VISUAL COGNITION, 2005, 12 (6), 938–953

Psychology Press
Taylor & Francis Group

The use of visual information in natural scenes

Maxine McCotter
Department of Psychology, University of Glasgow, UK

Frederic Gosselin
Dépt de Psychologié, Université de Montréal, Canada

Paul Sowden
Department of Psychology, University of Surrey, Guildford, UK

Philippe Schyns
Department of Psychology, University of Glasgow, UK

Despite the complexity and diversity of natural scenes, humans are very fast and accurate at identifying basic-level scene categories. In this paper we develop a new technique (based on *Bubbles*, Gosselin & Schyns, 2001a; Schyns, Bonnar, & Gosselin, 2002) to determine some of the information requirements of basic-level scene categorizations. Using 2400 scenes from an established scene database (Oliva & Torralba, 2001), the algorithm randomly samples the Fourier coefficients of the phase spectrum. Sampled Fourier coefficients retain their original phase while the phase of nonsampled coefficients is replaced with that of white noise. Observers categorized the stimuli into 8 basic-level categories. The location of the sampled Fourier coefficients leading to correct categorizations was recorded per trial. Statistical analyses revealed the major scales and orientations of the phase spectrum that observers used to distinguish scene categories.

Humans are remarkably fast at recognizing and classifying environmental scenes despite a large and varied number of component objects within a scene (Potter, 1975). Recent findings suggest that prior recognition of component objects is not essential for scene recognition, and that the overall gist of the scene may be more important (Henderson & Hollingworth, 1999; Oliva & Schyns, 1997, 2000; Oliva & Torralba, 2001; Sanocki & Epstein, 1997; Schyns

Please address all correspondence to: M. V. McCotter, Department of Psychology, University of Glasgow, 58 Hillhead St., Glasgow, G12 8QB, UK. Email: maxine@psy.gla.ac.uk

This research was supported by BBSRC Grants (nos. SP3185 and S1386) awarded to Philippe Schyns and Paul Sowden. The authors wish to thank Aude Oliva, for lending us the scene stimuli that were used in our experiment, and Bosco Tjan, for useful discussions concerning this research.

http://www.tandf.co.uk/journals/pp/13506285.html DOI:10.1080/13506280444000599

& Oliva, 1994), even though detection of component objects can be achieved in as little as 150 ms (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Thorpe, Fize, & Marlot, 1996) and in the near absence of attention (Li, VanRullen, Koch, & Perona, 2002).

Studies of the structure of scene categories in memory have identified three particularly useful levels of scene categorization (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Tversky & Hemenway, 1983): Superordinate (e.g., artificial/natural), basic (e.g., city/highway), and subordinate level (e.g., a particular example of a city). Gosselin and Schyns (2001b) proposed that basic-level categories are those that minimize the overlap of properties between categories (i.e., strategy length) and maximize the number of properties that are unique to this category (i.e., internal practicability), and therefore give rise to faster recognition (see Gosselin & Schyns, 2001b, for a discussion of the other properties associated with the basic-levelness of a category). Accordingly, the gist of one basic-level scene category should be the scene information that minimizes the overlap of properties with other categories, and maximizes the number of properties specific to this category.

There has so far been no systematic study of the structure of information responsible for basic-level scene categorizations. Part of the problem arises from the complexity and diversity of these stimuli, making it difficult to assess common information use. Here, we develop a new technique (based on *Bubbles*, Gosselin & Schyns, 2001a; Schyns, Bonnar, & Gosselin, 2002) to determine some of the information requirements of basic-level scene categorizations.

INFORMATION FOR SCENE CATEGORIZATION

Before addressing the issue of the information human use for basic-level categorizations, we must address the issue of the information that is available in the image statistics to perform the task. Analysis of a scene in the Fourier domain results in Fourier coefficients, representing the energy and phase relationships of each frequency in the image (Campbell & Robson, 1968; de Valois & de Valois, 1988; see Figures 1b and 1c for example of a Fourier transform). The energy in a Fourier coefficient is the contrast energy of this frequency in the image. Numerous psychophysical studies have shown that the human visual system is selectively sensitive to limited bands of spatial frequencies and to the orientation bandwidth of image components (Campbell & Kulikowski, 1966; Campbell & Robson, 1968; de Valois & de Valois, 1988). In addition to these physiological restrictions on information content, the spectral distribution of natural scenes also imposes constraints on the available information. Natural scenes are known to have a characteristic energy spectrum, with a linear decrease of energy with increasing spatial frequency (Field, 1987; Parraga, Troscianko, & Tolhurst, 1999). The distribution of energy in natural scenes has been shown to characterize their structure, albeit as a first approximation (e.g., Guérin-Dugué &

Oliva, 2000; Oliva & Torralba, 2001; Schwartz & Simoncelli, 2001; Schyns & Oliva, 1994; Simoncelli, 2003; Simoncelli & Olshausen, 2001; Switkes, Mayer, & Sloan, 1978; Tadmor & Tolhurst, 1993; Torralba & Oliva, 2003). For example, the 2-D image of a city has dense vertical and horizontal organization and the occurrence of the horizon line in coastal scenes produces a bias towards horizontal organization. The reasoning is that if these represent distinctive properties of scene categories, then they should be represented with higher energies in their amplitude spectra. Oliva and Torralba (2001) examined the amplitude spectra of artificial and natural scenes to formulate the ''spatial envelopes'' of scene categories. The average slope and dominant orientations of the amplitude spectra corresponded to degrees of scene ''openness'', ''expansion'', ''roughness'', and ''ruggedness''. These characteristic amplitude spectra were compared with basic-level scene categories. For example, mountain scenes scored highly on the ''ruggedness'' parameter, whereas coastal scenes and landscapes scored highly on the ''openness'' parameter. However, this approximation is valid for all visual stimuli compatible with the amplitude spectra of these scenes, and visually meaningful scenes are only a small subset of this set. Most other stimuli are simply noise (see Figure 1e).

This arises because phase relationships describe how spatial frequencies of varying energy linearly contribute to represent the structures of the image (the blobs, contours, and edges; Morrone & Burr, 1988; Oppenheim & Lim, 1981; Piotrowski & Campbell, 1983; see Figure 1d and 1e). The importance of phase is illustrated by the effect of disrupting the phase of spatial frequencies, which renders a scene unrecognizable (compare original image in Figure 1a to image in Figure 1e; see also Sadr & Sinha, 2004; Schwartz, Tjan, & Chung, 2003; Sekuler & Bennett, 2001; Thomson & Foster, 1997; Thomson, Foster, & Summers, 2000). While amplitude spectra vary from scene to scene (Tadmor & Tolhurst, 1993, 1994), the statistics captured by phase information contain the majority of the visual information used to discriminate scenes. Natural image statistics differ primarily from each other in terms of the higher order correlations that structure their phase spectra (Thomson, 1999; Thomson & Foster, 1997; Thomson et al., 2000), which allow sparse linear coding of higher order image statistics across spatial frequencies (Field, 1994; Morrone & Burr, 1988; Olshausen & Field, 1996; Sekuler & Bennett, 2001; Simoncelli & Olshausen, 2001). Consequently, everyday scene categorizations must use the information represented in the phase spectra.

To determine the phase information required for basic-level scene categorizations, we used a technique of selective alteration of phase in the Fourier coefficients while normalizing (whitening) the amplitude of each component frequency (see Figure 2c; Simoncelli & Olshausen, 2001; Tadmor & Tolhurst, 1994; Thomson, 1999). The algorithm randomly samples the Fourier coefficients of the phase spectrum (see Figure 2b). Sampled Fourier coefficients retain their original phase while the phase of nonsampled coefficients is

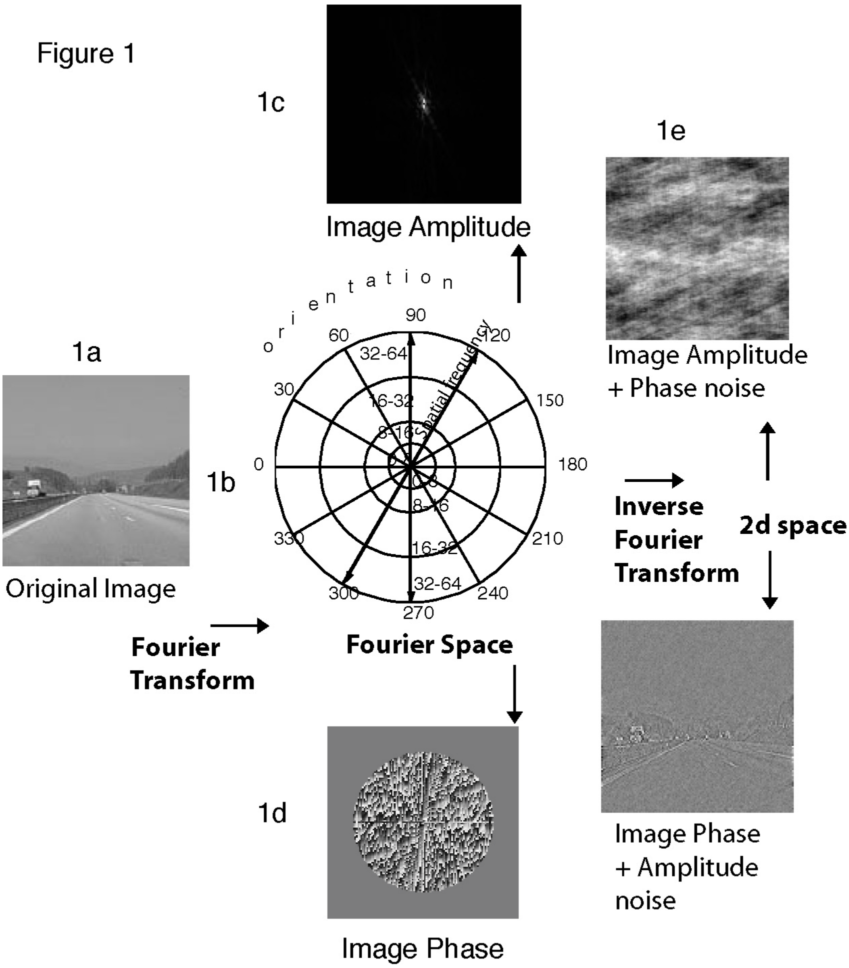

Figure 1. (a) The original image in 2-D space. (b) Representation of Fourier space with spatial frequency bandwidths of 0–8, 8–16, 16–32, and 32–64 cycles per image, and orientations from 0 to 359°. Orientations from 180 to 359° are a mirror symmetric sample of the phase and amplitude components at 0–179°. (c) Representation of image amplitude component in Fourier space. (d) Representation of image phase component in Fourier space. (e) Reconstruction of image amplitude with the phase component scrambled (above), and image phase with the amplitude component scrambled (below) in 2-D space.

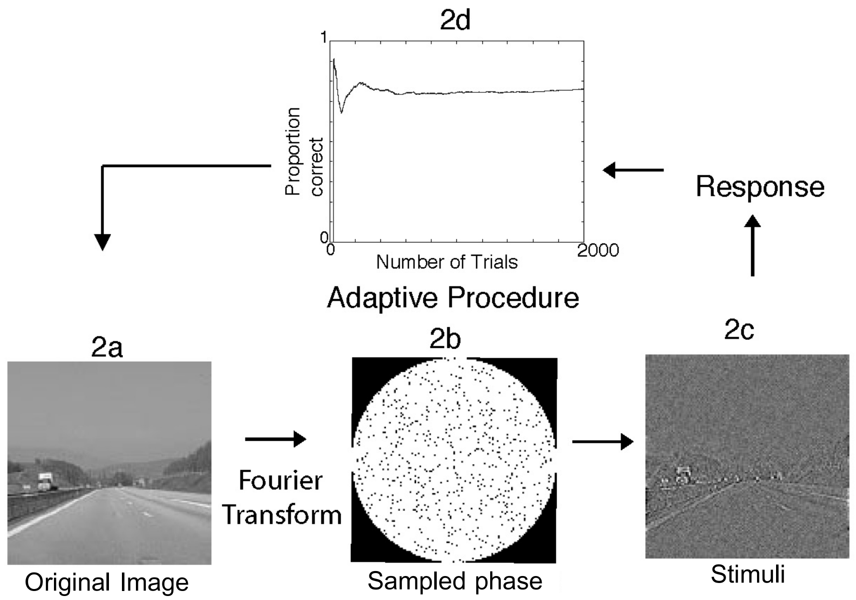

Figure 2. (a) Original image. (b) Random sampling of phase component. (c) Stimuli consisting of sampled phase with remaining phase scrambled and image amplitude replaced by white noise. (d) Adaptive procedure, which determines density of phase sampling per trial.

replaced with that of white noise. An inverse Fourier transform reconstructs the experimental stimulus of that trial (see Figures 1e, 2c, & Figure 3 for examples). To maintain categorization performance below ceiling (at 75% correct), the ratio of image phase to noise phase is adjusted online, on a trial per trial basis, independently for each category. For each category, the Fourier coefficients leading to correct categorizations are registered independently of those leading to incorrect categorizations. To ensure that the experimental task does not trivialize the complexity of real-world scene recognition, we chose a wide a variety of scenes and scene categories from an established database (Oliva & Torralba, 2001).

METHOD

Subjects

Twenty-four male and female observers aged between 18 and 35 took part in the experiment. All observers had normal or corrected to normal vision.

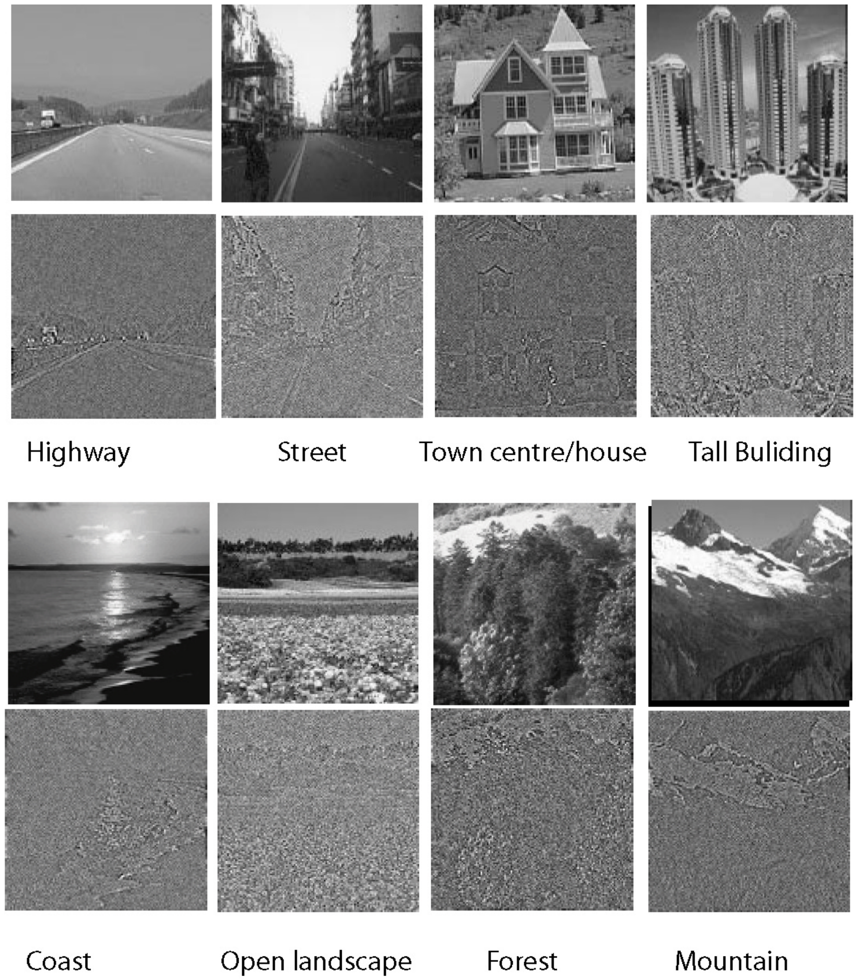

Figure 3. Examples of stimuli from each of the eight scene categories used in the experiment in their original format and in amplitude noise with phase density of .95.

Stimuli

Images from a scene database (Oliva & Torralba, 2001), 2400 in total, were used as stimuli. There were 300 examples of scenes for each category, and eight categories in total (highway, street, town centre/house, tall building, coast/ beach, open landscape, forest, and mountain; see Figure, 3). The taxonomy of the scene categories had been validated previously (see Oliva & Torralba, 2001). The 127×127 pixel 256 greyscale images subtended $9.3 \times 9.3^\circ$ of visual angle on the screen. Using a fast Fourier transform, we extracted the Fourier coefficients of each scene and ''whitened'' its amplitude spectrum by replacing it with the amplitude spectrum of white noise—resulting in an average energy slope equal to zero ($\alpha = 0$) across scenes.

In the whitened images, we introduced phase noise by randomly sampling the Fourier coefficients (see Figure 2b). The sampling range spanned all cycles per image from 1 to 63 (corresponding 0.11 to 6.76 cycles per degree of visual angle; 64 cycles per image = DC, which is not sampled), and for all orientations between 0° and 179°. For each sample, a mirror symmetric sample was constructed to extract spatial frequencies at orientations between 180° and 359°. We randomly transformed the phase ($-\pi$ radians to π radians) of nonsampled Fourier coefficients between 0° and 179° (orientation) at each frequency by replacing it with the phase of white noise (with a different white noise image computed for each stimulus). The phase information of all Fourier coefficients between 180° and 359° orientation was the same as the phase of the coefficients between 0° and 179° orientation, respectively. An inverse Fourier transform reconstructed a sparse experimental stimulus (see Figures 1e, 2c, and 3 for examples).

During the experiment, to maintain categorization accuracy at 75%, we adjusted online the density of the sampled phase, independently for each category (see Figure 2). Phase density was fixed at 95% for the first 50 trials per category to obtain a stable estimate of performance accuracy. Stimuli were constructed and the experiment was run using MATLAB version 5.0 and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), on a Macintosh G4 computer.

Procedure

Practice. Observers completed a practice session prior to the experiment to ensure they were familiar with the categorization task. In the task 160 images of scenes from the same scene database (20 examples per category) were used. None of the images used in the practice session were used in the experiment. Observers were presented with a greyscale image of a scene and asked to name it using one of the eight possible categories. Presentation of 160 images constituted one practice block. Each observer repeated practice blocks until they reached a criterion of 95% correct for one block.

Experiment. In a within-subjects design, a total of 2400 experimental stimuli were presented to each observer. Presentation was segmented into four blocks of 600 trials each. Order of presentation of experimental stimuli was randomized across observers. In all, the experiment lasted approximately 2.5–3 hours. Each scene was only presented once. On each experimental trial an observer categorized the sparse stimulus into one of the eight basic-level categories by pressing the appropriate labelled keyboard key. There were no constraints on response time, and the stimulus remained on the screen until their response.

RESULTS AND DISCUSSION

Figure 4a summarizes the average phase density required for observers to reach the 75% categorization correct performance criterion. Figure 4b shows a confusion matrix indicating the errors made with each scene category. Note that performance with coast scenes and landscape scenes fall below the performance criterion, and these scenes were often confused—this was true even when density of phase sampling was at the maximum allowed in the algorithm, 99.5% of phase information in phase, 0.5% of the phase scrambled. Figure 4c shows the average density of phase sampling per trial for each scene category averaged across subjects. Although the highest level of phase sampling was required for coast and landscape scenes, observers' performance remained well above chance level (performance accuracy of 45–50% when performance at chance level equals 12.5% correct for an eight alternative forced choice task).

It is possible that the visual features that typically occur in coast and landscape scenes are particularly sensitive to disruption of amplitude, and thus, cannot tolerate the effects of phase noise to the same extent as the other scene categories. Previous studies have shown that amplitude noise is most disruptive for perception of textured and shaded components occurring in natural scenes (e.g., the border between the coastline and skyline, contours of hills, and surface of a lake) that predominate coast and landscape scene categories (Morgan, Ross, & Hayes, 1991; Tadmor & Tolhurst, 1993). The four artificial scene categories, and the forest and mountain scene categories, consisted mainly of well-defined edges and were less affected by amplitude noise, and, consequently, able to tolerate higher levels of phase noise (see Figure 3 for a comparison of sparse stimuli). It is also likely that the visual characteristics of the basic-level scene categories coast and landscape are not ''redundant'' enough, or have low internal practicability (Gosselin & Schyns, 2001b, 2002). That is, the features typically occurring in coast scenes, (e.g., horizon line between sky and sea, ripples of sea) also occur frequently in landscape scenes (horizon line between sky and landscape, ripples of lake), leading to confusions between the two categories (e.g., 38.5% of responses to coast scenes were in the landscape category; see Figure 4b).

Now, we turn to examine the spectral information (spatial scale and orientation) that was effective for scene categorization. For each trial, we recorded the location of all the sampled Fourier coefficients together with the accuracy of the observer (correct or incorrect). Across the trials of a category, regularities should emerge in these paired locations and accuracies if the corresponding phase information represents a discriminative property of this scene category. The dual information of correct and total Fourier coefficients was kept

Figure 4a

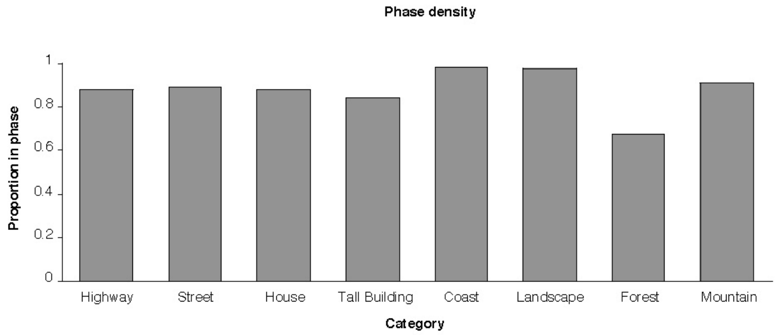

Figure 4b

	Highway	Street	House	Tall Building	Coast	Landscape	Forest	Mountain
Highway	70.97	15	2.78	.89	1.06*	5.92	2.48	.9
Street	5.16	70.76	14.02	3.32	.25	1.29	4.19	1.02*
House	.67	12.4	71.4	10.37	.33*	.85	3.4	.59
Tall Building	.52	3.81	10.9	75.01	.35	1.02*	7.1	1.29*
Coast	1.59*	1.14	.89*	.56	43.35	38.05	5.32	9.11
Landscape	4.77	2.60	1.57	.59*	9.79	49.72	15.24*	15.71
Forest	.33	1.95	4.37	3.6	.68	4.41*	78.13	6.52
Mountain	.73	1.3*	1.86	1.46*	2.9	9.92	10.7	71.13

Figure 4c

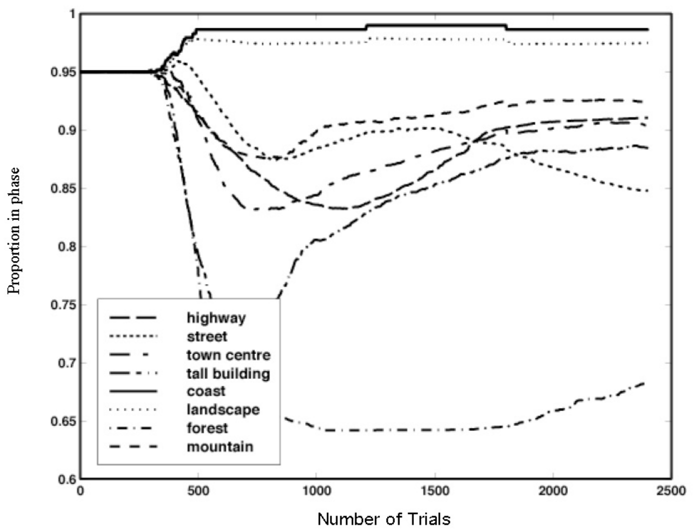

Figure 4. (a) Proportion of phase spectrum required for 75% accuracy per scene category. (b) Distribution of responses per scene category in percentages, including error responses. (c) Mean phase density (specified by a gradient descent algorithm) across trials for each scene category.

separately for each for the eight basic-level categories, and averaged across all observers. For each category, we then computed the proportion of correct over total (correct/total), for each Fourier coefficient. This proportion is the observer probability that a given coefficient leads to a correct categorization. To the extent that the amplitude information of this Fourier coefficient was whitened, the probability isolates the contribution of phase information.

For each scene category, we have a total of 12,644 Fourier coefficients to examine, hardly a small dimensional space! To simplify the data, we averaged the proportions associated with each Fourier coefficient according to 12 orientations of spatial frequency (from 0° to 179° by increment of 15° intervals) and four spatial frequency bandwidths (0–8, 8–16, 16–32, and 32–64 cycles per image; see Figure 1b for illustration). This segmented the data into 48 dimensions, each representing a different bandwidth and orientation given, respectively, by the radius length and angle in the semidiscs of Figure 5. A vertical orientation in the Fourier spectrum corresponds to horizontally orientated components in a scene, for example, the horizon line in a coast or highway scene.

We then transformed the segmented data for each category into 48 *Z*-scores (by computing an average and standard deviation from the 48 averaged proportions, independently for each category). A *Z*-score > 1.65 ($p < .05$) was considered ‘‘diagnostic’’. Figure 5 represents in red these diagnostic regions. All bandwidths and orientations were transformed into *Z*-scores, but all significant *Z*-scores (> 1.65) occurred in the 0–8 and 8–16 cycles per image bandwidths (0–1.74 cycles per degree), thus Figure 5 shows only these bandwidths.

How do these diagnostic regions in Fourier space correspond to 2-D image features in natural scenes? A striking aspect of our results is that the diagnostic bandwidths of the phase spectra for all scene categories occurred at relatively low spatial frequencies. Low spatial frequency information can provide a quick and rough estimate of a scene sufficient for fast recognition (Schyns & Oliva, 1994). For example, the localized structure (phase) of the components of a highway scene should provide sufficient information to discriminate a highway from the localized structure of components in mountain scene, even if image energy is obscured by white noise. To better relate the diagnostic orientations in the phase spectra to image features, we compared our *Z*-score data (Figure 5) with the averaged energy spectrums of scene categories taken from the same database of scenes (Oliva & Torralba, 2001). Remember that a vertical orientation in the Fourier spectrum corresponds to horizontally orientated components in a scene, for example, the horizon line in a coast or highway scene. The diagnostic phase spectra (present study) and the energy spectra (Oliva & Torralba, 2001) for coast and highway scenes were biased to vertical orientations. The visibility of the horizon line was described as the degree of ‘‘openness’’ in energy spectra. Horizontal phase and amplitude components correspond to vertically structured components in a scene, for example, the outline of a house

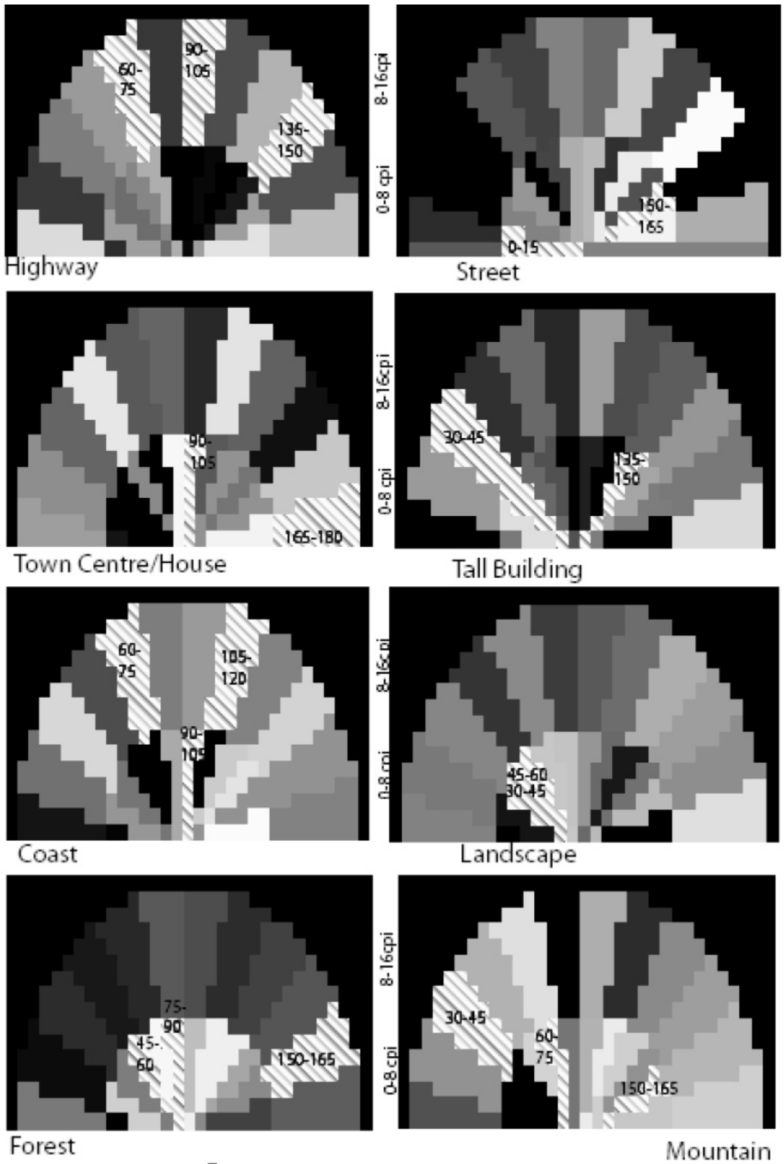

Figure 5. Plot of diagnostic phase in Fourier space (0–16 cycles per image, with orientation in degrees) for each scene category.

or building. In the present study, horizontal orientations in the phase spectra were diagnostic for town centre/house and street scene categories. This coincides with properties found in the averaged energy spectra of artificial (e.g., ''urban close up'' and ''city centre'') scene categories and is also described by Oliva and Torralba (2001) as degree of ''roughness'' in the energy spectra. Previous studies of image statistics have also shown that vertically structured features are common to artificial scene categories (Baddeley, 1997; Switkes et al., 1978; van der Schaaf & van Hateren, 1996). Diagonal orientations featured in the diagnostic phase spectra of mountain scenes, highways, streets, and tall buildings. Diagonal orientations in the Fourier spectrum correspond to scenes containing sloping edges, for example, the outline of a mountain, or perspective view of a street. Diagonal orientations occurred in the averaged energy spectra of mountain scenes (described as degree of ''ruggedness'') and described the degree of ''expansion'' in artificial scenes (e.g., vanishing lines in the perspective view of a scene; Oliva & Torralba, 2001). These comparisons suggest that the diagnostic phase spectra for each scene category coincide with characteristic amplitude spectra of the same scene categories reported by Oliva and Torralba.

However, a direct correlation of the results of Oliva and Torralba (2001) with our findings is not practical for the following reasons: First, we used the phase spectra in our study, not the energy spectra. While established methods exist for averaging the energy spectra of a set of images (e.g., van der Schaaf & van Hateren, 1996), averaging the phase information of a collection of images in a scene category does not provide a meaningful description. Previous studies of image phase in natural scenes have used higher order statistics to describe image phase (e.g., measures of skewness or kurtosis; Thomson, 1999). Thus, we cannot compare the ''average'' image phase of scene categories with our diagnostic phase spectra. Second, our data is not correlated directly with the energy spectrums of Oliva and Torralba because their study reveals the amplitude information *available* in the scene categories, whereas our diagnostic scene spectrums reveal the *potent* phase information—the subspace of available information used effectively—in these scene categories (Gosselin & Schyns, 2002).

How effective is this diagnostic phase information for discrimination of one scene category from another, and to what extent do diagnostic regions overlap between scene categories? The third phase of the analysis tested the effectiveness of the diagnostic regions of the phase spectrum to distinguish the images used in the experiment. To this end, we reconstructed the images used in the experiment using only the ''diagnostic'' regions of the phase spectrum, replacing nondiagnostic regions with the phase of white noise, and cutting off frequencies above 16 cycles per image (e.g., a coast scene with the diagnostic spectrum of *coast*). For each scene picture (e.g., one coast) we constructed seven distractors with the diagnostic spectra of the other scene categories (e.g., one coast with the phase of landscape, forest, highway, etc.). We then correlated the reconstructed scenes (both diagnostic and distractors) with the original images for each of the eight categories. A *t*-test (paired samples, $df = 7$) applied to the correlation coefficients of the diagnostic reconstructed scenes, and the correlation coefficients of the distractors, revealed higher correlations with the original image for images reconstructed with diagnostic phase spectra than for distractors, $t = 2.617$, $p < .05$. The correlational data demonstrates that even in high levels of phase noise, the diagnostic phase spectra for each scene category distinguished each scene from the nondiagnostic distractors. This implies that different regions of the phase spectra are diagnostic for different scene categories.

To examine the extent to which scene categories shared the same diagnostic orientations and bandwidths, the diagnostic regions for each scene category (see Figure 5) were added together and each region expressed as a proportion of the maximum possible overlap (from 0 to 8 categories). Figure 6 shows the frequency of diagnostic regions common to more than one scene category. Diagnostic regions shared by more than one category have a value above 0.125, those diagnostic for one scene category only have a value of 0.125 and regions used by none of the scene categories have a value of zero. The asterisked boxes in the table in Figure 4b indicate which scene categories overlapped. If observers were using one common area of the phase spectrum nonspecific to scene category, the number of regions shared by scene categories should be relatively high. Figure 6 shows that no one region is shared by more than two scene categories. This low level of overlap suggests that the local structures and edges described by the diagnostic phase spectra of a scene category are not common to many other

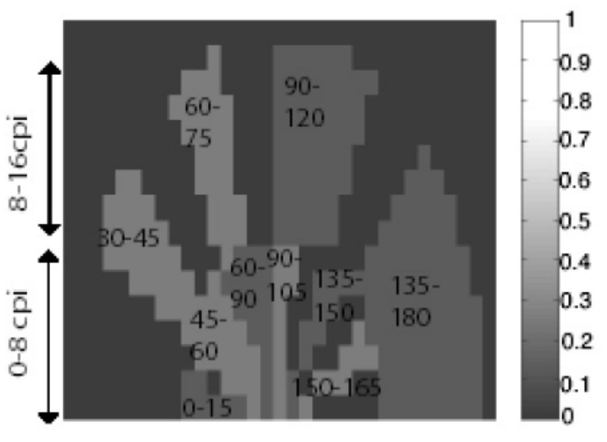

Figure 6. The diagnostic phase of all scene categories weighted by the frequency of occurrence for each bandwidth and orientation (0–16 cycles per image).

scene categories. For example, diagonal orientations at 30–45° and at 8–16 cycles per image are diagnostic for a mountain scene. This phase information should outline the sloping edge of the mountain, and differentiate it from images containing sloping edges that describe component features not specific to the ''mountain'' category (e.g., a highway).

CONCLUSION

In sum, we applied the bubbles technique (Gosselin & Schyns, 2001a) to the phase spectra of scenes to determine the spectral information that is effective for scene categorization. Analyses of the spectral information that led to correct categorizations produced diagnostic regions of the phase spectra that were category specific. According to the properties associated with basic-level categories (Gosselin & Schyns, 2001b), it is likely that these diagnostic orientations and bandwidths contain the scene information that minimizes the overlap of properties with other basic-level categories, and maximizes the number of properties specific to this category.

REFERENCES

Baddeley, R. (1997). The correlational structure of natural images and the calibrations of spatial representations. *Cognitive Science*, *21*, 351–372.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Campbell, F. W., & Kulikowski, J. J. (1966). Orientation selectivity of the human visual system. *Journal of Physiology*, *187*, 437–445.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, *197*, 551–556.

De Valois, R. L., & de Valois, K. K. (1988). *Spatial vision*. Oxford, UK: Oxford University Press.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra rapid visual categorisation of novel natural scenes. *Journal of Cognitive Neuroscience*, *13*, 171–180.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, *4*, 2379–2394.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, *6*, 559–601.

Gosselin, F., & Schyns, P. G. (2001a). Bubbles: A technique to reveal the use of information in recognition. *Vision Research*, *41*, 2261–2271.

Gosselin, F., & Schyns, P. G. (2001b). Why do we SLIP to the basic level? Computational constraints and their implementation. *Psychological Review*, *108*, 735–758.

Gosselin, F., & Schyns, P. G. (2002). RAP: A new framework for visual categorization. *Trends in Cognitive Sciences*, *6*, 70–77.

Guérin-Dugué, A., & Oliva, A. (2000). Classification of scene photographs from local orientations features. *Pattern Recognition Letters*, *21*, 1135–1140.

Henderson, J. M., & Hollingworth, A. (1999). High level scene perception. *Annual Review of Psychology*, *50*, 243–271.

Li, F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596–9601.

Morgan, M. J., Ross, J., & Hayes, A. (1991). The relative importance of local phase and local amplitude in patchwise image reconstruction. *Biological Cybernetics*, *65*, 113–119.

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: A phase dependent energy model. *Proceedings of the Royal Society of London, Series B*, *235*, 221–245.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*, 72–107.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*, 176–210.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, *69*, 529–541.

Parraga, C. A., Troscianko, T., & Tolhurst, D. J. (1999). The human visual system is optimized for processing the spatial information in natural visual images. *Current Biology*, *10*, 35–38.

Pelli, D. (1997). The video toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Piotrowski, L., & Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial frequency amplitude and phase. *Perception*, *11*, 337–346.

Potter, M. C. (1975). Meaning in visual search. *Science*, *187*, 965–966.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Sadr, J., & Sinha, P. (2004). Object recognition and random image structure evolution. *Cognitive Science*, *28*, 259–287.

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science*, *8*, 374–378.

Schwartz, N. Z., Tjan, B. S., & Chung, S. T. L. (2003). *Spatial frequency phase noise in fovea and periphery*. Paper presented at the annual meeting of the Vision Sciences Society.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, *4*, 819–825.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*, 402–409.

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

Sekuler, A. B., & Bennett, P. J. (2001). Visual neuroscience: Resonating to natural images. *Current Biology*, *11*, R733–R736.

Simoncelli, E. P. (2003). Vision and the statistics of the environment. *Current Opinion in Neurobiology*, *13*, 144–149.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.

Switkes, E., Mayer, M. J., & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research*, *18*, 1393–1399.

Tadmor, Y., & Tolhurst, D. J. (1993). Both the phase and the amplitude spectrum may determine the appearance of natural images. *Vision Research*, *33*, 141–145.

Tadmor, Y., & Tolhurst, D. J. (1994). Discrimination of changes in the second order statistics of natural and synthetic images. *Vision Research*, *34*, 541–554.

Thomson, M. G. A. (1999). Higher order structure in natural scenes. *Journal of the Optical Society of America, A*, *16*, 1549–1553.

Thomson, M. G. A., & Foster, D. H. (1997). Role of second and third order statistics in the discriminability of natural images. *Journal of the Optical Society of America*, *14*, 2081–2090.

Thomson, M. G. A., Foster, D. H., & Summers, R. J. (2000). Human sensitivity to phase perturbations in natural images: A statistical framework. *Perception*, *29*, 1057–1069.

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*, 391–412.

Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology*, *15*, 121–149.

Van der Schaaf, A., & van Hateren, J. H. (1996). Modelling the power spectra of natural images: Statistics and information. *Vision Research*, *36*, 2759–2770.