

**Представление полутоновых объектов
с многоуровневым разрешением для ускоренного
распознавания образов**

Ганебных С. Н., Ланге М. М.

lange_mm@ccas.ru

Москва, ВЦ РАН

Соотношение эффективности и сложности является важной характеристикой большинства методов распознавания образов. Один из перспективных подходов к решению задачи распознавания в терминах соотношения «эффективность-сложность» базируется на применении представлений образов с многоуровневым разрешением [1, 3]. В настоящей работе предлагается способ построения пирамидальных представлений для широкого класса образов, заданных на изображениях двумерными объектами с неоднородной яркостной окраской. Приводятся оценки вычислительной сложности распознавания с использованием пирамидальных представлений и стратегии поиска решений по схеме последовательных приближений [2, 4]. Эффективность развитого подхода продемонстрирована экспериментальными результатами распознавания жестов языка ASL (American Sign Language).

Представление образов и база эталонов

В рассматриваемой модели образ определяется множеством пикселей $P = \{p_k : k = 1, \dots, N\}$, яркости которых принимают значения $z_k = z(p_k) = 1, 2, \dots, q$, а уровень $z = 0$ соответствует яркости фона изображения. В декартовых координатах (X, Y) любой образ P рассматривается как двумерное твердое тело (возможно многосвязное) с тензором инерции $\mathbf{G} = \begin{pmatrix} g_{yy} & -g_{xy} \\ -g_{yx} & g_{xx} \end{pmatrix}$, который образован ковариациями g_{xx} и g_{yy} и корреляционными моментами $g_{xy} = g_{yx}$, в виде соответствующих математических ожиданий по всем пикселям $p_k \in P$ с распределением «масс» $w(z_k) = z_k \left(\sum_{k=1}^N z_k \right)^{-1}$. Класс допустимых образов ограничен объектами, удовлетворяющими условию $(g_{xx} - g_{yy})^2 + 4g_{xy}g_{yx} > 0$.

Пирамидальное представление R любого образа P из указанного класса строится посредством рекурсивного разбиения образа на сегменты (на два сегмента на каждом шаге рекурсии) и аппроксимации сегментов эллиптическими примитивами [3, 5]. В результате формируется пирамида представлений $R = \{R_l : l = 0, \dots, L\}$, содержащая $L + 1$ уровней, в которой представление R_l l -го уровня содержит 2^l примитивов (Рис. 1а). Пример пирамиды представлений ASL-жеста, содержащей девять уровней ($L = 8$) дан на Рис. 1б.

R_0						
R_1		ASL-жест «а»	R_0	R_1	R_2	R_3
R_2			R_4	R_5	R_6	R_7
R_3						
R_4						
R_5						
R_6						
R_7						
R_8						
	а	б				

Рис. 1. Пирамидальное представление образа эллиптическими примитивами: а — структура представления; б — уровни представления ASL-жеста.

Если заданы M пирамидальных представлений эталонных образов, где представления однотипных эталонов образуют семантические группы $\mathbf{R}_i^m = \{R_{ij} : j = 1, \dots, m\}$, $i = 1, \dots, M/m$ (по m реализаций в каждой группе), то база эталонов строится на объединении всех семантических групп $\mathbf{R}^m = \bigcup_i \mathbf{R}_i^m$. Благодаря пирамидальной структуре представлений, база эталонов образует многослойную иерархическую структуру $\hat{\mathbf{R}}^m = \{\hat{\mathbf{R}}_l^m : l = 0, \dots, L\}$, в которой каждый l -ый слой $\hat{\mathbf{R}}_l^m$ содержит M эталонов, объединенных в M/m семантических групп, и каждый эталон слоя $\hat{\mathbf{R}}_l^m$ представлен 2^l примитивами.

Стратегия поиска решений

Решение о распознавании образа P по его представлению R основано на нахождении в базе эталонов \mathbf{R}^m семантической группы $\mathbf{R}_s^m \in \mathbf{R}^m$, которая удовлетворяет заданному критерию и идентифицирует принадлежность образа к s -му классу. Критерий распознавания предполагает, что для любой пары (R, \mathbf{R}_i^m) определена мера

$$D_i^{(l)}(R) = \frac{1}{m} \sum_{j=1}^m D^{(l)}(R, R_{ij}), \quad l = 0, 1, \dots, L, \quad (1)$$

где $D^{(l)}(R, R_{ij}) \geq 0$ — заданная мера различия R и R_{ij} , вычисляемая по соответствующим представлениям l -го уровня. С учетом (1) критерий распознавания сводится к нахождению для предъявленного R группы эталонов \mathbf{R}_s^m , удовлетворяющей условию

$$D_s^{(L)}(R) = \min_i D_i^{(L)}(R) \leq D_s^*, \quad (2)$$

где $D_i^{(L)}(R)$ — мера отклонения представления R от группы \mathbf{R}_i^m , вычисляемая по L -му уровню с максимальным разрешением, а $D_s^* > 0$ — заданный допустимый порог отклонения для группы эталонов \mathbf{R}_s^m .

Поиск решения по критерию (2) основан на последовательном отборе семантических групп в слоях $\hat{\mathbf{R}}_0^m, \hat{\mathbf{R}}_1^m, \dots, \hat{\mathbf{R}}_L^m$ базы \mathbf{R}^m и нахождении ближайшей группы в слое $\hat{\mathbf{R}}_L^m$. Число отбираемых групп в слое $\hat{\mathbf{R}}_l^m$ определяется экспоненциально убывающей функцией с параметром α :

$$\mathcal{N}_l = \begin{cases} \mathcal{N}2^{-\alpha l}, & l = 0, \dots, L, & \alpha \leq \frac{1}{L} \log \mathcal{N}; \\ \mathcal{N}2^{-\alpha(l-L+k)}, & l = L-k, \dots, L, & \alpha > \frac{1}{L} \log \mathcal{N}; \end{cases} \quad (3)$$

где $k = \lfloor \frac{1}{\alpha} \log \mathcal{N} \rfloor$, $\mathcal{N} = M/m$, а логарифм берется по основанию 2. В l -ом слое отбираются \mathcal{N}_l групп, ближайших к R по мере (1), причем в $l+1$ -ом слое отбор \mathcal{N}_{l+1} ближайших групп ведется среди \mathcal{N}_l групп, отобранных в l -ом слое.

Стратегия (3) позволяет оценить вычислительную сложность C_α поиска решения в терминах числа обрабатываемых примитивов в представлениях базы \mathbf{R}^m . Указанная сложность определяется соотношениями

$$C_\alpha = \begin{cases} M(2^{(1-\alpha)(L+1)} - 1)(2^{1-\alpha} - 1)^{-1}, & \alpha \leq \frac{1}{L} \log \frac{M}{m}, \\ M2^{L-k}(2^{(1-\alpha)(k+1)} - 1)(2^{1-\alpha} - 1)^{-1}, & \alpha > \frac{1}{L} \log \frac{M}{m}. \end{cases} \quad (4)$$

При большом числе эталонов ($M \rightarrow \infty$) асимптотические оценки сложности (4) имеют вид

$$C_\alpha \leq \begin{cases} O(M^{1/\alpha}), & \alpha < 1, \\ O(M \log M), & \alpha = 1, \\ O(M), & \alpha > 1. \end{cases} \quad (5)$$

Для сравнения вычислительная сложность поиска решения по критерию (2), полученного перебором всех семантических групп эталонов в L -ом слое $\hat{\mathbf{R}}_L^m$ базы \mathbf{R}^m , определяется величиной $C = M2^L$, которая при $L \leq \frac{1}{\alpha} \log \frac{M}{m}$ и $M \rightarrow \infty$ дает оценку $C \leq O(M^{1+1/\alpha})$.

Результаты распознавания ASL-жестов

Разработанный подход опробован для распознавания ASL жестов, доступных по адресу: www.vision.auc.dk/~tbm/Gestures/database.html. Исходные ASL жесты соответствуют 25-и буквам английского алфавита, в котором каждая буква задана 30-ю реализациями одного и того же жеста, образующими семантические классы. Эксперименты проводились

с пирамидальными представлениями, содержащими $L+1 = 6, 7, 8, 9$ уровней разрешения и базами эталонов \mathbf{R}^m с параметрами $m = 2, 3, 4$. Распознавание проводилось по критерию (2) с оптимизированными значениями пороговых коэффициентов D_s^* для каждой семантической группы эталонов. При различных значениях α вычислялась доля ложных решений (P_{false}) и доля отказов (P_{reject}), не удовлетворяющих критерию (2). Значения величин P_{false} и P_{reject} при $\alpha = 1$ представлены в таблице.

L		5	6	7	8
\mathbf{R}^2	P_{false}	0,005	0,000	0,005	0,011
	P_{reject}	0,112	0,101	0,066	0,059
\mathbf{R}^3	P_{false}	0,013	0,011	0,008	0,005
	P_{reject}	0,053	0,040	0,032	0,019
\mathbf{R}^4	P_{false}	0,005	0,000	0,005	0,000
	P_{reject}	0,027	0,032	0,016	0,016

Для указанных баз эталонов и уровней разрешения реальный вычислительный выигрыш $\gamma = C/C_{\alpha=1}$ лежит в диапазоне от 7 до 15 и увеличивается с ростом уровня максимального разрешения L .

Работа выполнена при поддержке РФФИ, проект № 06-01-00524.

Литература

- [1] *Berretti S., Del Bimbo A.* Multiresolution spatial partitioning for shape representation // IEEE Proceedings of ICPR-2004, Cambridge, England: IAPR, 2004. — Vol. 2, — pp. 775–778.
- [2] *Equitz W.E., Cover T.M.* Successive refinement of information // IEEE Transactions on Information Theory. — 1991. — Vol. 37, — pp. 269–275.
- [3] *Lange M.M., Ganebnykh S.N.* Tree-like Data Structures for Effective Recognition of 2-D Solids // IEEE Proceedings of ICPR-2004, Cambridge, England: IAPR, 2004. — Vol. 1, — pp. 592–595.
- [4] *Ланге М. М., Ганебных С. Н.* Многоуровневая структура данных и быстрый поиск на основе последовательных приближений // всеросс. конф. ММРО-12. — Москва, 2005. — Рр. 153–156.
- [5] *Lange M.M., Ganebnykh S.N.* Moment-Based Pattern Representation Using Shape and Grayscale Features // Proc. of the Iberian Conf. on Patt. Recogn. and Image Analysis, IbPRIA-2007, Spain. — Springer, 2007. — Рр. 523–530.