Spatiotemporal templates for detecting orientation-defined targets

Masayoshi Nagai JSPS Research Fellow (Department of Psychology, McMaster University, Canada)
Patrick J. Bennett Department of Psychology, McMaster University, Canada
Allison B. Sekuler Department of Psychology, McMaster University, Canada

Using the classification image technique, the present experiments revealed several characteristics of human observers' spatiotemporal templates for the detection of texture-defined targets. The stimulus consisted of a five frame movie of a five by five spatial array of elements. The target was defined by the first- or the second-order characteristics of orientation-defined textures. When a target signal was presented across all five frames, human observers typically relied on the most reasonable cues in all five frames for detecting targets. In other words, they used the first-order cue for detecting the first-order target and used the second-order cue for detecting the second-order target. When the target signal was presented just during the third temporal frame, the temporal profile of the observers' spatiotemporal templates changed, so that only information presented near the third temporal frame was used. In addition, the type of spatial cue utilized also changed, so that for first-order target detection observers used second-order cues as well as first-order cues. This strategy was sensible, because both first- and second-order cues were available in this condition. There also was a trend toward increasing the extent of spatial information used when the temporal information was restricted, perhaps indicating that there is a space-time tradeoff in the information that can be used in these tasks. In addition, we showed that the classification image is useful way to reveal individual differences that are not shown with traditional psychophysical techniques.

Keywords: classification image, texture segregation, first second-order cues, spatiotemporal templates.

Introduction

The segregation of visual scenes is a critical process in early vision. In natural scenes, segregation is achieved mainly by extracting luminance- or color- defined edges separating different objects. However, the human visual system can segregate scenes into different regions even when no such cues are available. For example, texture patterns with line arrays at different orientations segregate from each other (e.g., Beck, 1966), and other cues can be used as well. These psychophysical studies have well described the visual attributes that can serve as cues for visual segregation. However, with traditional psychophysics techniques, it is not easy to show how each local element contributes to the visual segregation.

Neurophysiological studies have investigated the texture segregation process at a more local level, and have found evidence for both spatial and temporal modulation of V1 neurons. For example, the firing rate of V1 neurons is stronger at the border between two texture regions than at within a single region (Nothdurft, Gallant, & Van Essen, 2000). Moreover, Lamme (1995) has suggested that there are three temporal stages in the responses of V1 neurons. The first stage is a basic orientation tuning around the latency of 60 ms. The second is a border detection stage at the latency around 80 ms. Neurons responded more strongly when their receptive field were at the border of textures than when there was no texture-defined region. The third is a surface representation stage at about 120 ms.

Although neurophysiological studies clearly showed the spatial and temporal modulation of texture segregation of V1 neurons' responses, it is unclear how well activities of V1 cells contribute to the "whole" processing system for the visual segregation. For example, one higher visual system function, attention, selects which visual region or object should be processed faster and more profoundly than others. This means that higher visual stages could use each V1 cell's activity differently, depending on its attentional weights. Some V1 cells would contribute strongly to the texture segregation system, but others would not. It also remains unclear what the time course of the whole visual segregation processing system is.

Let's simply define 'human observers' as the whole visual segregation processing system. The present study tried to determine the spatial and temporal characteristics of human observers' visual segregation processing. Specifically, we asked: Which regions in the texture are actually critical for the segmentation? At which times are the regions actually used? These questions are very difficult to answer using traditional psychophysical methods. However, one technique that has recently become more popular, response classification (Ahumada & Lovell, 1971), may be useful in helping us answer these questions. This technique is characterized with noise presentation and response classification. In each trial, unique external noise is added to stimuli that an observer must classify (e.g., A or B). On some trials, the observer's classifications will be correct. However, on other trials the noise may make one stimulus (e.g., stimulus A) more like the other stimulus (e.g., stimulus B), leading to incorrect classifications. After many trials, these noise fields can be classified into four stimulus/response classes (N_{AA}, N_{AB}, N_{BA}, and N_{BB}). Here, N_{AB} represents all samples of noise fields where stimulus A was presented and the observer classified it stimulus B. The mean classification image (C_{mean}) is calculated as follows:

$$C_{mean} = [\mathrm{Mean}(N_{AA})+\mathrm{Mean}(N_{BA})]-[\mathrm{Mean}(N_{BB})+\mathrm{Mean}(N_{AB})] \quad (1)$$

The classification image is a map that shows the locations in the stimulus that have affected an observer's responses, or a correlation between the noise magnitude at each location in the stimulus and the observer's response to that stimulus. The classification image is an estimate of the linear template used by an observer (Murray, Bennett, & Sekuler, 2002). Moreover, it also can estimate the effects of nonlinear mechanisms(Neri & Heeger, 2002). The variance classification image (C_{var}) , which estimates one kind of nonlinear template, is calculated as follows:

$$C_{var} = [\mathrm{Var}(N_{AA})+\mathrm{Var}(N_{BA})]-[\mathrm{Var}(N_{BB})+\mathrm{Var}(N_{AB})] \quad (2)$$

Experiment 1: Detection of orientation-defined the 1st- and 2nd-order sustained target

We used orientation-defined textures of five by five elements in five temporal frames. The target was presented in the center three rows, and it was defined by the first- or the second-order orientation cues. The target was presented across all five frames. We used response classification to derive spatiotemporal templates for detecting differently defined targets.

Method

Observers. Four undergraduate students and one graduate student at McMaster University participated. All had normal or corrected-to-normal vision and were naïve as to the purpose of the experiment.

Apparatus. Stimuli were displayed on a 21 inch AppleVision monitor (resolution: 1152 x 870 pixels, size of screen: 38.0 cm x 28.5 cm, refresh rate: 75Hz), controlled by an Apple G3 computer. Observers viewed the stimuli binocularly from the distance of 100 cm, and head position was stabilized with a chin-and-forehead rest.

Stimuli. A stimulus in each trial consisted of a movie of five by five arrays of oriented-line blobs (Figure 1A). A movie consisted of five frames of arrays. Each frame was presented for 80 ms. Each oriented-line blob was displayed within an area of about 0.264 x 0.264 degrees and the center to center separation between the blobs was approximately 0.340 degrees. Thus, the total stimulus size was approximately 1.623 x 1.623 degrees. After the presentation of the 5-frame movie, a five by five array of circular blobs was presented as a mask. The lines and circular blobs had a negative contrast of 50% against the background (46.63 cd/m^2).

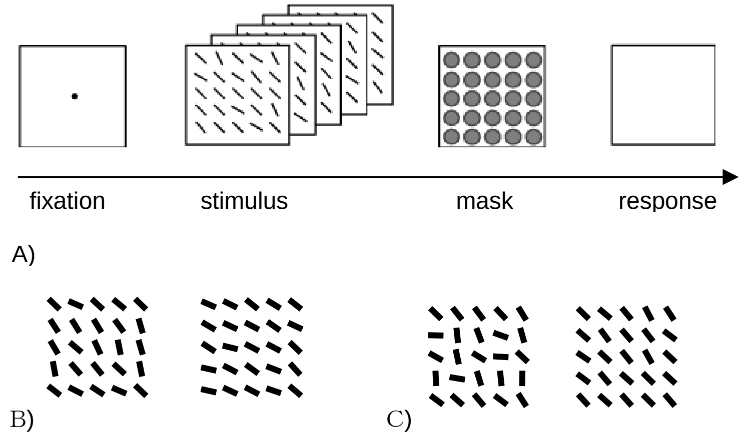

Figure 1. The stimuli use in the present study. A) The time course of a trial. B) An example of the first-order orientation-defined target (left) and non-target (right). C) An example of the second-order orientation-defined target (left) and non-target (right).

Observers attempted to discriminate between textures that contained a target from textures that did not. In a non-target texture, the orientations of all elements were drawn randomly from the same distribution. In a target texture, the orientations of elements in the middle three rows were selected from one distribution whereas the orientations of the remaining elements were drawn from a different distribution. In the first-order condition, the distributions of target and non-target orientations were uniform distributions (width = 40 deg) that differed only in mean orientation. Specifically, the mean target and non-target orientations were (135+d) and (135-d) degrees, respectively. The difference between means, 2d, was adjusted so that each observer responded correctly on approximately 75% of the trials. For the second-order condition, the distributions of the target and non-target orientations were uniform distributions that had the same mean (135 deg) but different widths (i.e., variances). Specifically, the non-target uniform distribution had a width of 40 deg and the target distribution had a width of 40+w. The difference in distribution width, w, was adjusted so that each observer responded correctly on approximately 75% of the trials.

Each trial began with the fixation point at the center of the screen. The fixation point was presented either 39 % negative or positive contrast. Pressing the space bar started each trial. 547 ms after the key press, 80 ms of the blank screen was presented, followed by the 80 ms x 5 frames of stimuli, and 507 ms of mask. After that, observers were required to judge whether the target was presented or not. Auditory feedback indicated whether the observer's response was correct or incorrect. 1000 ms after the response the next trial began (Figure 1A).

Procedures. Observers started with either the first- or the second-order sustained target detection task and then switched to the other task. The order of the tasks was counterbalanced across observers. Each observer participated in three one-hour experimental sessions of 1200 trials with the level at 75 correct thresholds. Before the experimental sessions, 75% correct thresholds of the first- and the second-order target detections were determined based on three to seven training session with method of constant stimuli.

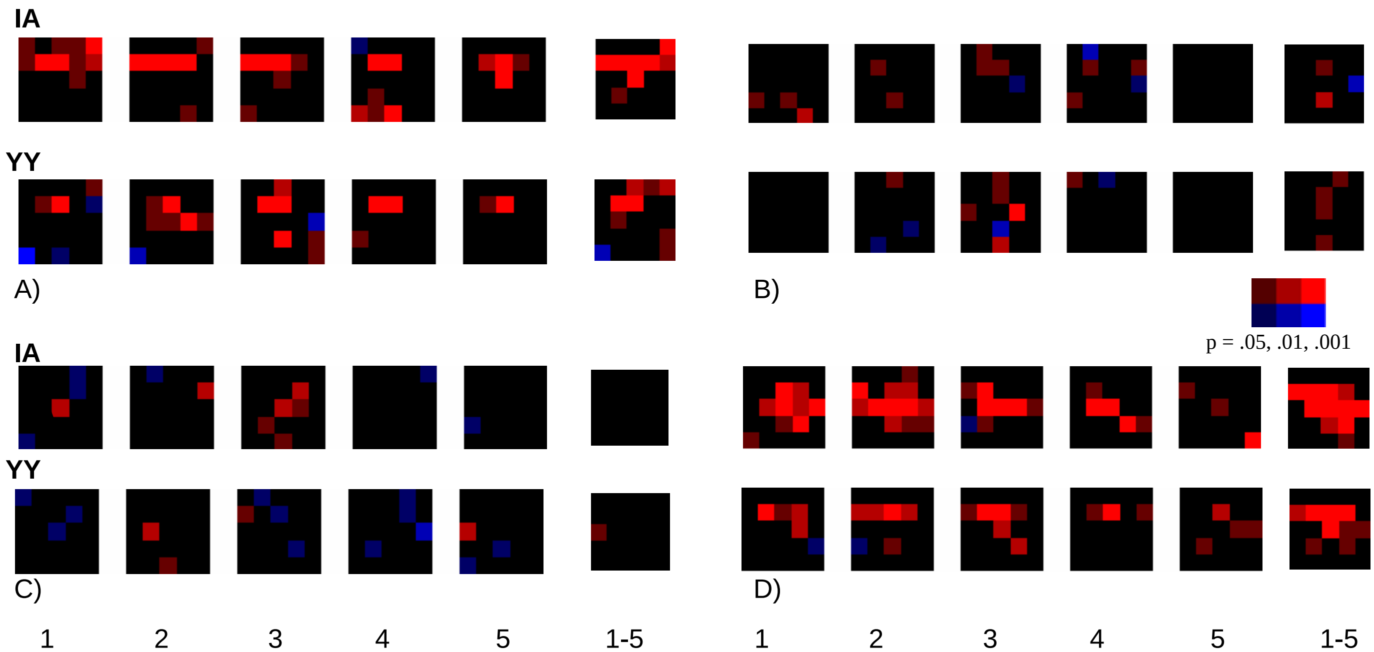

Figure 2. The results in Experiment 1 A) The mean classification images for the first-order sustained target detection in each time frame: the first to the fifth frames. The rightmost one shows the mean classification image from each spatial position after averaging across five temporal frames. B) The variance classification images for the first-order sustained target detection. C) The mean classification images for the second-order sustained target detection. D) The variance classification images for the second-order sustained target detection. The different brightness of red and blue pixels represents the different significance levels (p < .05, .01, and .001).

Results

Classification images for the first- and the second-order sustained target detections from observers IA and YY are shown in Figure 2. In this figure the numbers represent the temporal fames: classification images for frame 1,2,3,4,5. “1 - 5” means the collapsed images across all temporal frames. Red and blue pixels represent spatiotemporal loci which were significantly different from chance (p < 0.5, 0.1, and 0.01), and therefore were reliably associated with the observer's responses. In first-order (mean) classification images, red pixels indicate that the probability of an observer responding "target present" was significantly and positively correlated with the steepness (i.e., orientation more vertical than 135 deg) of that spatiotemporal element, whereas blue pixels indicate that the probability of "target present" responses were significantly and negatively correlated with the orientation steepness of that element. In second-order (variance) classification images, red pixels indicate spatiotemporal locations where large deviations (positive or negative) away from the mean orientation were significantly and positively correlated with the probability of observer's responding "target present", whereas blue pixels indicate spatiotemporal locations where large deviations from the mean orientation were significantly and negatively correlated with "target present" responses.

Figure 2A shows that, for the first-order sustained target detection, observers IA and YY did not use all available information, but instead used the upper and/or bottom edges across all the frames in the mean classification images. However, they did not consistently use any elements in variance classification image (Figure 2B). These results indicated that observers used a reasonable strategy for detecting the first-order sustained target. Figure 2C shows that, for second-order sustained target detection, observers IA and YY did not consistently use any element in the mean classification images, but did use the center and border elements, especially in the first three frames, in the variance classification images (Figure 2D). Again, the results indicate that observers used a reasonable cue to detect the second-order sustained target. It is interesting to note, however, that the spatial characteristics of the classification images differed slightly across conditions: observers used greater spatial areas for the second-order sustained target detection than for the first-order sustained target detection. Another observer JM replicated the results described above.

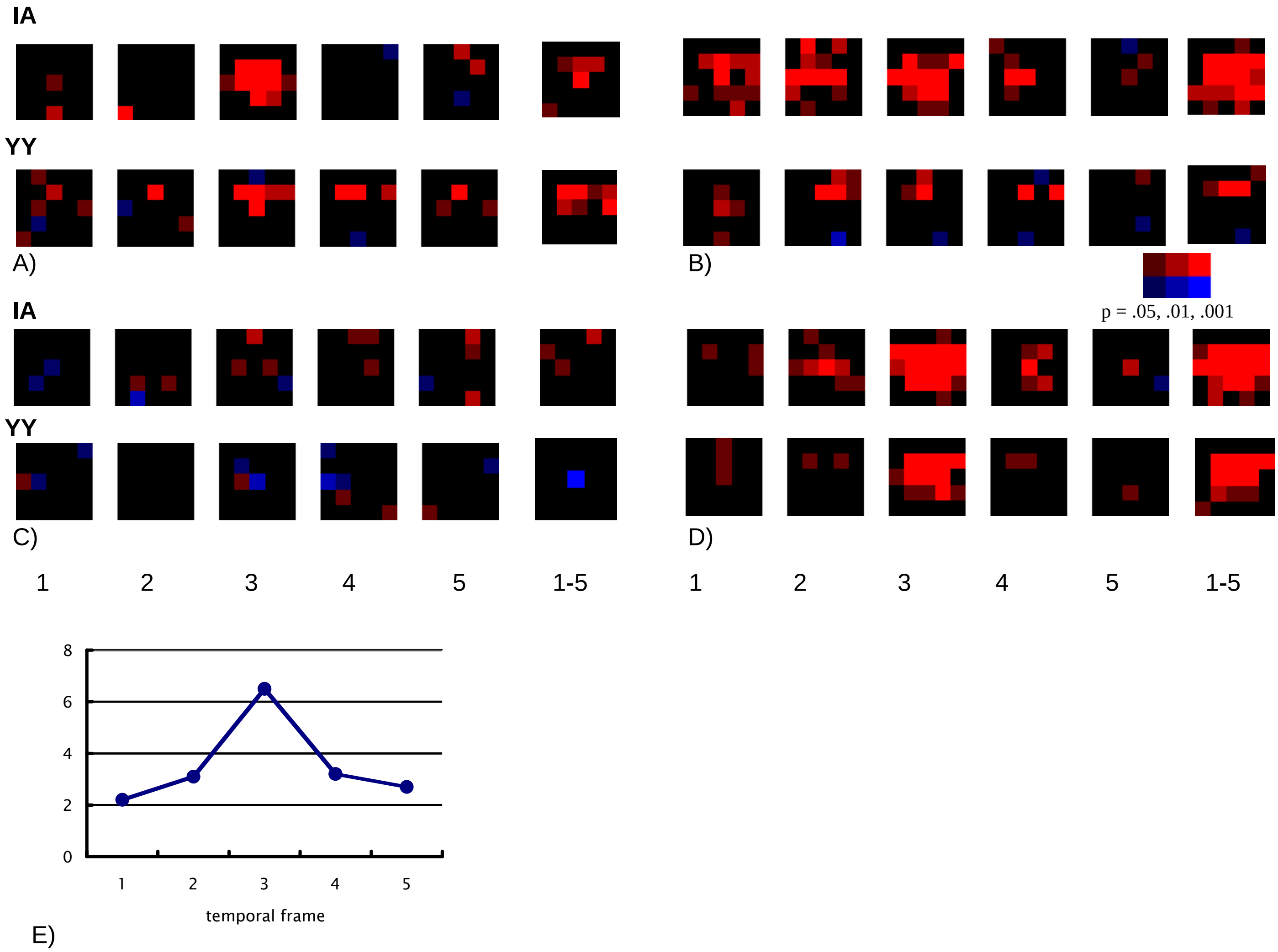

Figure 3. The results in Experiment 2. A) The mean classification images for the first-order flashed target detection in each time frame: B) The variance classification images for the first-order flashed target detection. C) The mean classification images for the second-order flashed target detection. D) The variance classification images for the second-order flashed target detection. The different brightness of red and blue pixels represents the different significance levels ($p < .05$, .01, and .001). E) The time course of the magnitude of the center of the second row elements in the mean classification image from observer YY.

Experiment 2: Detection of orientation-defined the 1st- and 2nd-order flashed target

The classification images in Experiment 1 did not provide any evidence of temporal tuning: information in nearly all frames was correlated with observers' responses. This failure to find temporal tuning may have been due to the fact that the stimulus was presented on every temporal frame. In Experiment 2, therefore, the target signal was presented only in the third frame ("flashed" target presentation). All other aspects of the procedure were the same as in Experiment 1. Observers IA, and YY participated in this experiment.

Results

Figure 3A and 3B shows the results for the first-order flashed target detection. The classification images were markedly different from those in Experiment 1. First, observers IA and YY used the second-order information (Figure 3B) as well as the first-order information (Figure 3A) for detecting the first-order flashed target. The use of the second-order information began at the first temporal frame and ceased at the forth frame. This trend is very interesting because the target was defined in the same way as Experiment 1. However, only in the flashed stimulus condition did observers use a second-order cue to detect a first-order target. Second, there were pronounced individual differences in the first-order (mean) classification image.

Observer IA used the first-order signal only during the third frame and significantly used 10 of 15 elements in the target presented area. In other words, Observer IA's first-order template exhibited narrow temporal summation but broad spatial summation. Note that in the previous experiment this observer used just the upper edge in the target presented area. In contrast, observer YY's first-order template exhibited broad temporal summation (i.e., all temporal frames were used) but narrow spatial summation (i.e., only a few spatial locations were used). Therefore, in general, this individual difference might illustrate a kind of space-time tradeoff in the spatiotemporal tuning in accessing the first-order information. However, it is important to note that even observer YY exhibited some first-order temporal tuning. For example, one element of the classification image in the center of the second row shows clear temporal dynamics (Figure 3E). The noise magnitude peaked at the third frame, which means observer YY the most strongly used that element in the third frame.

Figure 3C and 3D shows the results from observers IA and YY for the second-order flashed target detection task. For both observers, significant pixels were found primarily in the second-order (variance) classification images (Figure 3D). These significant pixels were clustered in time, occurring mostly during the one frame that contained the target (i.e., frame 3), but were distributed spatially over the entire target. It is interesting to not that the spatial structure of the second-order classification image was similar for flashed and sustained targets. Therefore, these results suggest that the spatial structure of first and second order templates are not affected by changes in the temporal characteristics of the stimulus.

General Discussion

Using the classification image technique, the present experiments revealed several interesting characteristics of the spatiotemporal templates that observers used to detect texture-defined targets. In general, although observers did not use all of the available stimulus information, the sources of information that observers did use were well matched to the type of target. For example, in the sustained conditions, observers used information that was distributed temporally across all five stimulus frames, whereas in the flashed target conditions observers relied most heavily on information on the one frame that contained the stimulus. Also, in the sustained target condition, observers used first-order cues to detect first-order targets, and second-order cues to detect second order targets. The results were slightly different with the flashed target: in that condition, observers used second-order cues to detect a second-order target, but used both first- and second-order cues to detect a first-order target. Finally, the results of Experiment 1 suggest that first- and second-order templates may differ in interesting ways in some conditions. For example, when detecting the first-order, sustained orientation-defined target, observers relied on cues near the target-background boundary. However, when detecting the second-order, sustained target, observers used collected information from a broader spatial area within the target itself.

Conclusion

Classification images can be used to characterize the spatiotemporal templates for detecting texture-defined targets. This technique successfully shows how elements localized in space and time contribute to decisions, and complement studies of figure-ground segregation that use standard psychophysical and neurophysiological methods. Moreover, this technique enables us to clearly visualize individual differences may not be apparent in more global measures of performance (e.g., thresholds).

Acknowledgement

This study was supported by JSPS Research Fellowship for Young Scientists (No. 200201154), NSERC Discovery grants (42133 & 105494), and the Canada Research Chair Program.

References

Ahumada, A. J., Jr., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America, 49,* 1751-1756.

Beck, J. (1966). Effects of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics, 1,* 300-302.

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Cerebral Cortex, 9*, 406-413.

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision*, *2*, 79-104.

Neri, P. & Heeger, D. J. (2002). Spatiotemporal mechanism for detecting and identifying image features in human vision. *Nature Neuroscience, 5*, 812-816.

Nothdurft, H. C., Gallant, J. L., & Van Essen, D. C. (2000). Response profiles to texture border patterns in area V1. *Visual Neuroscience, 17,* 421-436.