

Conceitos Básicos de Percepção de Cena

Basic Concepts of Scene Perception

Oséas Jarmouch Brito*

Paço da Imagem, Rio de Janeiro, RJ, Brasil

Resumo

A pesquisa sobre percepção de cenas é um esforço da comunidade científica em superar as dificuldades apresentadas ao estudo de estímulos complexos. As consistentes descobertas provaram a viabilidade do tema e encorajam uma abordagem mais holística e integrada nas investigações em percepção visual. Esse esforço e seus achados são, entretanto, pouco conhecidos pelos leitores em língua portuguesa e, até mesmo, entre alguns especialistas que se interessam pelo tema. O objetivo deste trabalho é fazer uma apresentação do campo, esboçar seu desenvolvimento, identificar suas principais questões e demonstrar sua pertinência e importância. Para isso, uma revisão seletiva de textos foi executada entre os autores que construíram os fundamentos da percepção de cenas. Ao concluir a leitura, espera-se que o leitor constate o papel relevante dos processos mnemônicos e do sujeito que observa no processo de captação visual do ambiente natural.

Palavras-chave: percepção de cenas; percepção visual; percepção visual de estímulos complexos; representação de cenas.

Abstract

Scene perception research is a scientific community effort to overcome difficulties presented to the study of complex stimuli. Consistent findings support feasibility of the theme and encourage a more holistic and integrated approach to visual perception research. This effort and its findings are, however, poorly known by readers in Portuguese language, even among experts interested in the subject. The purpose of this paper is to introduce this field of research, outline its development, identify its main issues and demonstrate its relevance and importance. This task was accomplished by a selective review of papers from authors who built the foundations of scene perception. Upon completion of the reading, it is expected that the reader understands the relevant role of mnemonic processes and of the observer in the process of visual capture of natural environments.

Keywords: scene perception; visual perception; perception of complex visual stimuli; scene representation.

1. Introdução

O estudo do processamento visual de estímulos complexos, com raras exceções, sempre foi evitado pelos pesquisadores interessados pela área da percepção visual.

* O. J. Brito - E-mail: oséasbrito@gmail.com

Dificuldades metodológicas e a crença de que o fenômeno pode ser melhor estudado a partir de seus constituintes mais simples, resultaram em experimentos com objetos isolados, apresentados distintamente no centro de uma tela ao observador experimental. Por décadas, a pesquisa de reconhecimento visual se concentrou mais em objetos do que em seus contextos (Bar, 2014). Entretanto, em nosso meio ambiente real, os objetos não se apresentam de forma tão pura e ordenada: oclusões, movimentos, cor, fragmentos, contextos e uma diversidade de outros fatores, são percebidos junto com o objeto. Assim, a cada recorte que nossos olhos efetuam em um cenário, nosso sistema visual recolhe uma multiplicidade de padrões de estímulos que precisam ser processados a fim de extrair sentido daquilo que se vê. Portanto, a percepção de estímulos complexos e a validade ecológica de suas investigações são desafios que passaram a ser sistematicamente enfrentados com o estudo da percepção do objeto inserido em um cenário e com o estudo da percepção do cenário que acolhe o objeto. Essa abordagem é conhecida como percepção de cenas.

O objetivo deste estudo é apresentar os conceitos essenciais sobre percepção de cenas. Para tanto, apresenta-se sua definição, resume-se seu desenvolvimento, identificam-se suas principais questões e pontua-se sua interação com outras áreas de conhecimento para, assim, enfatizar sua pertinência e importância. Com esse esforço, espera-se também preencher a lacuna existente em português de uma apresentação formal e organizada desse campo de pesquisa.

2. Método

A elaboração do presente trabalho baseou-se em revisões sobre percepção de cenas e nos textos dos autores que construíram seus principais fundamentos tais como Helene Intraub, John Henderson, Aude Oliva, Ronald Rensink, Mary C. Potter, Irving Biederman e Keith Rayner. Como o propósito era o de identificar os conceitos básicos do referido assunto, não se mostrou necessário adotar revisão sistemática em bases de dados a fim de inventariar a totalidade das pesquisas realizadas. As obras dos autores citados podem ser consideradas como sinalizadoras das discussões mais atuais e das possíveis tendências conceituais do campo.

3. Cena

O primeiro grande desafio do recente campo de investigação foi definir, com rigor, o que seria uma cena. Essa tarefa mostrou-se mais complicada do que parecia. Seria uma coleção de objetos, distribuídos circularmente, uma cena? Ou, então, um único objeto posicionado contra um fundo, tal como um pássaro contra o céu? Curiosamente, como acontece com outros conceitos, mesmo que nos escape uma clara definição do que seja uma cena, somos capazes de reconhecer quando estamos diante de uma cena real: ela “possuiria uma “cola” que mantém seus elementos juntos” (Bar, 2014, p. 1).

Existem, entretanto, certas semelhanças na operacionalização assumida por vários autores sobre cena. Aude Oliva (2010, 2013), por exemplo, a definiu como um recorte do ambiente no qual objetos e superfícies são arranjados de modo significativo como sala, rua ou floresta. Para a autora, uma cena, além de poder ser vista em diferentes escalas espaciais, tais como perto e longe, teria seus elementos organizados em um arranjo espacial. Na mesma linha, Henderson e Hollingworth (1999) a consideraram “uma visão semanticamente coerente

(e frequentemente nomeável), na escala humana, de um ambiente do mundo real, envolvendo elementos do fundo e múltiplos objetos discretos, organizados de forma espacialmente autorizado” (p. 244).

Como se pode notar, o conceito de arranjo ou configuração espacial é recorrente nas definições apresentadas e aí reside parte de seu valor nos enunciados definidores de uma cena: capacidade de captar uma preocupação central dos pesquisadores do tema. Importa, portanto, examiná-lo com cuidado.

Ronald Rensink (2000) desenhou, em *Scene Perception*, uma descrição do termo em discussão de fácil acesso e clareza ao estabelecer comparação com um objeto. Para ele, da mesma forma que o arranjo das partes de um objeto se apresenta de forma estável para o observador, assim deve ser o arranjo, disposição ou layout dos objetos e superfícies em uma cena: os elementos devem possuir uma posição relativa uns em relação aos outros. Assim, em uma sala, um sofá deverá estar no chão e não no teto. Henderson (2005), sem conflitar com essa noção, e incorporando aspectos já observados por Biederman, em seu trabalho de 1981, *On the semantics of a glance at a scene*, alicerce das atuais definições, explicou que ser espacialmente autorizado (*licensed*) significa respeitar as limitações das leis da física, incluindo gravidade, espaço, tempo e, também, adequação às restrições semânticas associadas à função do ambiente ou objetos. Como exemplo de conformidade às leis físicas, tem-se o fato de que mesas não flutuam no ar ou dividem o mesmo lugar no espaço com outro objeto. Como exemplo das restrições semânticas imposto pela função e identidade do objeto, cita-se a noção de que um cachecol não vestiria um poste de luz. Da mesma forma, uma cozinha, para ser considerada uma cena, deve respeitar as funções do ambiente cozinha; e uma faca, manteiga, copo de leite e pão, para conformarem uma cena, devem ser arranjados de acordo com a semântica de preparação de um sanduíche.

Relacionado ao conceito de configuração (*layout*) espacial e objeto, está a importante constatação de que cenas podem pertencer a diferentes categorias semânticas apesar de possuírem objetos ou layout semelhantes (Oliva, 2013). O corredor de uma aeronave comercial e o corredor de um ônibus de viagem possuiriam os mesmos objetos e o mesmo arranjo espacial sendo, entretanto, semanticamente distintos. A análise de uma cena envolveria perceber tanto os objetos, sua quantidade e localização como os tipos de superfícies (Oliva, 2013). As relações estabelecidas entre eles definiriam a cena.

Outro conceito chave na caracterização de uma cena é o de escala. A proposição acima apresentada de Henderson e Hollingworth (1999) refletiu o problema de designação enfrentado pelos pesquisadores e de seus esforços de tentar contorná-lo. Em geral, tendemos a classificar os elementos do mundo em duas categorias básicas: objetos e cenas (Oliva, 2010). Essa classificação, entretanto, sofre influência da escala (Henderson & Hollingworth, 1999). Em uma típica cena de sala de estar, a pequena mesa de centro seria um dos objetos da cena. Mas, em uma visão mais aproximada, a superfície da mesa de centro torna-se a cena e as revistas, cinzeiro e controle remoto seus objetos. Por ser difícil determinar quando uma escala espacial resultará em cena, os pesquisadores da percepção de cena assumem a visão de um ambiente ajustado à dimensão humana como escala espacial de referência: um ambiente que permita a locomoção humana e interação motora (Henderson, 2005; Henderson & Hollingworth, 1999). Ou, como diria Oliva (2013), agimos sobre os objetos, enquanto agimos em uma cena. Portanto, a visão de uma sala de jantar ou uma praça seriam consideradas boas cenas, enquanto a vista aérea da cidade e uma caixa de alfinetes não.

Apesar do caráter intuitivo e impreciso do uso da noção de escala e reconhecido por Henderson e Hollingworth (1999), ela possibilita formulações distintivas e classificadoras. Por exemplo, enquanto o objeto é contemplado pelo observador de um ponto de vista externo ao objeto, a cena é uma visão específica do ambiente que incorpora o observador. Portanto, a percepção de cenas difere da percepção dos estímulos visuais tipicamente empregados no campo da psicofísica e cognição visual (Henderson, 2005).

4. Perspectiva Histórica

A investigação dos processos envolvidos no reconhecimento visual evitou, em função da complexidade dos problemas envolvidos, a escolha de cenas como objeto de estudo. Ao longo da história, os pesquisadores assumiram, implícita ou explicitamente, que a pesquisa com objetos seria mais recomendável como ponto de partida ideal (Bar, 2014). Essa tradição manteve-se inalterada até que algumas realizações demonstraram a possibilidade de tratamento da questão. Mary Potter, David Navon e Irving Biederman são geralmente citados como pioneiros dessa mudança de paradigma (Oliva, 2013). Entretanto, foram os trabalhos desenvolvidos por Biederman, entre 1960 e 1970, que melhor expressam o início do deslocamento de uma abordagem centrada no objeto para uma abordagem centrada na cena (Bar, 2014). Em *On the semantics of a glance at a scene*, por exemplo, o autor apresentou sua pesquisa com a seguinte formulação: “Este capítulo trata da percepção e compreensão de cenas do mundo real” (p. 213). Nesse texto, pode-se constatar que Biederman assimilou e apostou nos postulados de Navon (1977) e Potter (1975), permitindo-lhe gerar os experimentos que sedimentaram a nova orientação. Por sua importância, expõe-se, a seguir, uma breve apresentação dos pioneiros estudos dos três autores.

Em 1975, Mary Potter apresentou uma tese em seu artigo *Meaning in Visual Search* que ainda é alvo de exame científico. Partindo da constatação de que o tempo médio de fixação do olho em uma cena nova dura cerca de 1/3 de segundo e de que imagens não familiares, quando observadas a essa velocidade, metade delas não são reconhecidas poucos minutos depois, a autora concluiu que, na observação de uma imagem, cada fixação seria rápida demais para que seja garantida a memória do que é visto. Determinada a melhor compreender a dinâmica das fixações, ela se propôs a investigar a possibilidade do observador detectar uma cena esperada, mas não conhecida, quando apresentada tão brevemente que seria difícil de ser lembrada. Seus dados confirmaram que indivíduos são capazes de identificar uma cena particular, entre várias apresentadas rápida e serialmente, apenas com uma descrição geral do alvo. Potter, assim, sublinhou a diferença entre detecção e memória e sugere que somos capazes de varrer o ambiente com breves fixações procurando não apenas por determinados padrões visuais, mas também por seus significados. Reside, aqui, um ponto de interesse de nossa breve exposição. Como especulou a autora, para que se possa “detectar um alvo definido pelo seu significado e não por um padrão visual específico, o observador, presumivelmente, teve que identificar cada cena.” (p. 966). E isso aconteceu com tempos de exposição de 125ms. Portanto, somos capazes de captar a essência de uma cena mesmo em apresentações tão rápidas como essa. Atualmente, designa-se o reconhecimento do significado geral de uma cena como *gist*.

Pouco tempo depois da publicação de Potter, como produto de sua pesquisa de doutorado, David Navon publicou, em 1977, *Forest Before Trees: The Precedence of Global Features in Visual Perception*. Nesse artigo, o autor enfrentou a questão dos níveis de

processamento de uma cena. Com rigor metodológico exemplar, ofereceu sustentação experimental para a tese de que o processamento das características mais globais de uma cena é feito antes que a análise mais local esteja concluída. Ou seja, o autor defendeu a precedência global do processamento visual do percepto. Portanto, um observador tende a reconhecer a floresta antes que se reconheça as árvores que a compõe.

O último pilar de sustentação do campo em estudo é o já referido *On the semantics of a glance at a scene*, de Irving Biederman, publicado em 1981. Nesse estudo, Biederman debruçou-se sobre a impressionante agilidade de caracterização visual de uma cena. Inspirado na pesquisa psicolinguística, como revelou o autor, resgatou o conceito de *esquema* de F. C. Bartlett atribuindo-lhe contornos mais nítidos. Por ser esclarecedor e aprofundar noções já abordadas, reproduz-se aqui parte de seu raciocínio.

Quando olhamos o mundo, mesmo brevemente, afirmou Biederman (1981), nossa impressão subjetiva é de conseguir, rapidamente e sem esforço, tanto uma percepção clara e integrada do que estamos vendo quanto sua compreensão. “A compreensão requer que não apenas as criaturas e objetos que constituem a cena sejam identificados, mas também que as relações entre essas entidades sejam especificadas” (p. 214). Essa constatação o levou a formular a pergunta norteadora de sua pesquisa: “Quais são as relações que distinguem uma cena bem formada de um mostruário de objetos aleatoriamente alocados?” (p. 216). Uma cena, analisou o autor, diferencia-se da imagem de uma coleção de objetos por possuir duas características: primeiro, relações especificadas, ou seja, para compreender como uma cena a imagem com os elementos homem, banco e harpa, é preciso que não apenas se identifique esses elementos, mas que saibamos que o homem está sentado sobre o banco enquanto toca a harpa; segundo, a possibilidade de que o observador produza uma “representação geral que integre as várias entidades e relações e permita acesso ao conhecimento do mundo sobre tais coisas” (pp. 214-215). Assim, o indivíduo sabe que ele está olhando para um concerto. Biederman (1981) usou o termo *esquema* para referir-se a essa representação. “O esquema especifica tanto os itens apropriados a uma dada cena quanto às relações semânticas e físicas que podem ser mantidas entre eles” (Biederman, 1981, p. 215).

É preciso observar que a exposição acima respondeu parcialmente à pergunta do autor. Embora ela apresente sua formulação para o conceito de *esquema*, deixa em aberto os tipos de relações que sustentam a diferença. Para tanto, o pesquisador postulou cinco tipos de relações: duas físicas e três semânticas. Recomenda-se a leitura do estudo para uma visão mais aprofundada da matéria. Cabe, entretanto, esclarecer que por relações físicas compreende-se as propriedades físicas que um objeto deve apresentar (não se pode ver através da maioria dos objetos, por exemplo) e por relações semânticas, de uma forma geral, entende-se como o conhecimento sobre a relação desse objeto com outros (é mais provável encontrar um hidrante em uma cena de rua do que em uma cena de cozinha). A grande descoberta de Biederman consiste na demonstração de que as relações semânticas são acessadas tão rapidamente quanto às relações físicas contrariando a tese de que um *esquema* só seria ativado após as relações físicas serem processadas. Na verdade, como enfatizou o autor, um grande volume de processamento semântico de uma cena pode ser executado a partir de uma única fixação sendo, portanto, desnecessário “postular uma sequência de fixações do olhar ou movimento para explicar a percepção de cena” (Biederman, 1981, p. 214). Em síntese, ao observador, é possível capturar o sentido geral e integrado das relações semânticas dos elementos que conformam uma imagem mesmo em brevíssimas exposições

delas. “É conveniente falar sobre o esquema de uma cena na tentativa de descrever esse rápido acesso à informação semântica” (Biederman, 1981, p. 253).

Em conjunto, esses três artigos não deixam dúvida de que uma quantidade significativa de informação de uma imagem pode ser capturada em décimos de segundo, que a cena deve ser considerada como um todo e que os aspectos globais da imagem podem ser acessados antes que ocorram os movimentos sacádicos.

5. A Pesquisa

A sistemática investigação da percepção de cenas mobiliza os mesmos métodos aplicados ao estudo dos processos cognitivos: comportamentais, neuropsicológicos, neuroimagéticos, computacionais, psicofísicos entre outros. Entre as questões mais importantes que esses métodos ajudam a tratar encontra-se o reconhecimento da cena: que tipos de representações os observadores constroem da cena?

Atualmente, duas propostas complementares se definiram para descrever a compreensão de cenas: abordagens centradas no objeto e abordagens centradas na cena ou espaço (Oliva, 2013). A primeira refere-se à segmentação dos componentes que passariam a funcionar como descritores da cena (*eis uma rua*, pois existem carros e prédios). A segunda refere-se a um tipo de representação na qual a configuração espacial e as propriedades globais da cena atuam como descritores (*eis uma rua*, pois é uma cena urbana, externa, cercada por superfícies altas com textura de padrões quadrados). Ao apreciar uma cena, portanto, o indivíduo teria disponível duas formas distintas, mas colaborativas, de construir a representação do estímulo e extrair informação dele. Sua capacidade de categorizar uma cena, portanto, se deve em função do reconhecimento imediato de objetos específicos ou da identificação das propriedades globais da imagem. O debate tende a girar em torno dessas perspectivas, mas alguns achados consistentes permitem fazer algumas afirmações. É conveniente apresentar um contexto mais geral para elas e esse é o objetivo das seções que se seguem.

5.1. Gist

Não seria um risco muito grande afirmar que o conceito de *gist* impulsionou a pesquisa em reconhecimento de cenas. Muito do esforço aplicado foi na tentativa de explicar a supracitada constatação de Potter e Biederman de que possuímos a habilidade de extrair o sentido geral de uma cena, sua essência (*gist*), mesmo com um tempo de exposição 125ms (Potter, 1975). Especificamente, estudos comportamentais demonstraram que esse tempo é suficiente para que os observadores reconheçam categorias de nível básico de uma cena (paisagem, urbana etc.), sua configuração espacial e outras características estruturais globais, tais como um grande volume em perspectiva (Oliva, 2005). Portanto, apesar do pouco tempo para sua elaboração completa, o indivíduo constrói uma representação da cena suficientemente rica para compreender seu significado, sua essência (*gist*) e reconhecer alguns dos objetos, antes mesmo que execute algum movimento do olhar (Rensink, 2000). Essa representação também possibilita que a atenção seja aplicada, controle os próximos movimentos dos olhos, identifique outros objetos (Oliva, Torralba, Castelano & Henderson, 2003) e permita o deslocamento do indivíduo no espaço (Rensink, 2000).

A habilidade de captar a essência de uma cena tão rapidamente levou os pesquisadores a examinarem o percurso temporal dos componentes envolvidos na percepção de cena. Como resultado, dois distintos estágios foram identificados: precoces e tardios (Oliva, 2013). Os estágios precoces (*early stages*) referem-se aos processos que ocorrem antes dos 100ms. Os estágios tardios (*late stages*) referem-se aos processos que ocorrem antes do observador movimentar seu olho (de 200 a 300ms). Como regra geral, é possível afirmar que com o avanço do tempo de exposição de uma imagem, mais detalhes de seu conteúdo são percebidos pelo observador. Por exemplo, no estágio inicial de processamento, uma cena pode ser classificada como uma paisagem larga e navegável sem que ainda se possa decidir se é uma montanha ou um lago (Oliva, 2013). A constatação alinha-se ao postulado de Navon (1977) sobre a prevalência da informação global sobre a local. De fato, o número de objetos em uma cena não afeta o tempo de reação de seu reconhecimento, o que sugere que seu significado seja inferido de sua configuração espacial (Oliva, 2010).

Importa, ainda, destacar que a percepção de cenas, como observou Rensink (2000), envolve os mesmos níveis de processamento da percepção em geral: baixo (*low*), intermediário (*intermediate*) e alto (*high*). A mesma consideração, como defende Oliva (2005), deve ser aplicada à captação da essência (*gist*) da cena.

5.2. Quando o olhar se move

A irregularidade da distribuição dos fotorreceptores na retina humana resulta em áreas da imagem captadas com diferentes níveis de detalhamento. Essa limitação na apreensão da informação visual é contornada com a movimentação do globo ocular que desloca o ponto de fixação do olhar de um ponto para outro a uma taxa de aproximadamente três vezes por segundo (Henderson, 2005). Diante disso, uma nova ordem de problemas se coloca para o claro entendimento da percepção de cena. E, obviamente, os achados dos estudos com o olhar estático devem ser levados em consideração. Assim, quando observamos uma cena, a cada fixação, uma *gist* é extraída e articulada com a próxima nos oferecendo uma coerente percepção do espaço (Oliva, 2010). Essa dinâmica abriga um importante ponto de discussão. Fenomenologicamente, experimentamos a cena de forma integrada, contínua e detalhada. Tal constatação parece sugerir que os detalhes de cada fixação são armazenados e integrados na memória de trabalho resultando em uma rica e fotográfica representação do mundo. Esse, porém, não parece ser o caso.

A pesquisa das informações mantidas na memória entre os movimentos sacádicos (memória transacádica) postula que a representação mental detalhada que é mantida entre um ponto de fixação e outro é uma ilusão: a memória transacádica é composta pelo significado da cena, o layout geral e apenas alguns detalhes (Intraub, 2002). Portanto, substitui-se o modelo de integração de representações detalhadas de tipo fotográfico por um modelo no qual o sistema visual usa a memória transacádica para relacionar as limitadas informações captadas nas fixações. Em uma apreciação de cena, o sistema visual nos oferece apenas informação suficiente para a compreendermos descartando as demais.

Entre as evidências experimentais que sustentam essa hipótese encontram-se os estudos sobre a capacidade do indivíduo em detectar mudanças em uma cena (Intraub, 2002; Rensink, 2000). Experimentos controlados demonstraram que mudanças efetuadas em um painel no exato momento em que o olho realiza uma sacada não são percebidas. Isso sugere que poucos detalhes da informação estariam sendo acumulados (Rensink, 2000). Sugere,

também, que as representações das cenas não são estruturas elaboradas a partir do movimento dos olhos e mudanças atencionais, mas antes estruturas que guiarão tais atividades (Rensink, 2000).

A limitação da memória visual não parece comprometer a adequação comportamental dos indivíduos ao meio. Essa constatação parece reforçar a noção de que outros fatores atuantes precisam ser considerados na percepção de cenas, como a memória de longo prazo e as expectativas (Biederman, 1981; Oliva, 2010). Helene Intraub investigou profundamente esses aspectos e encontrou um erro sistemático na percepção de cenas: observadores tendem a recordar os limites de uma dada cena de forma mais ampla que a original (2002, 2007, 2012). Isso quer dizer que quando uma pessoa é solicitada a desenhar uma cena anteriormente vista em uma foto, tenderá a reproduzir um desenho com o enquadramento da cena mais aberto, incluindo áreas que não foram observadas na imagem original. Esse fenômeno ficou conhecido como fronteira estendida (*boundary extension*). Ele revela a capacidade do sistema visual em prever a continuidade da configuração espacial da cena, o que, provavelmente, facilitaria a integração das sucessivas vistas geradas pelo movimento ocular (Intraub, 2002). A fronteira estendida sugere que as pessoas se baseiam em experiências anteriores para processar rapidamente os detalhes em uma cena. Ao se deparar com uma imagem pela primeira vez, o observador irá automaticamente incorporá-la em seu esquema, que inclui memórias semelhantes do que é visto e as expectativas sobre o que é provável de ser percebido em seguida (Oliva, 2010). Esse tipo de representação provavelmente reflete um tipo de mecanismo adaptativo que permite perceber o mundo além dos limites dos dados sensoriais de entrada (Park & Chun, 2014).

As evidências apresentadas estabelecem como uma ilusão a vivência de que captamos instantaneamente a totalidade da cena em seus ricos detalhes. Isso só é possível com múltiplas fixações do ambiente e, mesmo assim, poucos detalhes serão preservados na memória visual. E, como explica Rensink (2000a), eles nem são necessários, pois, caso seja preciso, um detalhe poderia ser imediatamente colhido pela rápida fixação da região em questão. Para o autor, é a rápida disponibilidade da informação que cria a experiência fenomenológica de um campo visual uniforme.

6. Revisões em curso

À medida que a pesquisa em percepção de cena avança, é natural que novos resultados sejam confrontados com os anteriores permitindo, assim, o progresso da área. Embora recentes questionamentos tenham surgido, o debate sobre a representação da cena permanece como central. As hipóteses de Rensink (2000a, 2000) e Intraub (2002, 2007) que tentam conciliar as limitações da retina e da memória transacádica com a experiência fenomenológica da visão, estão entre as mais influentes no campo. Intraub (2012), entretanto, é levada a reinterpretar seus achados em uma perspectiva mais ampla, resultando em um novo modelo de representação visual da cena. O amadurecimento de sua pesquisa a conduziu a classificar os modelos de representação visual em duas principais correntes: o tradicional, e já exposto, modelo visual-cognitivo (*visual-cognitive model*) e o modelo de fontes múltiplas (*multisource model*).

Basicamente, como explicou Intraub (2012), os modelos se diferenciam em termos de fonte de informação da cena, estrutura fundamental da representação da cena e o tipo de relacionamento que se estabelece entre a informação visual e conceitual. O modelo visual-

cognitivo, para a autora, postula que a representação visual da cena possui uma única fonte: as informações visuais de entrada. A cena visual captada evocaria rapidamente um conhecimento semântico que influenciaria as subseqüentes fixações. A integração dessas sucessivas vistas se basearia na recordação de representações esquemáticas da cena. Entretanto, como destaca Intraub (2012), foge ao modelo a possibilidade de explicar um erro construtivo como fronteira estendida, onde o observador “lembra-se” de informações que não estavam presentes no estímulo original.

Outro aspecto assumido pelo modelo tradicional é a separação entre a informação visual e o conhecimento conceitual (Intraub, 2012). Embora ambas sejam representações, elas seriam diferentes: uma concreta, visual-perceptual, e a outra abstrata. O modelo de múltiplas fontes, por outro lado, supera essa questão ao considerar como estrutura central na percepção de cenas não mais a informação visual, mas espacial: a representação egocêntrica do espaço. O modelo parte do pressuposto de que, no mundo real, o observador está incorporado à cena. Objetos e espaços estarão sempre em frente, à direita, ao lado, acima ou abaixo do observador: à minha frente, atrás de mim etc. Ao perceber uma cena, o estímulo visual informa ao observador um recorte egocêntrico que será completado pela percepção amodal¹, que constituirá uma representação que incorpora a informação visual dentro de um contexto esperado maior. Conceito e informação contextual preencheriam a estrutura egocêntrica com expectativas sobre o mundo além das fronteiras de uma fotografia, por exemplo. Isto é, o sentido egocêntrico do espaço elaborado pelo observador é uma estrutura organizadora que seria “preenchida” por múltiplas fontes, e não apenas pela informação visual, tais como o reconhecimento do objeto, percepção amodal além dos limites da visão e expectativas sobre a configuração do ambiente ao redor derivada da rápida classificação da cena. Assim, concluiu Intraub (2012), sobre o novo modelo, que: “

Ao invés de ser construída a partir da integração de visualizações sucessivas ao longo do tempo, a primeira fixação em uma cena desencadeia uma representação de fontes múltiplas do ambiente que pode ser refinada e atualizada por fixações subseqüentes. (Intraub, 2012, p. 118)

7. Estudos de neuroimagem

O mapeamento da atividade cerebral vem oferecendo relativo suporte ao modelo que considera a representação de uma cena como resultado do processamento dos diferentes tipos de informações que a constitui. Algumas áreas cerebrais se mostram mais nitidamente responsivas a características específicas da cena e, por isso, estariam associadas ao processamento dessa informação. Park e Chun (2014) realizaram valioso resumo dessa matéria e, apesar de reconhecerem que algum espaço para debate permanece, assinalam como regiões envolvidas na representação da cena a área para-hipocampal, o complexo retrosplenial, o sulco occipital transverso e o complexo occipito-lateral. O giro para-hipocampal responde preferencialmente às propriedades visuoespaciais da cena. O complexo retrosplenial e o sulco occipital transverso estariam envolvidos na análise da navegação espacial, criando uma representação integrada da cena. Em contraste com as regiões anteriores, que respondem preferencialmente às propriedades estruturais da cena, o complexo occipito-lateral ativa-se fortemente em resposta aos objetos.

Park e Chun (2014) apresentaram um quadro teórico do funcionamento distinto, mas complementar, das estruturas listadas. Os autores afirmaram que a observação de uma cena envolve a percepção de diferentes vistas. Cada vista, separadamente, terá sua propriedade

visuoestrutural representada pela região para-hipocampal, codificando propriedades tais como volume, perspectiva, configurações espaciais abertas ou fechadas, independentemente dos tipos de objetos que preenchem o espaço. Paralelamente, o complexo occipito-lateral representaria as propriedades do objeto, tais como conteúdos urbanos ou naturais. Já o complexo retrosplenial responderia pela integração das vistas de modo independente, representando como a mesma cena, as vistas espaço-temporalmente relacionadas. Colaborando com esse processo de integração e como reflexo das expectativas do observador, o hipocampo atuaria na formação de uma representação antecipatória ao próximo layout estando, portanto, diretamente relacionado ao fenômeno da fronteira estendida.

O modelo de Park e Chun (2014) demonstrou que a percepção de cenas envolve diferentes níveis ou estágios de processamento que integram espaço, tempo e memória e como esses estágios estariam relacionados às citadas regiões corticais. O valor do modelo reside no fato de recuperar as discussões abordadas até o momento e organizá-las em uma perspectiva neurologicamente plausível. Seus três estágios (construção estrutural, integração perceptual e construção da memória) e as propriedades processadas seguem as explicitadas discussões sobre a configuração espacial, a importância e reconhecimento do objeto, a experiência fenomenológica de uma imagem uniforme e o fenômeno da fronteira estendida.

8. Considerações finais

A intimidação imposta pelo estudo de estímulos complexos parece ser coisa do passado. Pesquisadores aceitaram o desafio de enfrentar as dificuldades impostas pela perspectiva holística da pesquisa em processamento visual. Seus resultados ecoaram sobre outros estudiosos do sistema visual encorajando-os a reorganizar os conteúdos do campo sob a nova perspectiva e em conceder-lhe destaque, como pode ser constatado em manuais introdutórios como os de Goldstein (2016) e de Wolfe et al. (2014). Por seu caráter interacionista, a percepção de cenas permite compreender com maior clareza a articulação dos fatores envolvidos em sua construção, tais como os processos mnemônicos e perceptuais. Ela, também, reconcilia o papel ativo do observador, pois leva em consideração suas expectativas e experiências.

É arriscado antecipar os próximos passos de uma área de pesquisa tão recente, que ainda está interpretando as implicações de seus resultados, entretanto é razoável supor que o problema da representação de cenas mobilizará a energia e atenção de uma parcela significativa de pesquisadores envolvidos com o campo.

Por fim, é preciso observar que a percepção de cenas evidencia tanto o aspecto dinâmico e construtivo do processo perceptivo, quanto sua elegância: percebemos e relembramos o mundo visual com uma profunda economia de recursos (Intraub, 2002).

Referências

- Bar, M. (2014). The current scene. In: K. Kveraga, & M. Bar, (Eds.), *Scene vision: Making sense of what we see* (pp. 1-3). Cambridge, MA: The MIT Press. Doi: 10.7551/mitpress/9780262027854.001.0001
- Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy and J. R. Pomerantz (Eds.), *Perceptual Organization* (pp. 213–253). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

- Goldstein, E. B. (2016). *Sensation and perception* (10th ed.). Belmont, CA: Wadsworth, Cengage Learning.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual review of psychology*, 50(1), 243-271. Doi:10.1146/annurev.psych.50.1.243
- Henderson, J. (2005). Introduction to real-world scene perception. *Visual Cognition*, 12(6), 849-851. Doi:10.1080/13506280444000544
- Intraub, H. (2002). Visual scene perception. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp.524-527). London, UK: Nature Publishing Group. doi: 10.1002/0470018860.s00632
- Intraub, H. (2007). Scene perception: The world through a window. In M. A. Peterson, B. Gillam, & H. A. Sedgwick (Eds.), *In the mind's eye: Julian Hochberg on the perception of pictures, films and the world* (pp. 454-466). New York: Oxford University Press.
- Intraub, H. (2012). Rethinking visual scene perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(1), 117-127. doi:10.1002/wcs.149
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*, 9(3), 353-383. doi:10.1016/0010-0285(77)90012-3
- Oliva, A. (2005). Gist of the scene. In: L. Itti, G. Rees, & J. K. Tsotsos (Eds), *Neurobiology of Attention* (pp.251-256). San Diego, CA: Elsevier. doi:10.1016/B978-012375731-9/50045-8
- Oliva, A. (2010). Visual scene perception. In: E. B. Goldstein (Ed.), *Encyclopedia of perception* (Vol. 2, pp. 1111-1115). Thousand Oaks, CA: SAGE Publication.
- Oliva, A. (2013). Scene perception. In: J. S. Werner, & L. M. Chalupa (Eds.), *The New Visual Neurosciences* (pp. 725-732). Cambridge, MA: The MIT Press.
- Oliva, A., Torralba, A., Castelhano, M. S., and Henderson, J. M. (2003). Top-down control of visual attention in object detection. *Proceedings of the IEEE International Conference on Image Processing*, Vol. I, 253-256.
- Park, S., & Chun, M. M. (2014). The constructive nature of scene perception. In: K. Kveraga, & M. Bar (Eds.), *Scene vision: Making sense of what we see* (pp. 45-72). Cambridge, MA: The MIT Press.
- Potter, M. C. (1975). Meaning in visual search. *Science*, 187(4180), 965-966. doi: 10.1126/science.1145183
- Rensink, R. A. (2000a). The dynamic representation of scenes. *Visual Cognition*, 7(1-3), 17-42. Doi:10.1080/135062800394667
- Rensink, R. A. (2000b). Scene perception. In: A. E. Kazdin (Ed.), *Encyclopedia of psychology*, Vol. 7, pp. 151-155. New York: Oxford University Press.
- Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., Lederman, S. J., & Merfeld, D. M. (2014). *Sensation and Perception* (4th ed.). Sunderland, MA: Sinauer Associates.

Notas

(1) *Percepção amodal* refere-se à percepção de toda estrutura física de um objeto mesmo que apenas uma parte dela tenha estimulado os receptores sensoriais.