

## Мгновенное восприятие естественных сцен и объектов

И.С. Уточкин

Ежесекундно зрительный анализатор получает и обрабатывает огромное количество информации, выделяя из распределенного по сетчатке паттерна световой энергии множество признаков – цвет, пространственную частоту, ориентацию, скорость и пр. В дальнейшем на основе этих признаков, по-видимому, выполняется серия операций, которая, в конечном счете, рождает осознанный перцептивный образ. Для этого необходимо каким-то образом соединить разрозненные признаки в целостные объекты, обособить одни объекты от других и опознать. В реальных условиях, однако, восприятие не ограничивается опознанием объектов. Пространственное расположение (компоновка) и взаимодействие объектов между собой обеспечивают информацию более высокого порядка, необходимую для восприятия и понимания целостных событий, явлений и ситуаций в окружающем мире. Именно эту способность извлекать информацию о взаиморасположении и свойствах больших множеств естественных объектов мы и будем понимать под *восприятием сцен*.

### *Легкость восприятия сцен*

В классических экспериментах с восприятием и запоминанием сложных фотографических изображений было показано, что схватывание сути сцены (*scene gist*) происходит довольно легко и требует совсем небольшого времени. Для категоризации основных частей сцены и даже многих ее деталей и последующего их узнавания достаточно в сущности одного короткого взгляда, которому соответствует краткое предъявление на 100-200 мс (Intraub, 1981; Potter, 1975). В потоке быстро сменяющихся друг друга фотографий испытуемые могут с высокой точностью определить, был ли им показан, например, дом, а также впоследствии узнать этот дом среди фотографий других домов. А увидев фотографию морского побережья, человек с одного взгляда запоминает, что в небе он видел стаю чаек, а на воде – яхта с желтым парусом.

Восприятие основных элементов сцены с одного взгляда – поистине удивительная способность, если учесть, что осознанное и детальное восприятие и запоминание объектов – процесс довольно медленный, к тому же требующий последовательной фиксации отдельных объектов с помощью внимания, объем которого ограничен (Luck & Vogel, 1997; Rensink, 2000; Treisman, 2006). По-видимому, это возможно за счет восприятия некоторых глобальных признаков изображения, минуя локальные признаки отдельных объектов.

Первичность восприятия глобальных признаков по отношению к локальным – старая идея, в явном виде продекларированная гештальтпсихологами и подвергнутая строгой проверке в когнитивной психологии (Navon, 1977). Однако значительное продвижение в понимании того,

каким образом происходит восприятие этих глобальных признаков, было сделано в последние 25-30 лет. Результаты этих исследований показывают, что восприятие глобальных признаков – это не один, а несколько феноменов, за которыми стоят и разные механизмы. В данном тексте мы последовательно рассмотрим три отдельных феномена, составляющих мгновенное восприятие сцены:

1. Оpozнание ландшафта, который составляет естественный фон сцены (например, природный или городской пейзаж, море или пустыня, сад или комната);

2. Различение и опознание сложных объектов, к которым относятся почти все естественные предметы и животные;

3. Оценка свойств и признаков множественных объектов (например, степени спелости ягод на малиновом кусте, средней скорости автомобильного потока на шоссе).

#### *Опознание ландшафта: модель «пространственного конверта»*

Понятие *ландшафта* применяется, как правило, к крупномасштабным сценам, где восприятие настроено на охват относительно большого пространства (лесов, полей, улиц, комнат и пр.). Для восприятия небольших пространств крупным планом (например, поверхности письменного стола с разложенными на ней книгами и канцелярскими принадлежностями) понятие ландшафта мы использовать не будем.

По-видимому, для опознания ландшафтов зрительной системе достаточно очень небольшого количества предельно глобальных признаков, не включающих никакой детальной информации о конкретных объектах. В самом деле, если бросить беглый взгляд на рисунок 1а, то легко будет опознан ландшафт улицы, хотя размытое изображение не дает детальной информации о типичных объектах: домах, фонарях, автомобилях. Если оставить видимыми только отдельные объекты, закрыв руками остальную часть изображения, то однозначно опознать их будет затруднительно (например, если закрыть нижнюю и верхнюю части рис. 1а, оставив только область автомобилей, то сами автомобили могут показаться чем угодно – например, камнями). Даже при достаточной четкости изображения объектов (рис. 1б) наличие уличного ландшафта играет с нашим восприятием шутку. Так, при беглом взгляде мы действительно видим улицу с нормальными домами и автомобилями, как и на рис. 1а. Но стоит присмотреться чуть внимательнее – и на месте домов мы ясно увидим кухонную мебель – шкафы, холодильник, плиту! Наличие глобальных признаков, таких как перспектива, открытость (голубой участок вверху, соответствующий открытому небу), упорядоченность объектов (дома стоят рядами, а не разрозненно) и др., задают типичный ландшафт улицы. Примеры, приведенные на рис. 1, демонстрируют нам не только легкость, с которой эти глобальные признаки извлекаются из изображения даже при дефиците информации о деталях.

Они также показывают нам, что скорее ландшафт влияет на категоризацию конкретных объектов, чем отдельные объекты формируют образ ландшафта.



Рисунок 1. Влияние глобальных ландшафтных признаков на восприятие сцены и объектов: а) несмотря на нечеткость контуров и отсутствие деталей, сцена легко опознается как ландшафт улицы; б) даже при достаточной четкости изображения глобальные признаки заставляют нас видеть объекты, типичные для данного ландшафта (предметы, кажущиеся домами, на самом деле являются кухонной мебелью) (Oliva & Torralba, p. 25, Fig. 1).

Какие же глобальные признаки формируют образ ландшафта? Серия исследований, которые провели О. Олива и А. Торральба (Oliva & Torralba, 2001; Torralba & Oliva, 2002), показывает, что эти признаки действительно доступны еще до выделения отдельных объектов; это признаки организации физического пространства, а не объектов. Вместе они образуют «пространственный конверт» (*spatial envelope*) – глобальный каркас сцены, который затем наполняется конкретными объектами и действующими лицами. В своих экспериментах Олива и коллеги обнаружили, что для различения сотен разнообразных ландшафтов человек пользуется весьма ограниченным набором признаков – всего порядка 5-6. Эти признаки доступны на очень ранних стадиях зрительного анализа (примерно после 34 мс просмотра) и предшествуют более точному опознанию ландшафта (например, в качестве пустыни, озера, горы и др.) (Greene, Oliva, 2009). К основным глобальным признакам ландшафта можно отнести следующие:

1. *Естественность*. Наблюдатели легко различают природные и рукотворные ландшафты. Для рукотворных ландшафтов характерно наличие прямых горизонтальных и вертикальных линий, в то время как естественные ландшафты содержат много текстурированных зон и преимущественно волнистые контуры поверхностей.

2. *Открытость*. Пространство сцены может быть ограничено множеством элементов (таких как деревья, горы, небоскребы), что создает закрытый пространственный конверт. Напротив, отсутствие загромождающих элементов, уходящая в бесконечность поверхность (земля или вода) и наличие ясно различимой линии горизонта создают открытый пространственный конверт.

3. *Сложность*. Данный признак описывает наличие иерархически организованных элементов сцены. Сложным можно считать ландшафт, включающий в себя множество разнородных элементов, а также встроенность этих элементов в другие, более крупные элементы (например, булыжники на склоне горы). Напротив, простыми ландшафтами можно считать те, где присутствует минимум выделяющихся элементов, а сам ландшафт максимально однороден и фактически воспринимается как единая текстура (например, ландшафт леса, видимый с высоты, где каждое дерево представляет всего лишь текстурный элемент).

4. *Перспектива*. В различении рукотворных сцен, где присутствует много прямых линий, отмечают Олива и Торральба, важную роль играет перспектива. Так, наличие сходящихся прямых линий легко распознается и дает стойкое переживание глубины (что характерно, скажем, для восприятия ландшафтов улиц, шоссе, парков и др.). Преобладание параллельных линий над сходящимися задают фронтальную плоскость, что характерно для фасадов домов, внутренних стен помещений и др.

5. *Неровность*. Этот признак является диагностическим, прежде всего, для естественных ландшафтов, поскольку рукотворные ландшафты обычно создаются на ровных поверхностях. Неровность характеризует степень отклонения контуров земной поверхности от горизонтальной ориентации – наличие наклонов, впадин и возвышений.

6. *Средняя глубина*. Этот важный структурный признак позволяет оценивать масштаб представленного ландшафта и различать локальные и панорамные сцены. По мнению Торральбы и Олива (2002), информация о глубине передается простыми частотными характеристиками изображения, не апеллируя ни к каким объектным признакам вроде знакомого размера. Однако условием для успешной передачи такой информации является опознание ландшафта в качестве естественного или рукотворного. Так, для естественных ландшафтов увеличение средней пространственной частоты коррелирует с более локальной точкой обзора и меньшей глубиной. Действительно, естественная панорама включает в себя небольшое количество глобальных образований (земля, вода, лес) с довольно однородной текстурой. При сильном приближении эта текстура становится весьма пестрой и разнообразной (трава, листья, рисунок коры деревьев и т.п.). Обратная закономерность наблюдается для рукотворных сцен. Как правило, небольшие искусственные поверхности (столы, стены, асфальт) сравнительно однотонны и не содержат большого количества высокочастотных

элементов. Однако панорамный вид городского пейзажа представляется гораздо более разнообразным. Теснящиеся дома, улицы, парковки и газоны создают очень плотную и разнородную глобальную текстуру.

Каждый ландшафтный признак биполярен и непрерывен. Таким образом, образуется психологическое пространство ландшафтов с осями-признаками, наподобие геометрического пространства с тремя измерениями. Любой типичный ландшафт представляет собой точку (или скорее компактное скопление точек) в этом психологическом пространстве. Узнавание конкретного ландшафта, соответственно, может быть представлено как 1) локализация сцены в пространстве признаков и 2) последующее соотнесение координат полученной точки с записью в памяти, где каждый типичный ландшафт также «оцифрован» (т.е. обозначен координатами).

#### *Различение и опознание сложных объектов*

В предыдущем параграфе мы убедились в том, что опознание ландшафта способно повлиять на категоризацию объектов, наполняющих этот ландшафт (рис. 1). Однако в восприятии сцен возможен и в некотором смысле обратный ход, когда только конкретные объекты могут прояснить, что за сцена перед нами. Очевидно, что информации, содержащейся в глобальных признаках, зачастую недостаточно для тонкой дифференциации сцен, на которую также способен человек буквально с одного взгляда. Например, мы можем видеть типичный ландшафт комнаты: закрытый, с небольшой средней глубиной и т.п. При этом довольно легко мы опознаем эту комнату как спальню, кухню или кабинет. Но ведь пространственный конверт у этих помещений одинаковый. Значит, дело в объектах, которые наполняют этот конверт. Объекты, которые помогают категоризовать сцену (просто за счет своей уместности в этой сцене), мы будем называть диагностическими объектами. Например, кровать будет диагностическим объектом, который поможет распознать спальню в ландшафте комнаты, а чашки и тарелки помогут распознать кухню. Однако для того, чтобы с одного взгляда отличить спальню от кухни, объекты также должны быть распознаны с одного взгляда.

Способность различать несколько разнотипных и сложных объектов с одного взгляда особенно поражает в восприятии *сюжетных сцен*, т.е. таких, где запечатлено динамическое событие или действие. Обычно в сюжет вовлечены несколько объектов (живых или неживых). Например, за 100-200 мс наблюдатель способен увидеть на фотографии ребенка, играющего в мяч с собакой. В этом сюжете фигурируют как минимум три объекта: ребенок, собака и мяч, и все они должны быть опознаны за столь короткое время. Более того, довольно тонкая детализация образа нужна для того, чтобы определить, что перед нами ребенок, а не взрослый (например, необходимо оценить относительный размер человеческой фигуры в координатах ландшафта). Как же возможно столь быстрое опознание сложных объектов, заведомо состоящих из множества элементов и признаков?

Быстрота, с которой извлекается информация о сложных объектах, порождает несколько важных вопросов. Во-первых, как зрительной системе удастся выделить объекты из весьма пестрого фона, каковым являются элементы ландшафта, а иногда и другие объекты? Во-вторых, насколько детальным будет образ самого объекта и его свойств? В более радикальной формулировке это вопрос о том, действительно ли сложный объект воспринимается как целое или мы «выхватываем» лишь отдельные простые части, по которым наше сознание достраивает все остальное.

*Выделение объекта из фона.* Как правило, исследователи рассматривают два класса факторов, определяющих легкость выделения какого либо объекта из фона или из других объектов (которые в этом случае называют дистракторами).

Первый класс факторов – это соотношение некоторых физических особенностей фигуры и фона, определяющее меру локального контраста вдоль границы между фигурой и фоном (Wolfe, 1994). Участок зрительной сцены, сильно отличающийся от соседних участков, как правило, будет обладать *перцептивной очевидностью (saliency)* и автоматически привлекать к себе внимание, что является необходимым условием для быстрого и легкого выделения объекта. Это легко видеть на двух примерах. На рис. 2а показан пример *легкой сегментации текстуры (effortless texture segmentation)*. Мелкие объекты, сгруппированные по гештальтистским принципам близости и сходства, образуют более крупные, которые легко воспринимаются на текстурированном фоне, образованном другими мелкими объектами, также похожими между собой. Второй пример – *эффект «выскакивания» (pop-out effect)* – показан на рис. 2б. Буквально с одного взгляда можно легко увидеть две черные линии, несмотря на разнообразие размеров и ориентаций всех остальных линий. Нетрудно также видеть, что выделение перцептивно очевидных сегментов (потенциальных объектов) идет параллельно по всему зрительному полю: например, глядя на фигуру, образованную вертикальными линиями на рис. 2а, краем глаза Вы уже можете различить два «выскакивающих» черных участка на рис. 2б. На рис. 2в и 2г представлены эффекты сегментации текстуры и «выскакивания» в естественных сценах.

Однако, как правило, для быстрого выделения объекта по принципу перцептивной очевидности фигура должна отличаться от окружающего фона, по крайней мере, одним явным признаком (Bergen & Julesz, 1983; Treisman & Gelade, 1980). К таким признакам могут относиться цвет, размер, пространственная ориентация, скорость движения и др., а при наличии трехмерного стереоизображения – и признаки глубины (Nakayama & Silverman, 1986). Если признаки или элементы фигуры не уникальны, то легкого выделения объекта не происходит. Так, на рис. 2д увидеть четкие контуры квадрата уже не так просто, хотя его текстура состоит только из вертикальных черных и горизонтальных белых линий, а текстура фона – только из

вертикальных белых и горизонтальных черных. Аналогичным образом, единственная вертикальная черная линия на рис. 2е не «выскакивает» из горизонтальных черных и вертикальных белых.

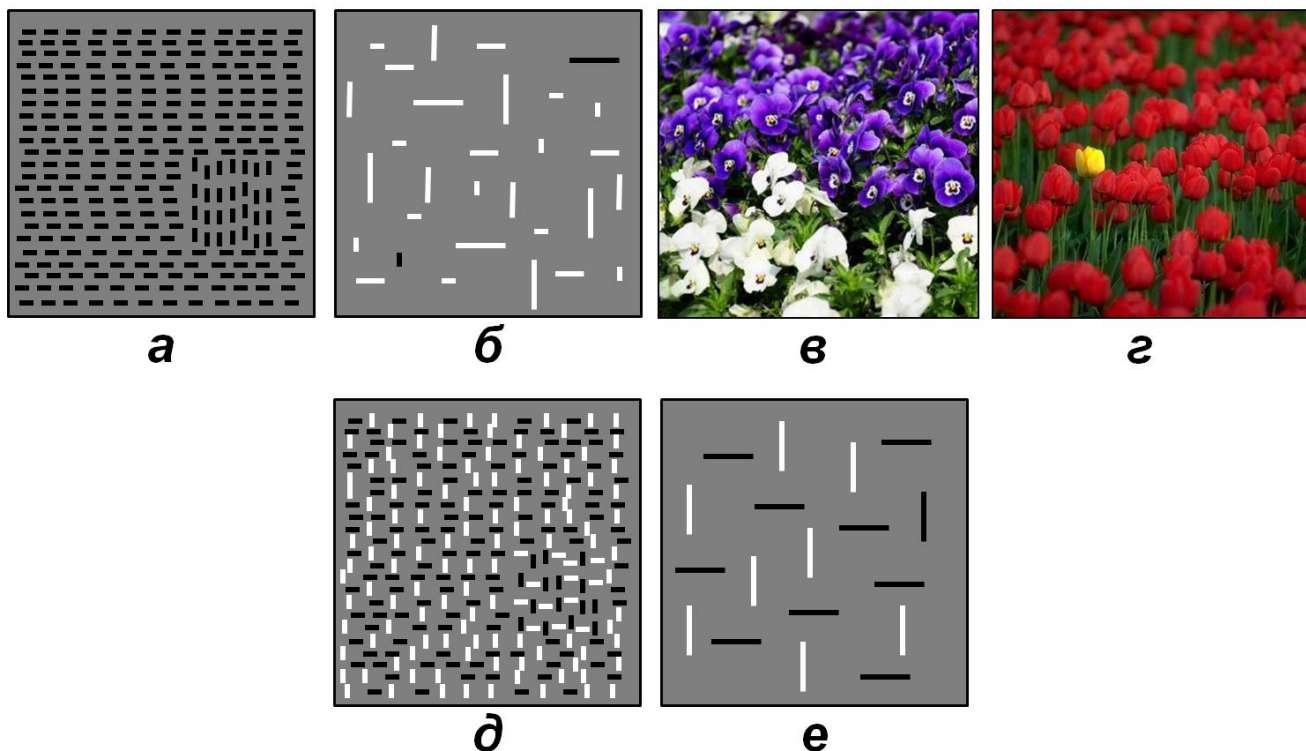


Рисунок 2. Выделение объекта из фона: а) легкая сегментация текстуры по очевидному признаку – ориентации линий; б) эффект «выскакивания» уникальных объектов по признаку цвета; в-г) сегментация текстуры и эффект «выскакивания» в естественных сценах; д) сложная сегментация текстуры, заданной сочетанием ориентации и цвета; е) отсутствие «выскакивания» уникального объекта, заданного сочетанием ориентации и цвета.

Второй класс факторов, влияющие на легкость выделения объектов – это различные эффекты, связанные с установкой, ожиданиями, текущими целями, устойчивыми интересами и т.п. Например, в потоке быстро сменяющих друг друга фотографий испытуемый гораздо точнее определит, был ли ему показан велосипед, если этот вопрос ему задать до начала показа, а не после (Potter, 1976). Также в целой серии экспериментов было показано, что систематически повторяющиеся уникальные объекты (заданные как одним особым признаком, так и их сложным сочетанием), в конечном счете, начинают «выскакивать» или, по крайней мере, находиться быстрее (Kristjansson et al. 2002; Maljkovic & Nakayama, 1994; Schneider & Shiffrin, 1974). Предположительно, это происходит благодаря активации некоего образца в памяти, который ведет внимание к объектам, более всего напоминающим этот образец (Maljkovic & Nakayama, 1994; Wolfe, 1994).

*Детализация образа объекта.* Читатели, наверняка, легко вспомнят детскую головоломку из серии «Найди 10 отличий», где на двух очень похожих картинках нужно найти

ряд деталей, которыми эти картинки различаются. Несмотря на то, что обе картинки находятся прямо перед глазами, найти очевидные различия с одного взгляда не получается; для этого необходимо сравнивать между собой объекты по отдельности, последовательно направляя на них внимание. Эта простая демонстрация эффекта, названного психологами «слепотой к изменению» (*change blindness*), свидетельствует о том, что при глобальном восприятии сцены с одного взгляда образы объектов остаются весьма неполными, хотя и достаточными для почти безошибочного опознания.

Для объяснения того, каким образом недетализированный образ объекта, тем не менее, легко опознается, И. Бидерман (1987) выдвинул теорию распознавания по компонентам. Согласно теории, основой распознавания объектов являются геометрические характеристики их контуров: кривизна, параллельность линий, симметрия и др. Различные сочетания этих характеристик образуют простые трехмерные геометрические формы, такие как сфера, конус, цилиндр, параллелепипед и т.п. (всего порядка 36). Эти простые формы, названные *геонами*, в разных пространственных сочетаниях дают огромное многообразие схематичных образов объектов, которые могут быть опознаны без лишних деталей. При мгновенном восприятии сцены разнообразные сложные вариации форм реальных объектов приводятся к схематичному геонному шаблону по гештальтскому принципу *прегнантности*. Как подчеркивает Бидерман, его теория позволяет объяснить константность многих свойств воспринимаемых объектов, таких как размер и форма. Даже при изменении видимого положения объектов пространственные характеристики их частей меняются, но идентичность геонов и их взаиморасположение остаются постоянными. Вместе с тем, как отмечает Х. Шиффман (2003), теория геонов не способна в полной мере объяснить разнообразие форм, которое способен различить человек с одного взгляда. Например, геонные схемы птицы и самолета будут практически одинаковы, так же как и геонные схемы почти всех четвероногих животных.

Альтернативный подход предлагает Э. Трейсман (2006) в рамках своей теории интеграции признаков. По ее мнению, в ходе перцептивного развития у животных и человека формируются системы довольно разнообразных детекторов для отдельных частей объектов, которые начинают со временем функционировать, подобно базовым сенсорным признакам (цвету, яркости, ориентации и т.п.). Например, гладкий компактный контур, наличие глаз, крыльев, головы, конечностей являются признаками животных, а контур, состоящий из большого количества прямых линий, или наличие колес являются признаками транспортных средств. Эти сложные признаки подчиняются законам функционирования элементарных признаков. Они могут быть автоматически выделены из любой части зрительного поля, что соответствует эффекту «выскакивания», на их основе может производиться сегментация текстуры. По физиологическим данным, эти признаки регистрируются на уровне локальных



нейронных ансамблей. Эта особенность также характерна именно для признаков, в то время как репрезентация целостного объекта, состоящего из этих признаков, требует вовлечения распределенной нейронной сети, в которой, предположительно происходит пространственное связывание признаков.

Существование подобных прижизненно сформированных признаков позволяет с легкостью судить о наличии того или иного сложного объекта в зрительной сцене, не фокусируя внимание на этом объекте. В отличие от понятия геонов, которое предлагает Бидерман, трейсмановские сложные признаки могут основываться не только на геометрических формах, но также на других простых признаках, таких как цвет, размер и др. Это дает несравненно больше отличительных особенностей, по которым можно отличить один сходный объект от другого. Например, птицы и летучие мыши обладают крыльями, но этих животных можно относительно легко различить с одного взгляда, поскольку текстуры крыльев различны. Несомненно, чем более отдаленными являются классы объектов, тем больше вероятность встретить отдельные признаки у отдельных представителей этих классов. Вероятно, по этой причине грубая категоризация (например, насекомое или млекопитающее, мебель или транспорт) при кратком предъявлении происходит гораздо легче и успешнее, чем категоризация близких объектов (кошка или собака, яблоко или абрикос).

Теория Трейсман также не требует, чтобы различные части объекта образовывали правильную пространственную конфигурацию для его опознания (что, по Бидерману, является практически единственным шансом для образования узнаваемой геонной структуры). В экспериментах, которые описывает Трейсман, испытуемые в режиме быстрой смены слайдов легко обнаруживали присутствие животных, но затруднялись с локализацией этих животных (т.е. не могли правильно оценить, было ли животное в правой, левой или центральной части сцены). Кроме того, в подобных экспериментах часто наблюдаются *иллюзорные соединения* (*illusory conjunctions*), когда признаки, обнаруженные на одних изображениях, как бы мигрируют в другие. Например, полосатая текстура дивана, показанная на одном слайде, может «перескочить» на последующий или даже предыдущий слайд, изображающий, к примеру, белую лошадь. Результатом такой миграции признака становится ошибочное, но вполне явственное восприятие полосатой зебры. По мнению Трейсман, эти факты свидетельствуют о том, что зрительная система сравнительно легко выделяет признаки сложных объектов, и эти признаки сами по себе могут использоваться для опознания объектов хотя бы в первом приближении. Однако в условиях широко распределенного внимания (которое обычно и связывают с мгновенным восприятием сути сцены) эти признаки свободно «плавают» в перцептивном пространстве и времени, т.е. не привязаны четко ни к определенному месту, ни к определенному объекту. Только локально сфокусированное на определенной части сцены

внимание правильным образом связывает эти признаки и дает жизнь подлинной сознательной репрезентации этого объекта. Однако сам процесс связывания довольно медленный и требует более активного, тщательного и длительного рассматривания.

*Оценка свойств и признаков множественных объектов: репрезентация ансамблей*

В предыдущем разделе мы упомянули о сегментации текстуры, когда сходные объекты при наличии пространственной близости образуют целостную фигуру, которую легко увидеть на фоне других объектов, образующих иную текстуру. Однако часто сходные объекты бывают довольно широко разбросаны в пространстве и даже перемежаются непохожими объектами, что не способствует группировке. Как же возможно в этом случае мгновенное схватывание всего этого множества?

Интенсивные исследования в этой области начались чуть более десяти лет назад под общим термином «*статистическая репрезентация множеств, или ансамблей*». В основе подхода лежат эксперименты Д. Ариэли (2001), в которых было показано, что испытуемые способны достаточно точно оценивать средний размер набора «разнокалиберных» кругов, предъявленных на короткое время; при этом точность оценки среднего превышала точность оценки размера отдельных членов множества, осуществленной по памяти. По мнению Ариэли, этот результат свидетельствует о том, что зрительная система оценивает размеры индивидуальных объектов, но не сохраняет их в виде отдельных записей (что задействовало бы огромный ресурс внимания и рабочей памяти), а строит на их основе более компактную репрезентацию.

Альтернативную версию того, как происходит усреднение, предлагают К. Мышек и Д. Саймонс (2008). По их мнению, для успешного усреднения вообще нет необходимости анализировать все объекты. Достаточно сфокусировать внимание на нескольких объектах (например, самом большом и самом маленьком или на компактной группе объектов вблизи точки фиксации), чтобы осуществить усреднение и экстраполировать полученную оценку на весь остальной ансамбль. Однако, в противовес этой гипотезе, было установлено, что точность и скорость оценки среднего увеличивается по мере роста объема множества (Chong et al., 2008; Robitaille & Harris, 2011). Этот эффект «статистической мощности» служит вполне убедительным доказательством того, что усреднение происходит по всем объектам сразу, а не выборочно, как предполагали Мышек и Саймонс.

В других исследованиях было показано, что подобное усреднение с легкостью осуществляется и для ряда других признаков: пространственной ориентации (Dakin & Watt, 1997), яркости (Bauer, 2009), скорости движения (Watamaniuk & Duchon, 1992) и даже эмоциональных выражений лиц в толпе (Haberman & Whitney, 2007). Информация о среднем извлекается, по разным подсчетам, примерно за 50-200 мс (Chong & Treisman, 2003; Whiting &

Oriet, 2011), что вполне подпадает под характеристику «с одного взгляда». Помимо оценки среднего, зрительная система с легкостью извлекает и другие виды статистик из ансамблей множественных объектов, такие как численность выборки или дисперсия (по: Alvarez, 2011). Одним словом, возможности, которые предоставляет нам репрезентация ансамблей при восприятии сложных сцен, поистине грандиозны.

*Сегментация подмножеств.* Представьте себе, что Вы собираете ягоды с малинового куста. Можно ли, кинув лишь беглый взгляд на куст, определить, много ли на нем красных спелых ягод и, следовательно, прикинуть, насколько урожайным окажется этот куст? Зная из предыдущей части о том, как осуществляется мгновенная репрезентация множественных объектов, мы можем предположить, что эта задача может быть решена путем оценки средней степени красноты «ягодного ансамбля». Проблема заключается лишь в том, что, помимо ягод, на кусте растут зеленые листья. Если рассматривать ягоды и листья как единый ансамбль, то оценка красноты ягод окажется неверной, поскольку гораздо более многочисленная масса листьев значительно сдвинет среднее в сторону зеленого цвета. Следовательно, для решения подобной проблемы зрительная система должна обладать способностью к разделению перемешанных между собой ансамблей. Эту способность мы будем называть *сегментацией подмножеств* (по аналогии с сегментацией текстуры).

Исследования последних лет показывают, что легкая сегментация подмножеств действительно возможна до некоторой степени. Так, С. Чонг и Э. Трейсман (2005) предъявляли своим испытуемым наборы, состоящие из кругов одного или двух цветов, и просили оценить среднее в одном из произвольно выбранных цветных подмножеств. В результате они обнаружили, что точность усреднения в «двухцветном» условии была не хуже, чем в «одноцветном». При этом было неважно, сгруппированы ли круги одного цвета в пространстве или перемешаны с кругами другого цвета. В других экспериментах было показано, что люди способны довольно точно оценивать абсолютную численность (Halberda et al., 2006) или пропорцию (Treisman, 2006) объектов определенного цвета среди объектов двух других цветов, т.е. считать/рассчитывать статистику одновременно по трем подмножествам.

Впрочем, похоже, что способность к мгновенной сегментации подмножеств ограничена. В этом легко убедиться, взглянув на рис. 3. Попробуйте, кинув беглый взгляд, определить, сколько разных видов ягод (т.е. подмножеств) Вы видите на каждой из трех фотографий. На рис. 3а эта задача кажется элементарной. Немного сложнее определить количество подмножеств и на рис. 3б. Однако на рис. 3в задача становится гораздо сложнее, и Вам, скорее всего, понадобится задержать взгляд на этой фотографии и попросту пересчитать количество разнотипных ягод. По экспериментальным оценкам, легкая сегментация возможна лишь для

двух (или, по самым оптимистичным оценкам, трех) перемешанных между собой подмножеств (Utochkin, under review).



**а**



**б**



**в**

*Рисунок 3.* Сегментация подмножеств. С одного взгляда попробуйте определить, сколько разных видов ягод представлено на каждой фотографии. Скорее всего, на фотографиях а) и б) такая сегментация подмножеств пройдет легко; на фотографии в) задача покажется более сложной.

Еще одно фундаментальное ограничение сегментации подмножеств, как указывает Трейсман (2006), заключается в том, что она, как и сегментация текстуры, может происходить только на основе явных признаков, но не их сочетаний. В своих экспериментах Трейсман показывала испытуемым наборы букв Т, Х и О трех разных цветов и просила их определить относительное количество либо определенных букв (например, только О), либо букв определенного цвета (например, только зеленых), либо их сочетания (например, только зеленых букв О). Испытуемые хорошо справлялись с оценкой количества объектов, заданных одним из двух признаков, но крайне плохо – с оценкой количества сочетаний. По мнению Трейсман, этот результат вполне закономерен. Напомним, что, согласно теории интеграции признаков,

развиваемой данным автором, для оперирования сочетаниями признаков необходимо связывание этих признаков в репрезентациях объектов, а это возможно только при последовательной фокусировке внимания на каждом занятом участке зрительного поля.

В данном обзоре мы вкратце рассмотрели основные феномены мгновенного восприятия сложных сцен и подходы к пониманию природы этих феноменов. Основной вывод, который можно сделать из обзора, заключается в том, что уже на ранних этапах зрительного анализа нам доступно довольно большое количество информации, необходимой для осмысленного восприятия окружающей нас среды и событий, происходящих в ней. Несомненно, способность к мгновенному «схватыванию» сути происходящего имеет важное биологическое значение, позволяя человеку и многим видам животных ориентироваться в сложной и быстро меняющейся среде. Однако не менее важен и другой вывод. При всей кажущейся полноте такого мгновенного образа, он отражает лишь глобальные свойства сцены и объектов в ней. Многие важные аспекты сцены, связанные, прежде всего с характеристиками конкретных объектов, могут оставаться неосознанными (мы убедились в этом на примере головоломки «Найди 10 отличий»). В этой связи трудно переоценить роль активного зрительного восприятия, связанного с тщательным рассматриванием зрительной сцены и опосредованного вниманием.

#### Литература

*Шиффман Х.* Ощущение и восприятие. 5 издание. М., Питер, 2003.

*Alvarez G.A.* Representing multiple objects as an ensemble enhances visual cognition // *Trends in Cognitive Science*. 2011. V. 15. P. 122-131.

*Ariely D.* Seeing sets: Representation by statistical properties // *Psychological Science*. 2001. V. 12. P. 157–162

*Bauer B.* Does Stevens's power law for brightness extend to perceptual brightness averaging? // *Psychological Record*. 2009. V. 59. P. 171–186.

*Bergen J.R., Julesz B.* Parallel versus serial processing in rapid pattern discrimination // *Nature*. 1983. V. 303. P. 696-698.

*Biederman I.* Recognition-by-Components: A theory of human image understanding // *Psychological Review*. 1987. V. 94. P. 115-147.

*Chong S.C., Treisman A.M.* Representation of statistical properties // *Vision Research*. 2003. V. 43. P. 393–404.

*Chong S.C., Joo S.J., Emmanouil T.-A., Treisman A.* Statistical processing: Not so implausible after all // *Perception and Psychophysics*. 2008. V. 70. P. 1327-1334.

*Chong, S.C., Treisman A.M.* Statistical processing: Computing average size in perceptual groups // *Vision Research*. 2005. V. 45. P. 891-900.

*Dakin S.C., Watt R.J.* The computation of orientation statistics from visual texture // *Vision Research*. 1997. V. 37. P. 3181–3192.

*Greene M.R., Oliva A.* The briefest of glances: The time course of natural scene understanding // *Psychological Science*. 2009. V. 20. P. 464-472.

*Haberman J., Whitney D.* Rapid extraction of mean emotion and gender from sets of faces // *Current Biology*. 2007. V. 17, P. R751–R753.

*Halberda J., Sires S.F., Feigenson L.* Multiple spatially overlapping sets can be enumerated in parallel // *Psychological Science*. 2006. V. 17. P. 572-576.

*Intraub H.* Rapid conceptual identification of sequentially presented pictures // *Journal of Experimental Psychology: Human Perception and Performance*. 1981. V. 7. P. 604-610.

*Kristjansson A., Wang D.L., Nakayama K.* The role of priming in conjunctive visual search // *Cognition*. 2002. V. 85. P. 37-52.

*Luck S.J., Vogel E.K.* The capacity of visual working memory for features and conjunctions // *Nature*. 1997. V. 390. P. 279-281.

*Maljkovic V., Nakayama K.* Priming of popout: I. Role of features // *Memory and Cognition*. 1994. V. 22. P. 657-672.

*Myczek K., Simons D.J.* Better than average: alternatives to statistical summary representations for rapid judgments of average size // *Perception and Psychophysics*. V. 70. P. 772–788.

*Nakayama K., Silverman G.H.* Serial and parallel processing of visual feature conjunctions // *Nature*. 1986. V. 320. P. 264-265.

*Navon D.* Forest before trees: The precedence of global features in visual perception // *Cognitive Psychology*. 1977. V. 9. P. 353-383.

*Oliva A., Torralba A.* Building the gist of a scene: The role of global image features in recognition // *Progress in Brain Research*. 2006. V. 155. P. 23-36

*Oliva A., Torralba A.* Modeling the shape of the scene: A holistic representation of the spatial envelope // *International Journal in Computer Vision*. 2001. V. 42. P. 145-175.

*Potter M.C.* Meaning in visual search // *Science*. 1975. V. 187. P. 965-966.

*Potter M.C.* Short-term conceptual memory for pictures // *Journal of Experimental Psychology: Human Learning and Memory*. 1976. V. 2. P. 509-522.

*Rensink R.A.* The dynamic representation of scenes // *Visual Cognition*. 2000. V. 7. P. 17-42.

*Robitaille N., Harris I.M.* When more is less: Extraction of summary statistics benefits from larger sets // *Journal of Vision*. 2011. V. 11. P. 1-8.

- Schneider W., Shiffrin R.M.* Controlled and automatic human information processing: I. Detection, search, and attention // *Psychological Review*. 1977. V. 84. P. 1–66.
- Torralba A., Oliva A.* Depth estimation from image structure // *IEEE Pattern Analysis and Machine Intelligence*. 2002. V. 24. P. 1226-1238.
- Treisman A.* How the deployment of attention determines what we see // *Visual Cognition*. 2006. V. 14. P. 411-443
- Treisman A.M., Gelade G.* A feature integration theory of attention // *Cognitive Psychology*. 1980. V. 12. P. 97-136.
- Utochkin I.S.* Seeing variety: Subset segmentation in human vision // *Vision Research*. Under review.
- Watamaniuk S.N.J., Duchon A.* The human visual system averages speed information // *Vision Research*. 1992. V. 32. P. 931–941.
- Whiting B.F., Oriet C.* Rapid averaging? Not so fast! // *Psychonomic Bulletin and Review*. 2011. V. 18. P. 484 – 489.
- Wolfe J.M.* Guided Search 2.0: A revised model of visual search // *Psychonomic Bulletin and Review*. 1994. V. 1. P. 202-238.