

動物検出課題において高次画像統計量は手がかりとなるか

坂野 逸紀

京都大学大学院人間・環境学研究科

日本学術振興会

齋木 潤

京都大学大学院人間・環境学研究科

We can rapidly judge whether images of natural scenes contain animals or not. So what information do we utilize to rapid animal detection? We investigated whether higher-order image statistics proposed by the texture synthesis algorithm of Portilla & Simoncelli (2000) could be a cue for animal detection. We compared detection performance between three types of distracter images: synthesized textures which shared statistics with animal images of natural scenes, synthesized textures having the same statistics as non-animal images and non-animal images themselves. Participants had to detect animals in pictures with brief presentation (40ms) as quickly and accurately as possible in the Yes/No paradigm. On each trial, a single image was located on the fixation point (0 deg) or at the eccentricity of 14 deg. The detection performance was significantly lower than other distracter conditions at both eccentricities when texture images sharing the statistics with animal images were used as distracter. We also carried out the same task using new images with their amplitude equalized. The result ruled out the possibility that amplitude difference between animal and non-animal image sets cause the performance impairment. These findings suggest that humans make use of higher-order statistics for rapid animal detection.

Keywords: animal detection, categorization, natural image, statistical summary, texture synthesis, eccentricity.

問題・目的

ヒトは、複雑な視環境の中からそこに存在する動物を極めて高速に検出できることが知られている。では、この検出にはどのような情報が寄与するのであるか？この問いに対し、画像統計量は有力な回答候補の1つとされる。しかしながら、具体的にどのような統計量がヒトにとって重要なものかははっきりとは分かっていない。自然画像を対象とした過去の研究では、画像をフーリエ変換した際に得られる振幅スペクトル情報が注目されていた。しかしながら、振幅それ自体はヒトの動物検出にとって本質的に重要な情報ではないという指摘が行動実験研究からはなされている (Wichmann et al., 2010)。

本研究では、Portilla & Simoncelli (2000) により提唱された高次画像統計量(P-S 統計量)を新たに取り上げ、これが瞬間呈示された画像の中から動物を検出する手がかりとなるかを検討した。高次画像統計量とは、フーリエ空間における画像の位相情報を考慮した統計量を指す。その一種であるP-S統計量は、ヒトがテクスチャ刺激から瞬時的に得ている情報を現在のところ最もよく表現しているものの1つとされる。我々はこの統計情報が動物検出にも寄与しているという仮説のもと、実験を行った。なお、近年の研究において、P-S統計量は周辺視野の視知覚を支える情報であるとの指摘がある (Balas et al., 2009)。P-S統計量の利用は周辺視野のみに限られる可能性も考慮し、本実験では画像の呈示視野を条件に組み込んだ。

方法

協力者 実験1では15人、実験2では14人の大学生が実験に参加した。全員が裸眼ないしは矯正した上での正常な視力を有していた。

刺激 実験1では、Serre et al. (2007) の画像データベースから動物を含む・含まない自然画像を120枚ずつ選択した。また、Portilla & Simoncelli (2000) の合成アルゴリズムにより、これらの自然画像とP-S統計量を(近似的に)等しくするテクスチャ画像を120枚ずつ合成した(Figure 1)。ただし、実験1の刺激では動物を含む・含まない自然画像間の振幅スペクトルの違いは維持されたままであった。また、それらから合成されたテクスチャ画像も元画像と同じ振幅スペクトルを有していた。実験2では、まず自然画像の振幅スペクトルを全画像の平均値に均等化する処理を施したもので、およびそれを元に合成したテクスチャ画像が用いられた。全画像は256階調のグレイスケールであり、その平均画像強度とRMSコントラストは均一であった。

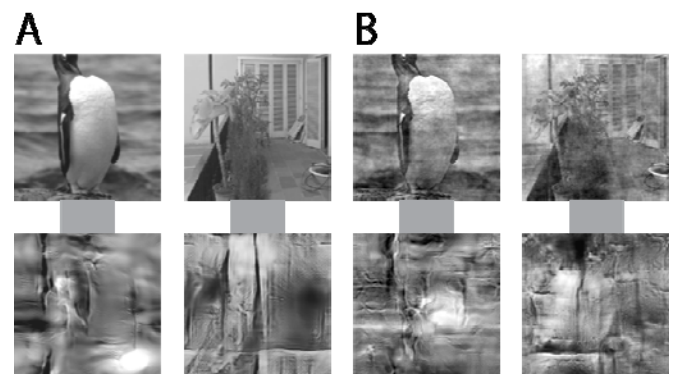


Figure 1. Examples of natural images (top) and textures (bottom) derived from them using the texture synthesis algorithm (Portilla & Simoncelli, 2000). A) Images used in Experiment 1. B) Images in Experiment 2 with amplitude spectrum equalized.

手続き

瞬間呈示された画像に動物が含まれているかを判断することが課題であった。画面中央に注視点が600 ms呈示された後、画像が注視点の位置、あるいは注視点から14° 右にずれた位置のいずれかに40 ms呈示された。協力者は、できる限り早く、かつ正確に動物の有無をボタン押しでYes/No回答するよう求められた。標的刺激は動物を含む自然画像であった。一方、妨害刺激は動物を含んだ画像を元に合成されたテクスチャ画像(Texture(Animal))、動物を含まない画像から合成されたテクスチャ画像(Texture(Non-animal))、動物を含まない自然画像(Non-animal)の3種類が用いられた。48試行を1ブロックとし、用いられる妨害刺激別に10ブロック、計1440試行が行われた。うち半数は0°、もう半数は14°の位置に画像が呈示された。

結果

検出成績として、 A' をFigure 2に示す。実験1において、両呈示視野でTexture(Animal)を妨害刺激とした条件の成績は他の2条件に比べて低かった。妨害刺激・呈示視野を要因とする2要因の分散分析を行ったところ(Greenhouse-Geisserの ϵ により自由度を補正)、両要因において主効果がみられた(妨害刺激: $F(1.35, 18.91) = 26.28, p < .0001$; 呈示視野: $F(1, 14) = 410.77, p < .0001$)。しかしながら、両要因の交互作用が有意であった($F(1.76, 24.64) = 8.95, p = .002$)。そのため、呈示視野別に妨害刺激条件ごとの成績を比較したところ、両視野でTexture(Animal)を妨害刺激とした条件の成績は他の条件よりも有意に低かった($p < .01$)。実験2においても、Texture(Animal)を妨害刺激とした条件は他の2条件よりも低い成績であった。妨害刺激・呈示視野の両要因で主効果がみられた(妨害刺激: $F(1.53, 19.94) = 14.72, p = .0003$; 呈示視野: $F(1, 13) = 81.05, p < .0001$)。これらの実験では交互作用は見られなかった($F(1.96, 25.48) = .53, n.s.$)。妨害刺激条件間を個々に比較したところ、Texture(Animal)を妨害刺激とした条件の成績は他の2条件よりも有意に低かった($p < .005$)。

上記した検出成績の差は反応時間とのトレードオフにより生じた可能性もある。そこで、各実験におけるヒット、コレクトリジェクション時の平均反応時間を同時に検討した。その結果、どちらの視野においてもTexture(Animal)を妨害刺激とした条件の反応時間は他の妨害刺激条件と等しい、あるいは他条件より遅い傾向にあった。このことから、トレードオフの可能性は除外された。

考察

実験結果から、動物を含む自然画像と高次画像統計量であるP-S統計量を共有するテクスチャ画像、すなわちTexture(Animal)を妨害刺激とした場合、動物検出はより難しくなることが分かった。この傾向は中心・周辺視野の双方において確認された。このことは、ヒ

トがP-S統計量に含まれる情報を中心・周辺視野で共に動物検出の手がかりとしていることを示唆している。

実験1では妨害刺激の種類と呈示視野に交互作用が見られ、実験2では見られなかったことは、今後の検討が必要であろう。これに関しては、振幅情報への依存度が中心・周辺視野で異なるために起こったという可能性が挙げられる。すなわち、課題を遂行する際、周辺視野は中心視野よりも振幅情報に相対的により依存しているというものである。実験1においては、振幅情報を標的刺激と共有していたのはTexture(Animal)のみであった。実験2では、他2条件においても標的・妨害刺激間の振幅情報の違いが無かった。その影響は周辺視野においてより強く、結果Texture(Animal)条件と他2条件の成績差が周辺視野においてより縮まった、と考えられる。

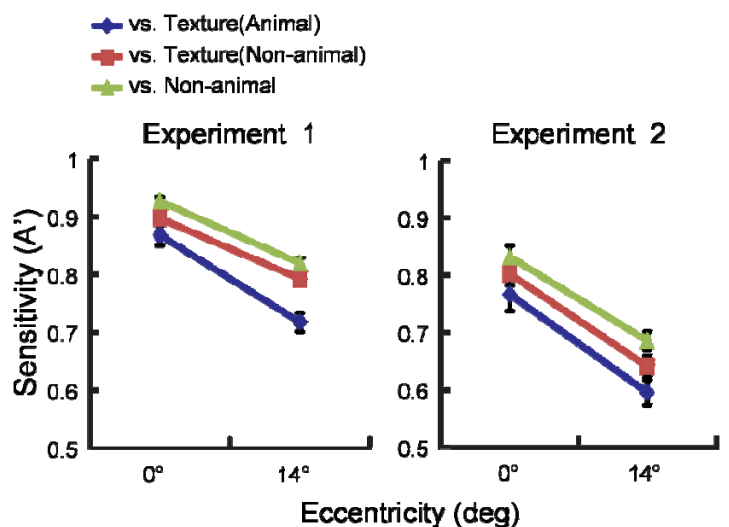


Figure 2. Detection performance in Experiment 1 and 2. The error bars indicate the standard errors.

引用文献

- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of vision*, 9(12), 13:1-18.
- Portilla, J., & Simoncelli, E. P. (2000). A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International Journal of Computer Vision*, 40(1), 49-71.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15), 6424-6429.
- Wichmann, F. A., & Gegenfurtner, K. R. (2010). Animal detection in natural scenes: Critical features revisited. *Journal of Vision*, 10, 1-27.