

## КОЛИЧЕСТВЕННОЕ ОПИСАНИЕ ЗАКОНОВ ПЕРЦЕПТИВНОГО ГРУППИРОВАНИЯ С ПОМОЩЬЮ ПРИНЦИПА РЕПРЕЗЕНТАЦИОННОЙ МИНИМАЛЬНОЙ ДЛИНЫ ОПИСАНИЯ

© 2008 г. А. С. Потапов\*, канд. техн. наук; В. Г. Петроченко\*\*

\* НПК “Государственный оптический институт им. С.И. Вавилова”, Санкт-Петербург

\*\* Санкт-Петербургский государственный университет информационных технологий, механики и оптики, Санкт-Петербург

Рассмотрена проблема количественного описания законов перцептивного группирования, разработанных в гештальт-психологии, и их использования в области компьютерного зрения. Для унифицированного описания законов гештальта предложен принцип репрезентационной минимальной длины описания. Проведены психофизические эксперименты, в которых установлено, что вероятность выделения стимула человеком зависит преимущественно только от информативности стимула, выражающейся в уменьшении длины описания (в рамках заданного представления) элементов зрительной сцены при описании стимула отдельно от элементов поля. Построена и экспериментально проверена компьютерная модель, выполняющая группирование структурных элементов изображения на основе информационного критерия и представления, воплощающего некоторые из законов гештальта.

Коды OCIS: 3000.30210.

*Поступила в редакцию 29.02.2008.*

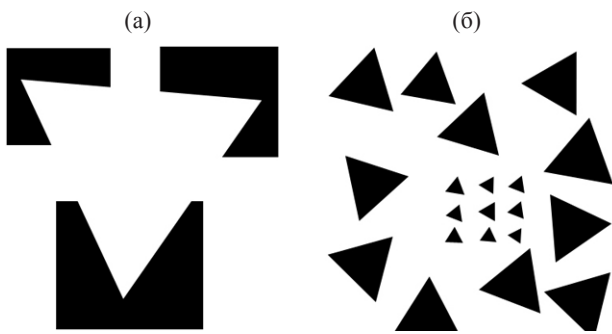
### Введение

Одним из классических вопросов в исследовании проблем как компьютерного зрения, так и зрительного восприятия человека является вопрос о том, как достигается целостное восприятие изображений из локальной информации, содержащейся в отдельных пикселах [1]. В психологии восприятия этот вопрос исследовался в рамках теории гештальта, в которой был обнаружен ряд “законов” перцептивного группирования элементов поля зрительной системой человека. Однако эти законы не носили количественный или тем более алгоритмический характер, поэтому долгое время в области компьютерного зрения им уделялось мало внимания [2].

С недавнего времени стали предприниматься попытки объединить результаты, полученные в теории гештальта и в области компьютерного зрения, в частности, в рамках теоретико-информационного [1], нейросетевого [2] и статистического [3] подходов. Для гештальт-психологии это может дать возможность количественного уточнения законов перцептивного группирования и возможность дополнительной экспериментальной проверки посредством компьютерного моделирования, а в области компьютерного зрения – послужить развитию новых идей и биологически обоснованных алгоритмов анализа изображений.

Одним из перспективных подходов как к проблеме количественного выражения законов гештальта [1], так и к проблеме интерпретации изображений в системах компьютерного зрения [4] является теоретико-информационный подход на основе принципа минимальной длины описания (МДО). Согласно этому принципу наилучшей моделью имеющихся данных является модель, для которой достигается минимум суммы длины описания данных с помощью модели и длины описания самой модели [5].

Однако этот принцип не позволяет строго вывести критерии перцептивного группирования без привлечения дополнительных эвристических соображений. В данной работе предлагается расширение этого принципа, учитывающее понятие представления изображений, – принцип репрезентационной минимальной длины описания (РМДО). Ставятся психофизические эксперименты, показывающие, что эффективность группирования, выполняемого зрительной системой человека, зависит преимущественно от информативности стимула в рамках заданного представления (по крайней мере, для использованных в эксперименте типов стимулов). Таким образом, проблема алгоритмизации законов перцептивного группирования оказывается связанной не столько с поиском адекватных количественных критериев группирования, сколько с поиском оптимальных представлений изображений.



**Рис. 1.** Примеры изображений, иллюстрирующих законы группирования по “хорошему продолжению” (а) и подобию (б).

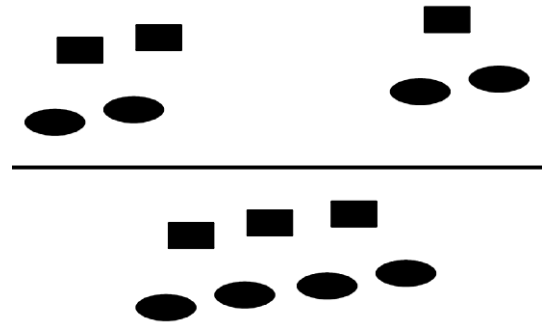
### Основные положения теории гештальта

Теория гештальта была разработана в первой половине XX века как попытка объяснить особенности человеческого восприятия. В то время об устройстве мозга было известно мало и гештальт-психологи пытались дать термодинамическую интерпретацию механизмов восприятия, полагая, что “перцептивное поле” под действием некоторых сил реорганизуется в состояние с минимальной энергией (т. е. обладающее наибольшей симметрией). Хотя такая интерпретация и не получила подтверждения, она позволила накопить большой эмпирический материал, в частности, по изучению зрительного восприятия и сформулировать ряд “законов” перцептивного группирования.

Примером закона группирования может служить закон группирования по подобию, согласно которому сходные или подобные элементы зрительного поля имеют тенденции объединяться в группы. Другие законы группирования – это группирование по близости, “хорошему продолжению” (объединяются элементы, располагающиеся на некоторой регулярной кривой, например, линии), замкнутости формы и т. д. На рис. 1 представлены типичные примеры, показывающие действие законов гештальта.

Указанные законы – результат обобщения наблюдаемых особенностей человеческого восприятия. Как правило, каждый закон относится к одному признаку: форме, размеру, направлению и т. д. Однако законы гештальта недостаточно конкретны, в связи с чем не ясны границы, в которых действует тот или иной закон, и возникает проблема разрешения конфликтов между законами [1]. Когда применимы несколько законов, которые дают разный результат группирования, теория гештальта не может надежно предсказать, как будет воспринято изображение человеком (см. рис. 2).

Более того, в рамках теории гештальта остается непонятным, почему именно эти законы использу-



**Рис. 2.** Пример конфликта законов группирования по сходству и “хорошему продолжению”; теория гештальта не дает предсказания, на каком расстоянии между элементами происходит перегруппирование.

ются в процессе восприятия, и являются ли они оптимальными и обязательными при реализации систем компьютерного зрения. Для ответа на эти вопросы рассмотрим принцип репрезентационной минимальной длины описания.

### Принцип репрезентационной минимальной длины описания

Согласно принципу МДО моделью  $m^*$ , наилучшим образом описывающей некоторый набор данных  $F$ , является модель, которая позволяет минимизировать сумму длины описания модели  $L(m)$  и длины описания данных в рамках модели  $L(f|m)$ , –

$$m^* = \operatorname{argmin}(L(m) + L(f|m)). \quad (1)$$

В качестве длин описания  $L$  в теоретических работах выступает алгоритмическая сложность строк символов, соответствующих данным  $f$  и модели  $m$ . При этом в качестве пространства моделей выступает множество программ для машины Тьюринга, а данные  $f$  содержат всю имеющуюся информацию.

При решении задач анализа изображений требуется их независимая интерпретация, причем при описании каждого изображения используется некоторая априорная информация. В этих условиях необходимо уточнить принцип МДО как принцип репрезентационной МДО, включающий два положения, рассмотренные ниже [6].

1. Моделью  $m^*$ , наилучшим образом описывающей некоторый набор данных  $f$  в рамках представления  $S$ , является модель, минимизирующая сумму

$$L_S(f, m) = L_S(m) + L_S(f|m). \quad (2)$$

Здесь  $L_S$  – длина описания в рамках данного представления, которая формально может быть определена как  $L_S(x) = L(Sx)$ , где  $L(Sx)$  – алгоритмическая сложность конкатенации строк  $S$  и  $x$ .

2. Лучшим для данной выборки изображений  $F = \{f_1, \dots, f_n\}$  является представление  $S$ , для которого минимизируется сумма длины представления  $L(S)$  и суммы длин описаний изображений

$$\sum_{i=1}^n MDL_S(f_i),$$

где  $MDL_S(f_i) = \min_m L_S(f, m)$ .

Представление  $S$  – это информация, общая для всех изображений выборки  $F$ . Таким образом, выбор оптимального представления изображений – это эмпирическая задача, решение которой может направляться средней по некоторой выборке длиной описания.

Рядом исследователей отмечается [1], что законы гештальта могут быть выражены в терминах длины описания. Однако, согласно принципу РМДО, для построения информационного критерия, количественно выражающего законы гештальта и корректно предсказывающего результаты восприятия конкретных стимулов, необходимо построение адекватного представления изображений.

### Психофизический эксперимент

Для определения характеристик перцептивного группирования в зрительной системе нами был проведен следующий эксперимент. Испытуемым предъявлялся стимул (см. рис. 3), а затем предъявлялись два изображения (см. рис. 4), одно из которых содержало стимул, а другое – нет (т. е. выступало в роли дистрактора). Каждое из изображений предъявлялось на 1 с.

Стимулы варьировались по числу элементов, по среднеквадратическому отклонению элементов от положения на прямой и по дисперсии ориентации. Для каждого набора параметров формирования стимулов проводилась серия экспериментов с испытуемыми, в результате которой устанавливался сред-

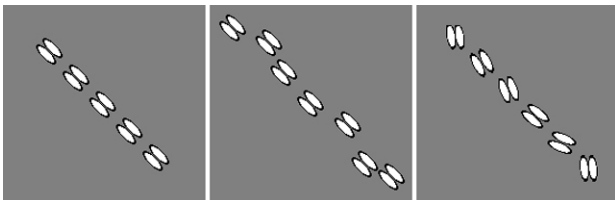


Рис. 3. Примеры стимулов с разными значениями параметров (числа элементов, ошибки положения, ошибки ориентации).

ний процент ошибок, допускаемых людьми при выборе изображения, содержащего стимул.

Рассмотрим вопрос о представлении информации об элементах на изображениях такого типа. Каждый элемент описывается набором параметров  $x_i, y_i, \varphi_i$ , где  $x_i, y_i$  – координаты  $i$ -го элемента,  $\varphi_i$  – угол его ориентации. Если элементы не разбиты на группы, то параметры каждого из них описываются независимым образом. Координаты некоторого элемента могут быть описаны относительно положения ближайшего к нему элемента, для чего необходимо указать его номер (для этого требуется  $\log_2 N$  бит, где  $N$  – общее число элементов) и относительные координаты (для их указания требуется в среднем  $\log_2 \sigma_r + C_1$  бит, где  $\sigma_r$  – среднее расстояние между элементами, а константа  $C_1$  определяет точность описания координат; также требуется описать азимут направления на ближайший элемент, однако в силу изотропности распределения азимутов этот компонент описания не учитывается). Пусть углы ориентации элементов поля равномерно распределены в диапазоне  $[-\varphi, \varphi]$ , тогда для их описания потребуется  $N(\log_2 \varphi + C_2)$  бит ( $C_2$  – константа, определяемая точностью описания ориентации). Для описания элементов поля, не разбитых на группы, требуется

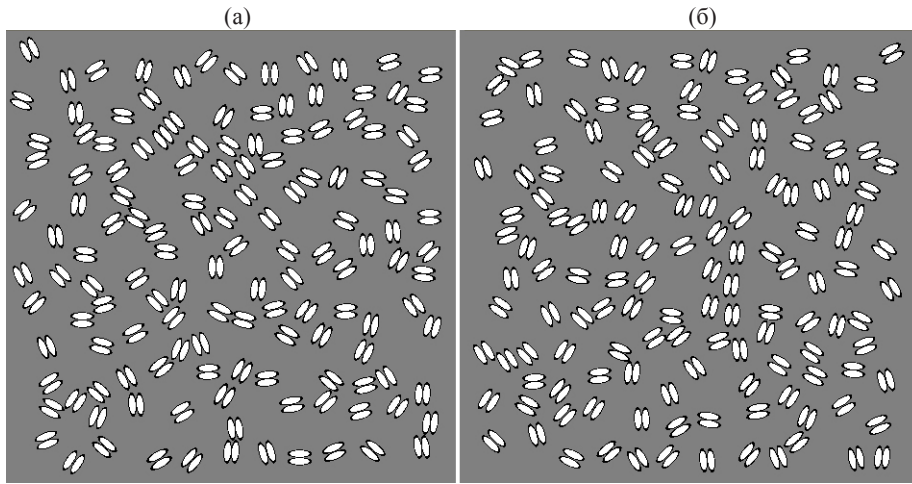
$$L_0 = N(\log_2 N + \log_2 \sigma_r + \log_2 \varphi + C_1 + C_2) \text{ бит.} \quad (3)$$

Из элементов поля выделим группу из  $n$  элементов. Для описания оставшихся элементов потребуется  $L = (N - n)(\log_2 N + \log_2 \sigma_r + \log_2 \varphi + C_1 + C_2)$  бит. Пусть априори известно, что эти элементы располагаются на некоторой кривой с определенным шагом (как было в эксперименте, см. рис. 3). Тогда можно предложить следующее представление информации для описания данной группы. В рамках данного представления указывается, какие именно элементы выделены в группу, для чего требуется  $n \log_2 N$  бит. Координаты элементов группы описываются через их отклонение от кривой, и при среднем отклонении  $\sigma_c$  для этого потребуется  $n(\log_2 \sigma_c + C_1)$  бит. Пусть углы ориентации элементов группы равномерно распределены в диапазоне  $[-\psi, \psi]$ . Тогда для их описания потребуется  $n(\log_2 \psi + C_2)$  бит. Для описания всех элементов при выделенной группе требуется

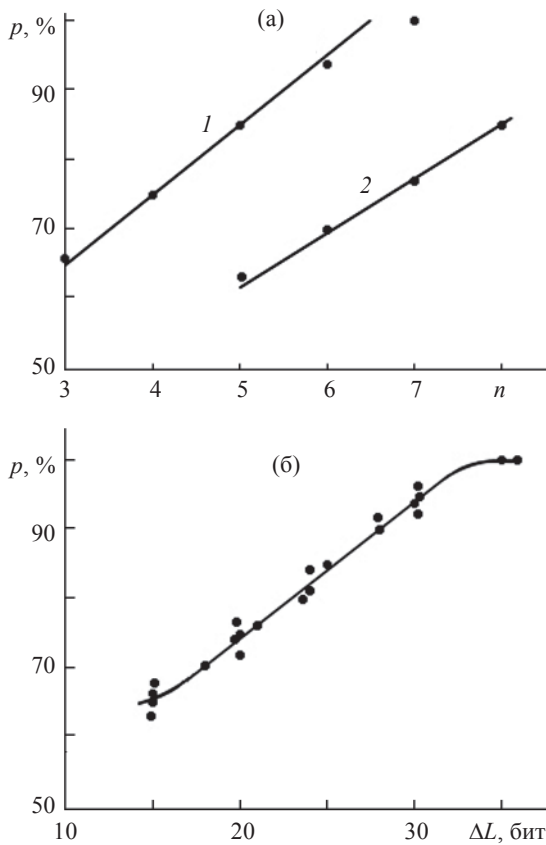
$$L_1 = (N - n)(\log_2 N + \log_2 \sigma_r + \log_2 \varphi + C_1 + C_2) + n(\log_2 N + \log_2 \sigma_c + \log_2 \psi + C_1 + C_2) \text{ бит.} \quad (4)$$

В рамках данного представления выигрыш в длине описания при выделении группы составит

$$\Delta L = L_0 - L_1 = n \left( \log_2 \frac{\sigma_r}{\sigma_c} + \log_2 \frac{\varphi}{\psi} \right) \text{ бит.} \quad (5)$$



**Рис. 4.** Примеры предъявляемых в ходе эксперимента изображений, одно из которых (а) содержит искомый стимул, а другое (б) – нет.



**Рис. 5.** Зависимость процента ошибок  $p$  от числа элементов  $n$  в стимуле (а) и от информативности  $\Delta L$  стимула (б) при различных значениях других параметров стимула. 1 –  $\psi = \pi/4$ , 2 –  $\psi = \pi/2$ .

Согласно принципу репрезентационной МДО выигрыш в длине описания является основным критерием выполнения группирования. На основе дан-

ных экспериментов была определена зависимость процента ошибок, допускаемых человеком при обнаружении стимула, от информативности стимула, т. е. выигрыша в длине описания  $\Delta L$ , достигаемого при проведении группирования.

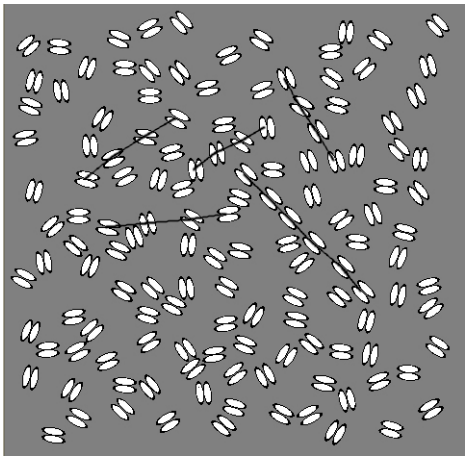
При генерации стимула и элементов поля (рис. 3 и 4) были использованы следующие значения:  $\varphi = \pi$  (элементы поля имели произвольную ориентацию),  $\sigma_r = 32$ ; число элементов в стимуле  $n$  варьировалось в диапазоне от 3 до 8, величины  $\psi$  и  $\sigma_c$  принимали значения в диапазонах  $[\pi/8, \pi/2]$  и  $[4, 16]$  соответственно.

На рис. 5а представлены примеры зависимости процента ошибок от числа элементов в стимуле для разных значений параметра  $\psi$ . Процент ошибок одновременно зависит от всех параметров стимула:  $n$ ,  $\psi$  и  $\sigma_c$ .

На рис. 5б приведена зависимость процента ошибок от информативности стимула  $\Delta L$ . Каждой точке соответствует усредненное по серии экспериментов значение процента ошибок при фиксированных параметрах  $n$ ,  $\psi$  и  $\sigma_c$ . Поскольку все точки ложатся на одну кривую, наиболее значительная часть вариативности в проценте ошибок может быть объяснена только различием в значении  $\Delta L$  для разных стимулов.

### Компьютерная модель

Предложенный критерий группирования (5) был использован при построении алгоритма поиска наборов элементов, объединение которых в группу приводит к уменьшению длины описания. Эксперименты показывают, что на изображениях, которые предлагались испытуемым, с помощью этого



**Рис. 6.** Примеры автоматического выделения стимула (а также нескольких групп регулярно расположенных элементов) из элементов поля

алгоритма также находился искомый стимул (см. рис. 6). Однако на данном этапе исследований сравнение процентов ошибок, допускаемых программой, реализующей алгоритм, и человеком, затруднительно, поскольку для человека время восприятия стимула было существенно ограничено, что оказывало большое влияние на вероятность обнаружения стимула.

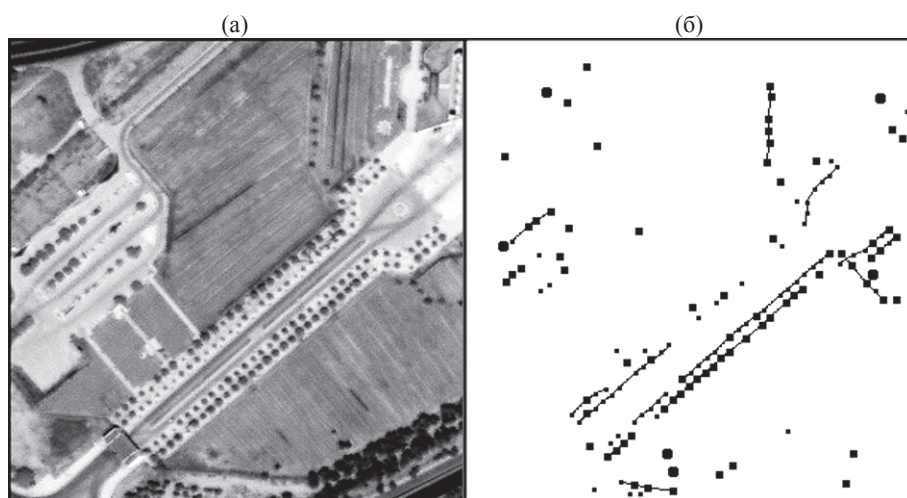
В то же время разработанный алгоритм может применяться для задач автоматического анализа реальных изображений (см. рис. 7), что подчеркивает важность исследования и перенесения механизмов зрительного восприятия человека в область компьютерного зрения.

## Заключение

Проведенные психофизические эксперименты и тестирование алгоритма группирования структурных элементов показали возможность применения принципа репрезентационной минимальной длины описания в приложении к проблеме перцептивного группирования, а следовательно, и возможность количественного описания законов гештальта с точки зрения уменьшения длины описания изображения в рамках определенного представления.

В данной работе, однако, использовался простой тип стимула и, соответственно, ограниченное представление изображений. Для расширения полученных выводов необходима постановка эксперимента с более широким набором стимулов. На основе принципа репрезентационной минимальной длины описания можно предположить, что процент ошибок зависит от выбранного представления, которое у разных людей может отличаться.

На выбор используемого человеком представления может также оказывать влияние количество сообщаемой априорной информации. В данном эксперименте человеку заранее сообщались все параметры стимула. Однако если, например, число элементов в стимуле заранее не известно, то для его описания необходимо выделить дополнительно примерно  $\log_2 n$  бит. Если ориентация или другие параметры стимула априори не известны, то они также должны быть описаны в процессе группирования, что уменьшит выигрыш в длине описания. Для ответа на вопрос, как это будет сказываться на вероятности обнаружения стимула, необходимо провести дополнительные эксперименты. Для расширения сферы применения алгоритмов группирования



**Рис. 7.** Аэрокосмическое изображение (а) и результат группирования выделенных на нем пятен (б).

в компьютерном зрении также требуется дальнейшее расширение представления, в рамках которого осуществляется описание изображений.

#### ЛИТЕРАТУРА

1. *Desolneux A., Moisan L., Morel J.-M.* Gestalt theory and Computer Vision // Seeing, Thinking and Knowing: Meaning and Self-Organisation in Visual Cognition and Thought, Carsetti A. (Ed.). 2004. 360 p.
2. *Robert A.* From Contour Completion to Image Schemas: A Modern Perspective on Gestalt Psychology // Technical Report 9702, Department of Cognitive Science, University of California. 1997. 58 p.
3. *Zhu S.-C.* Embedding Gestalt Laws in Markov Random Fields – a theory for shape modeling and perceptual organization // IEEE Trans. on Pattern Analysis and Machine Intelligence. 1999. V. 21. № 11. P. 1170–1187.
4. *Поманов А.С.* Распознавание образов и машинное восприятие: общий подход на основе принципа минимальной длины описания. СПб.: Политехника, 2007. 548 с.
5. *Vitanyi P.M.B., Li M.* Minimum description length induction, Bayesianism, and Kolmogorov complexity // IEEE Trans. on Information Theory. 2000. V. 46. № 2. P. 446–464.
6. *Поманов А.С.* Исследование представлений изображений на основе принципа репрезентационной длины описания // Изв. вузов. Приборостроение. 2008. Т. 51. № 7. С. 3–7.