble
46, Avenue Felix Viallet
38031 Grenoble, France
oliva@tirf.grenet.fr

Philippe G. Schyns
Dept. of Psychology
Montreal University
C.P. 6128, Succursale A
Montreal (Quebec) H3C-3J7
schynsp@psy.umontreal.ca

Abstract

Efficient categorizations of complex stimuli require effective encodings of their distinctive properties. In the object recognition literature, scene categorization is often pictured as the ultimate result of a progressive reconstruction of the input scene from precise local measurements such as boundary edges. However, even complex recognition tasks do not systematically require a complete reconstruction of the input from detailed measurements. It is well established that perception filters the input at multiple spatial scales, each of which could serve as a basis of stimulus encoding. When categorization operates in a space defined with multiple scales, the requirement of finding diagnostic information could change the scale of stimulus encoding. In Schyns and Oliva (1994), we showed that very fast categorizations encoded coarse information before fine information. This paper investigates the influence of categorization on stimulus encodings at different spatial scales. The first experiment tested whether the expectation of finding diagnostic information at a particular scale influenced the selection of this scale for preferred encoding of the input. The second experiment investigated whether the multiple scales of a scene were processed independently, or whether they cooperated (perceptually or categorically) in the recognition of the scene. Results suggest that even though scale perception is mandatory, the scale of stimulus encoding is flexibly adjusted to categorization requirements.

Introduction

Efficient categorizations of complex visual stimuli require effective encodings of their distinctive properties. In the object recognition literature, scene categorization is often pictured as the ultimate result of a progressive reconstruction of the input scene from simple measurements (e.g., Marr, 1982). Boundary edges, surface markers and other low-level visual cues are progressively integrated into successive layers of representations of increasing complexity, the last of which derives the identity of a scene from the identity of a few objects. For example, in the central picture of Figure 1, combinations of fine-grained edge descriptors and other local cues suggest the presence of cars, road panels, highway lamps and other objects which typically compose a highway scene. Precise categorization of a scene often requires that the local identification of component objects precedes the identification of the scene.

However, complex visual displays composed of many partially hidden objects are often recognized quickly, in a single glance--in fact, as fast as a single component object (Biederman, Mezzanotte, & Rabinowitz, 1982; Potter, 1976; Schyns & Oliva, 1994). This data suggests that there could be more direct routes to scene categorization than "object-before-scene." Categorization processes could sometimes directly extract global representations of the input scene; representations allowing "express," but comparatively less precise classifications of the input. To illustrate the different routes of scene categorization, squint or blink while looking at the central picture of Figure 1, another scene should appear (if this demonstration does not work, step back from the picture until you perceive a city).

Figure 1 simultaneously presents visual cognition with two scenes associated with a different spatial scale. Although it is possible to identify the background city scene from the spatial layout of its major "blobby" components, it is virtually impossible to identify each isolated blob as a building (a blob can potentially correspond to many objects). This example demonstrates that object-before-scene is not a mandatory route to successful scene categorizations. Scene-before-object can also characterize the recognition of complex pictures. In any case, scale-specific information can be used independently to achieve distinct categorizations of the same hybrid stimulus.

Recent studies on the relationships between categorization and perception have revealed that the high-level task being accomplished influences the low-level encodings of the stimuli (e.g., Goldstone, 1994; Schyns & Murphy, 1994). The availability of multiple levels of representation of the same scene could promote selective encodings of the scene at the scale best suited for the task considered. For example, while precise categorizations could reconstruct the input from local fine-grained measurements (e.g., boundary edges), express categorization processes could encode the same stimulus at a cruder resolution highlighting the global scene structure. In Schyns and Oliva (1994), we reported such a coarse-to-fine encoding of hybrid stimuli. However, in agreement with the demonstration of Figure 1, control stimuli also revealed that the two spatial scales were available at the onset of processing. Thus, the reported coarse-to-fine could result from task constraints rather than from a mandatory order of spatial scale perception.

In this paper, we report two experiments studying how categorization processes operate in the space of spatial scales made available by perception. The first experiment tested

(n-1)

(n)

(n-1)

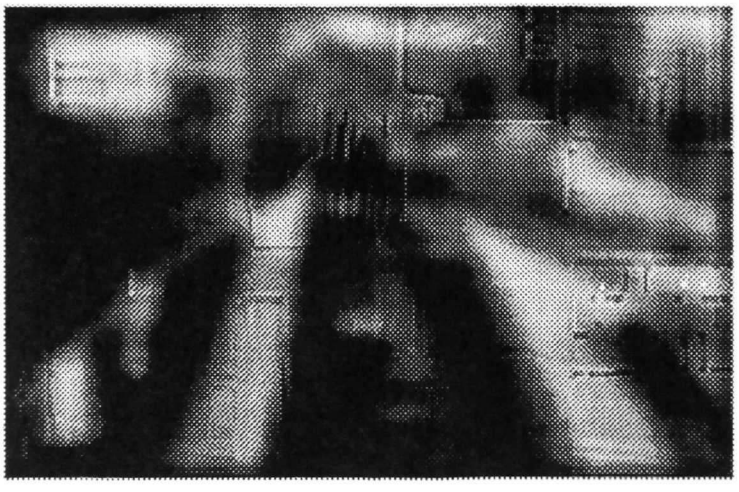

Figure 1: This figure illustrates the hybrid stimuli used in our experiments. n-1 and n+1 are LF/Noise hybids. n is an ambiguous hybrid. The succession of these three hybrids illustrate the gist of the cross-frequency priming in Experiment 2. The HF component of n always was of the same category as the LF component of n+1. The triple shown illustrates the perceptual condition. In the categorical condition (not shown on this figure), the HF component of n was a different scene of the highway category.

whether the expectation of finding diagnostic information at a particular scale influenced the selection of this scale for preferred encoding of the stimuli. The second experiment investigated whether the coarse and fine spatial components were processed independently, or whether they cooperated (perceptually or categorically) in the recognition of the input.

Experiment 1

To demonstrate that categorization influences the scale of stimulus encoding, we ran a simple experiment in which subjects were asked to categorize 18 hybrid stimuli (see the n picture of Figure 1). Hybrids are inherently ambiguous and so we should expect equivalent proportions of categorizations based on coarse and fine information. But if categorization processes expect diagnostic information to reside at only one scale, then input encoding could preferentially operate at this scale and influence the categorization of the hybrid. To induce such expectation through categorization, we initially exposed two groups of subjects to hybrids that were meaningful at only one scale, before presenting the groups with the same set of ambiguous hybrids.

The Low-Frequency (LF) (vs. High-Frequency, HF) group was initially sensitized to 6 hybrids whose HF (vs. LF) component was structured noise. We expected that these stimuli would sensitize categorization to the scale components containing diagnostic information. Without subjects being aware, the two scale components of the last 12 hybrids were both diagnostic. We expected mutually exclusive categorizations of these stimuli, without subjects being aware of the other meaningful scene. This result would provide evidence that the high-level constraint of finding diagnostic information for categorization induces scale-specific encodings, without an obligatory processing of one scale before the other.

Methods

Subjects. Twenty-four adult subjects with normal or corrected vision volunteered their time to participate to the experiment. They were randomly assigned to the LF (vs. HF) group with the constraint that the number of subjects be equal in each group.

Stimuli. Three types of hybrid stimuli were constructed (LF/Noise, HF/Noise and ambiguous) from different pictures of four categories (*city*, *highway*, *living-room* and *bed-room*). We synthesized a total of 6 LF/Noise (vs. 6 HF/Noise) sensitization stimuli by combining the LF (vs. HF) components of two distinct pictures of the categories with HF (vs. LF) structured noise (see the n - 1 and the n + 1 pictures of Figure 1). Test stimuli were ambiguous hybrids, computed as explained earlier by combining the LF and HF components of two different scenes. We synthesized a total of 24 hybrids by systematically combining 2 pictures of 4 distinct categories with the constraint that the two scenes of a hybrid were of a different category. Hybrids subtended 6.4 x 3.4 deg of visual angle on the monitor of an Apple Macintosh. (See Schyns & oliva, 1994, for a detailed description of the computation of hybrids).

Subjects did not directly experience these hybrids. Instead, they saw one animation per hybrid stimulus. Each hybrid (sensitization and test) was presented in a brief animation composed of three successive frames--at a rate of 45 ms per frame, to ensure that they fuse on the retina. The first, second and third frames presented the hybrid with low- and high-frequency Butterworth cut-off points set at 2 and 6, 3 and 5, 4 and 4 cycles/deg of visual angle, respectively. Although subjects saw brief animations, for ease of presentation, we will refer to the animations as hybrids in the remaining of the text.

Procedure. Sensitization Phase. LF subjects were exposed to 6 LF/Noise, and the HF group saw 6 HF/Noise. In a trial, subjects would see one hybrid for 135 ms on a CRT monitor. Order of trials were randomized with a 1.5 sec interval between trials. Subjects' task was to categorize the hybrid. As there was only one meaningful scene in LF/Noise and HF/Noise stimuli, subjects could only succeed by attending to the diagnostic scale (LF or HF).

Testing Phase. Testing stimuli were presented immediately after the sensitization stimuli, without discontinuity in their presentation. There are two ways to synthesize a hybrid from two scenes, depending on which picture is assigned to the LF (or HF) component. Half of the subjects of each group saw one version of each hybrid, and the other half saw the other version. For example the first half saw LF city1/HF highway1 (block A) and the other half saw LF highway1/HF city1 (block B). There were 12 hybrids in each block. This strategy ensures a balanced design, without repetition of trials. Note that the pictures used for sensitization were not used for testing. The 12 hybrids of the testing phase were each presented as explained above, and the entire experiment lasted for about 2 minutes. We recorded the number of LF (vs.) HF categorizations of the 12 ambiguous hybrids in each condition.

Debriefing. After the experiment, we asked subjects several questions about the experiment. One of these questions was particularly important for the interpretation of the results. Subjects were shown a hybrid stimulus composed of two meaningful scenes and were asked the following question: "Here is a stimulus composed of two scenes. Did you explicitly notice, or did you have the impression that there were such stimuli during the experiment?"

Results and Discussion

To ensure that the blocks of test hybrids did not influence performance, we first ran an ANOVA taking the LF- vs. HF-group, block A vs. B and LF vs. HF categorizations as factors. As neither the block factor nor the interactions with LF vs. HF categorizations were significant, we collapsed the two blocks in each group. Subjects sensitized to the LF scale categorized 73% of ambiguous hybrids according to their LF component, while HF subjects categorized 72% of the same stimuli on the basis of their HF information. A two way ANOVA revealed a significant interaction between sensitization (LF vs. HF) and categorizations (LF vs. HF), $F(1, 22) = 43.69$, $p < .0001$.

The data reveal mutually exclusive categorizations of identical stimuli. There are at least two possible accounts of

the opposite categorizations. Subjects could notice that there were two meaningful scenes in the 12 hybrids, but strategically decide to report only the scale information congruent with their sensitization phase. Another, perhaps more interesting interpretation would propose that the sensitization phase influenced the way stimuli were encoded for categorization. That is, although low-level perception would register both spatial scales, stimulus encoding and categorization processes would only operate at the scale imposed by the initial categorization constraints.

In the debriefing phase, one of the questions specifically asked subjects whether they noticed that two meaningful scenes composed a large number of the stimuli. All subjects (but one) reported seeing only one scene that was perceived as a noisy picture--as if the scene was observed through a dirty window. Subjects were surprised to learn that two-third of the hybrids were composed of two scenes.

Together, these results suggest that the selection of a spatial scale for categorization can be determined by the information content of that scale. It is doubtful that categorizations could be maintained at a single scale (when both scales were meaningful) if the selection of spatial scales for higher-level processing was mandatorily fixed by low-level processes. If the groups reliably categorized the same stimuli as different scenes, it seems likely that their stimulus encoding and categorization processes were driven by the constraint of finding diagnostic information.

Experiment 2

Experiment 1 provided evidence that high-level processes can actively "select" the spatial scale of stimulus encoding, and that subjects were not aware of the information at the other scale. Awareness, however, should not be equated with processing. Subjects could very well be unaware of the other scale, but implicit processing at this scale could influence explicit processing at the relevant scale. This influence could simply be perceptual, revealing a cooperation of the spatial scales in the low-level analysis of the input, or the influence could be categorical, suggesting that the two spatial scales of a hybrid could simultaneously activate high-level representations.

Experiment 2 was designed to address the issue of perceptual or categorical influences across spatial resolutions. Three groups of subjects were asked to categorize a series of hybrids. Most hybrids of the series were LF/Noise, so we expected categorization to operate mostly at this scale, as shown in Experiment 1. Once every 4 LF/Noise, on trial n, we introduced an ambiguous hybrid whose HF component was the same scene as the LF component of the next hybrid (see Figure 1). We hypothesized that although explicit categorizations were accomplished only at the LF scale, an implicit processing of HF information could influence explicit categorizations. In the *perceptual* group, the prime and the target were different scale representations of exactly the same scene. In the *categorical* group, the prime and the target were different scale representations of distinct pictures of the same scene category (e.g., two different highways). In the *control* group, all n stimuli were replaced by LF/Noise stimuli. If all spatial resolutions are mandatorily perceived, we should only observe a positive priming in the perceptual condition. If all spatial resolutions are perceived and encoded for categorization, priming should occur in two experimental conditions.

Methods

Subjects. Subjects were 44 Grenoble University students who were paid to participate to the experiment. Only 36 subjects (12 per group) we used for the analysis (see below).

Stimuli. Hybrids were the 24 ambiguous stimuli of Experiment 1 and sensitization stimuli were the 8 LF/Noise of 8 scenes (2 pictures of 4 categories). As in Experiment 1, three-frame animations of the hybrids were presented (at a rate of 45 ms per frame).

Procedure. In an initial sensitization phase, subjects categorized two times the 8 LF/Noise, to ensure that they would categorize the scenes consistently fast. In contrast to Experiment 1, LF/Noise were interleaved with ambiguous hybrids throughout this experiment, as explained below.

The priming situation used triples of hybrid stimuli (see Figure 1). Hybrid n 1 and n had identical LF components. This procedure was meant to prime a LF categorization of n (to reduce chances of HF categorizations). Hybrid n was ambiguous for the perceptual and the categorical groups. In the perceptual group (illustrated on Figure 1), the HF component of n and the LF component of n + 1 were different scale representations of the same scene. In the categorical group, these two components were different scenes of the same category. This situation allowed the testing of a cross-resolution priming (perceptual and categorical) of the HF of n on the LF of n + 1. In the control group, there was no correspondence across resolutions between n and n + 1 (i.e., n was a LF/Noise). Each of the 24 hybrids served as n stimulus in composing 24 triples, using the appropriate n 1 and n + 1 hybrids.

Triples describe the organization of Related (R) trials. Triples were separated from one another with one LF/Noise stimulus. These 24 separators were used as fillers to keep categorization at the LF scale. The n - 1 stimulus of a triple served to compute UnRelated (UR) trials. UR trials were always preceded by a LF/Noise stimulus--i.e., there was no scene correspondence across resolutions. The entire experiment was composed of a total of 96 trials (24 triples plus 24 LF/Noise). Note that the only difference across group is the nature of the HF component of a n hybrid. Subjects' task was to categorize stimuli (by naming them) as fast and as accurately as they possibly could. We recorded subjects' reaction times with a Lafayette vocal-key and also measured their categorization performances.

Debriefing. After the experiment, we asked subjects the same questions about the overall appearance of the stimuli as in Experiment 1.

Results and Discussion

As there were repetitions of trials in this experiment, the proportion of subjects who noticed two scenes in some hybrids grew accordingly (4 in 16 in the perceptual group and the same proportion in the categorical group). Their data were discarded from the analysis. In the remaining data, we also removed the triples in which the n stimulus was

categorized as HF, to ensure that priming was only measured after an explicit LF categorization of the ambiguous hybrid. On average, 2 triples (out of 24) were removed per subject. 91% of LF categorizations of the ambiguous hybrids indicate that categorization was reliably kept at only one spatial scale.

Cross-resolution priming rates were high (28 ms) between R and UR trials in the perceptual group, but non-existent in the categorical and control groups (0 and 1 ms, respectively). An ANOVA with groups (perceptual, categorical and control) and types of trials (R vs. UR) revealed a significant interaction $F(2, 33) = 3.77$, $p < .05$. A post-hoc test between R and UR trials of the perceptual group showed a significant effect of priming, $F(1, 33) = 11.52$, $p < .01$, but no such effect was observed for the control and the categorical groups.

The results provide further evidence that diagnostic information can bias explicit categorizations to the informative scale. But they also demonstrate that the uncategorized information is not lost. Instead, this information is implicitly registered and influences explicit categorizations, across resolutions. This influence only occurs when the prime and target are identical scenes represented at different scales. When the two spatial scales represented different exemplars of the same category, no priming was observed.

Together, these results suggest a mandatory processing of the complete scale space, even when diagnostic information is consistently associated with only one spatial scale. The constraint of finding relevant categorization information influences which spatial component is preferentially encoded and categorized. Implicit processing does not seem to go beyond the perceptual registration of the different scale components.

General Discussion

The aim of this paper was to investigate how the high-level constraint of categorizing complex scenes interacts with the materials made available by low-level scale perception. Results of Experiment 1 demonstrated that the scale for preferred processing was determined by the diagnostic information present at this scale. The second experiment showed that even when explicit categorizations were accomplished at the diagnostic scale, they were perceptually (but not categorically) influenced by implicit processing at the other scale. Together, these results indicate that scale processing mandatorily occurs at all spatial scales. The constraint of finding diagnostic information determines which aspect of the stimulus is encoded for categorization.

It is interesting to note that categorizations can be kept at a single level of resolution, even when diagnostic information is present at the other spatial scale. If object and scene recognition systematically resulted from a reconstruction of detailed and highly processed measurements of the input, one should expect a bias in favor of the HF categorization of a hybrid. The same bias should be observed if categorization was initiated after a low-level coarse-to-fine analysis of the scene--because categorization would preferentially operate on the detailed information. The fact that no *a priori* bias is observed for one particular scale when both scales are registered suggests a flexible usage of a mandatorily processed scale space.

In summary, our experiments suggest that scale perception constrains categorization to operate in a scale space, but a space sufficiently diversified to promote flexible (in the case of our experiments mutually exclusive) categorizations of the same input stimulus. Our experiments indicate that information, rather than lower-level processes, determines which aspect of the stimulus is encoded for categorization. We believe there is much to learn about the ways task constraints and perception participate to the encoding of complex visual stimuli for categorization.

References

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, **14**, 143-177.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, **123**, 178-200.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Potter, M. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, **2**, 509-522.

Schyns, P.G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science*, **5**, 195-200.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In Medin (Ed.). *The Psychology of Learning and Motivation*, **31**, 305-354. Academic Press: San Diego, CA.