

УДК 681.51:007.5

В. А. Гороховатский, Р. В. Гребинник

Харьковский национальный университет радиоэлектроники
проспект Ленина, 14, 61166 Харьков, Украина

Иерархические меры в структурном распознавании изображений с применением сжатия описаний

Рассмотрены вопросы обработки информации в интеллектуальных системах компьютерного зрения. С целью повышения быстродействия распознавания предложена иерархическая мера для сопоставления структурных описаний, включающая этап фильтрации множеств признаков. Обсуждены результаты компьютерного моделирования.

Ключевые слова: компьютерное зрение, распознавание изображений, характерные признаки, дескрипторы, иерархическая обработка, фильтрация, быстродействие, помехозащищенность.

Введение и постановка задачи

Путем перехода от изображения к множеству *характерных признаков* (ХП) удается достичь существенного снижения объема информации, при этом характеристики распознавания (достоверность, помехозащищенность) практически не снижаются. Визуальные данные при этом рассматривают как мультимножества [1]. Характерный признак представляет собой числовой вектор фиксированной размерности, связанный с точкой пространства координат и называемый *дескриптором*. Основной задачей сопоставления объектов в виде множеств ХП есть надежное функционирование распознающих систем в неполном пространстве признаков и исключение ложных компонент, возникающих в процессе анализа. Для этих целей необходима разработка специальных мер сходства. Из-за действия помех и фона имеем модель данных, в которой определенная часть описания относится к ложному классу и не отражает свойств распознаваемых объектов [2, 3].

Установление соответствий на базе одиночных ХП имеет высокое быстродействие, однако, при этом не гарантируются достаточные достоверность и помехозащищенность. Лучшие показатели распознавания характерны для множественных соответствий, которые получаются путем анализа альтернатив на основе пороговой обработки. Однако быстродействие анализа групповых соответствий значительно ниже, чем одиночных.

Выходом из ситуации является введение этапа предварительной обработки (фильтрации, сжатия) множеств одиночных соответствий и затем осуществление

© В. А. Гороховатский, Р. В. Гребинник

полного компонентного анализа на сжатых подмножествах. За счет этого достигается снижение объема информации при сохранении достаточной надежности решений. Введенная таким образом *иерархическая обработка* на первом этапе реализует фильтрацию ХП, а окончательное решение (второй этап) строится на помехозащищенном способе анализа отношений. Несколько возросшая вычислительная сложность первого этапа компенсируется надежностью и малым числом анализируемых компонент [1–6].

Цель работы — формализация и изучение свойств иерархического подхода к анализу множеств ХП применительно к задаче структурного распознавания изображений в условиях неполного представления и ложных воздействий.

Задачи исследования — построение иерархических способов определения мер близости для конечных множеств дескрипторов, конкретизация и изучение особенностей применения мер, моделирование и оценка эффективности предлагаемых подходов для реальных данных.

Формализация и задача оптимизации подхода

Пусть U — некоторый универсум дескрипторов $\lambda \in U$. Обозначим Λ^1, Λ^2 — конечные множества дескрипторов $\Lambda^1, \Lambda^2 \subset U$, $\Lambda^1 = \{\lambda_i^1\}_{i=1}^{\mu(1)}$, $\Lambda^2 = \{\lambda_i^2\}_{i=1}^{\mu(2)}$, $\mu(1)$, $\mu(2)$ — мощности множеств. Введем на U расстояние $\rho(\lambda, \lambda_0)$, $\lambda, \lambda_0 \in U$, и определим нормированное метрическое пространство $\mathfrak{Z} = (U, \rho)$, $\rho: U \times U \rightarrow [0, 1]$. В качестве ρ в компьютерном зрении чаще всего применяют евклидово или манхэттенское расстояние в пространстве R^n , $U \subseteq R^n$. Распространенными способами нормировки дескрипторов есть деление на норму вектора или на максимум среди компонент [7].

Рассмотрим правило, соответствующее попаданию дескриптора λ внутрь шара радиуса δ с центром в λ_0 :

$$\rho(\lambda, \lambda_0) \leq \delta. \quad (1)$$

Правило (1) определяет некоторое семейство шаров с параметром δ и устанавливает бинарное отношение неразличимости дескрипторов [1]. Принадлежность шару означает эквивалентность λ, λ_0 . Заметим, что $\delta \in [0, 1]$ в силу свойств метрики ρ . Правилом (1) задаются эквивалентные подмножества элементов, как между множествами Λ^1, Λ^2 , так и внутри множеств. Выполнение условия (1) проверяется посредством предиката

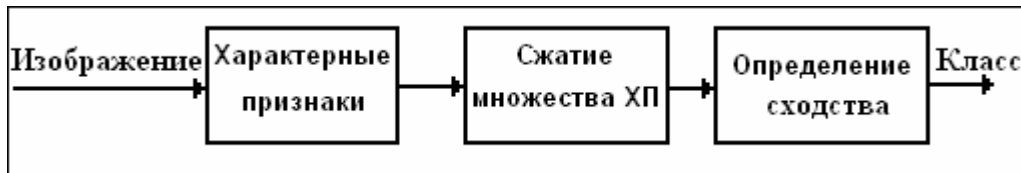
$$L[\rho(\lambda, \lambda_0), \delta] = \begin{cases} 0, & \rho(\lambda, \lambda_0) \leq \delta, \\ 1, & \text{else,} \end{cases}$$

равного 0 или 1 в зависимости от истинности неравенства. Вторым общепринятым способом анализа является пороговая обработка L^δ , отличающаяся от L тем,

что при выполнении (1) сходство равно значению $\rho(\lambda, \lambda_0)$. Области определения близких значений для L и L^δ совпадают, однако L^δ сохраняет значения метрики, которые можно использовать как веса элементов. Традиционно мера близости между множествами Λ^1, Λ^2 определяется путем анализа значений парной близости их элементов [7].

Рассмотрим модель подхода к сопоставлению множеств Λ^1, Λ^2 , где на первом этапе формируются *сжатые описания* Λ_*^1, Λ_*^2 , а на втором этапе вычисляется близость на их основе. Подход представим в виде последовательности двух отображений — Q_1 и Q_2 . Пусть Q_1 — это фильтрация множества в виде $Q_1: \Lambda^1 \rightarrow \Lambda_*^1, Q_1: \Lambda^2 \rightarrow \Lambda_*^2$, причем для мощностей множеств до и после преобразования должно выполняться условие $\mu(\Lambda_*^i) \ll \mu(\Lambda^i), i = 1, 2$. Второй этап традиционно реализуется отображением $Q_2: (\Lambda_*^1, \Lambda_*^2) \rightarrow R^1$, где результатом есть $d(\Lambda_*^1, \Lambda_*^2) \in R^1$ — мера близости.

Рассмотрим две схемы построения Q_1 : сжатие Λ^1, Λ^2 может осуществляться независимо либо путем совместного анализа в виде $Q_1: (\Lambda^1, \Lambda^2) \rightarrow (\Lambda_*^1, \Lambda_*^2)$. Применение независимой обработки эффективно на подготовительном этапе и может быть осуществлено, в частности, путем построения кластеров на множествах Λ^1, Λ^2 . При совместном анализе реализуется взаимосвязанный отбор элементов из множеств Λ^1, Λ^2 , т.е. осуществляется выбор их подмножеств, в необходимой мере отражающих близость исходных множеств. Независимая и совместная обработки могут быть применены также в комбинации друг с другом. Предлагаемая схема обработки данных показана на рисунке.



Последовательность обработки

В целом эффективность двухэтапной обработки в плане оптимизируемых вычислительных затрат определяется отношением величин $\mu(\Lambda_*^i)$ и $\mu(\Lambda^i)$ после и до обработки, и фактически задается значениями сформированной мощности $k^i = \mu(\Lambda_*^i), i = 1, 2$, т.к. $\mu(\Lambda^i)$ являются исходными величинами. Однако становится очевидным, что сокращение в признаковом описании, осуществленное на этапе Q_1 , в итоге может привести к снижению вероятности распознавания объектов.

Пусть база видеоданных представлена системой множеств $\Lambda_0 = \{\Lambda_0^j\}_{j=1}^J$, где J — число эталонов. Решение задачи распознавания визуального объекта, заданного множеством Λ , традиционно формализуется как $j_\Lambda = \arg \text{extr}_{j \in J} \mathcal{G}(\Lambda, \Lambda_0^j)$, где

\mathcal{Q} — мера близости описаний [1]. Представим задачу оптимизации предлагаемого двухэтапного подхода в задаче распознавания как

$$\begin{aligned} k(\Lambda, Q_1) \rightarrow \min, \quad k^j(\Lambda_0^j, Q_1) \rightarrow \min, \\ P(\Lambda, Q_1) \geq P_\varepsilon, \\ k^j / \mu(\Lambda_0^j) \geq \delta_\Lambda, \end{aligned} \quad (2)$$

где P — вероятность распознавания; P_ε — порог для величины P ; $\delta_\Lambda \in [0,1]$ — априорное ограничение для отношений мощностей в представлении j -го эталона. В задаче (2) минимизируются как значение мощности сжатого описания объекта $k = \mu(\Lambda)$, так и мощности сжатых описаний эталонов k^j . Выражения (2) не содержат зависимости от Q_2 , т.к. второй этап предполагается традиционным. Первое и второе ограничения в (2) могут быть связаны между собой, т.к. вероятность P зависит от числа элементов. С другой стороны, второе ограничение имеет самостоятельное значение, т.к. каждый объект имеет некоторое минимальное описание для различения его в базе видеоданных.

Из выражений (2) видим, что значения k, k^j, P являются функциями $\Lambda, \Lambda_0^j, Q_1, \delta_\Lambda$. В общем случае они зависят также от допустимого уровня шума, в условиях которого осуществляется распознавание. Таким образом, критерием в задаче оптимизации (2) есть минимальное число ХП для объекта и эталона, обеспечивающее надежное распознавание в базе видеоданных при заданном уровне помех.

Непосредственно аналитическое решение задачи оптимизации (2) затруднено в силу нелинейности выражений и отсутствия части строгого аналитического описания, в частности, для $P(\Lambda, Q_1)$. Будем решать задачу путем компьютерного моделирования.

Иерархическая мера на основе ближайших соседей

Как видим, ключевым моментом является этап Q_1 сжатия описаний. Рассмотрим подробно способ совместного анализа объекта и эталонов с целью получения значения $\mathcal{Q}(\Lambda, \Lambda_0^j)$ с применением Q_1 . Рассмотрим схему обработки, которая осуществляет формирование сжатых подмножеств на базе принципа ближайших соседей.

Традиционно на основе метрики ρ формируем последовательность оценок близости элемента $\lambda_i \in \Lambda$ к элементам $\lambda_v^j \in \Lambda_0^j$, $v = 1, \dots, \mu^j$, где $\mu^j = \mu(\Lambda_0^j)$ в виде набора $\rho_1(\lambda_i, \lambda_1^j), \rho_2(\lambda_i, \lambda_2^j), \dots, \rho_{\mu^j}(\lambda_i, \lambda_{\mu^j}^j)$, затем выполняем его ранжирование:

$$\rho_1(\lambda_i, \lambda_1^j) \leq \rho_2(\lambda_i, \lambda_2^j) \leq \dots \leq \rho_{\mu^j}(\lambda_i, \lambda_{\mu^j}^j), \quad (3)$$

где для упрощения записи использована исходная нумерация элементов из Λ_0^j . Да-

лее из (3) формируем кортеж близких к λ_i элементов. Метод ближайшего соседа и метрика Хаусдорфа выбирают один элемент на базе значения ρ_1 , метрика m ближайших соседей формирует кортеж на основе значений ρ_1, \dots, ρ_m , $m < \mu^j$ [7].

Путем анализа множества ранжированных наборов (3) для элементов $\lambda_i \in \Lambda$ вычисляем значение меры близости для множеств. Например, в метрике Хаусдорфа определяется максимум по i среди $\rho_1(\lambda_i, \lambda_i^j)$, а в одной из ее модификаций — среднее вида $\rho_c = [\mu(\Lambda)]^{-1} \sum_{i=1}^{\mu(\Lambda)} \rho_1(\lambda_i, \lambda_i^j)$ [8]. Другие варианты получения близости множеств используют пороговый принцип формирования [1, 2], когда элемент относится к кортежу в соответствии с предикатом L , т.е. если $\rho(\lambda_i, \lambda_i^j) \leq \delta_\rho$, где δ_ρ — порог.

Так как для вычисления близости множеств определяющую роль играют именно их эквивалентные элементы, возьмем за основу анализа кортежи близких элементов. Мощность этих кортежей должна удовлетворять последнему ограничению в задаче (2). Однако при пороговом подходе число элементов в сформированных кортежах может оказаться значительным из-за традиционной близости локальных характеристик изображений. Кроме того, выбор оптимального порога также является проблемой. Еще более ситуация усугубляется в связи с необходимостью построения отношений на кортежах близких элементов, что дополнительно увеличивает число признаков. В работе [4] уменьшение вычислений достигается путем ограничения числа высокоуровневых признаков. В качестве одной из решаемых таким способом проблем является обеспечение нужного уровня помехозащищенности без увеличения объема вычислений.

Одним из выходов в сложившейся ситуации является использование фиксированных мощностей m_Λ^*, m_j^* для кортежей близких элементов сравниваемых множеств, что гарантирует нужное быстродействие. Учитывая, что большинство методов структурного сопоставления основано на применении однозначных соответствий элементов, целесообразно выбрать условие $m = m_\Lambda^* = m_j^*$. Наиболее подходящим способом является принятие глобального решения о классе объекта на основе m ближайших соседей, т.е. применение ранжированного набора ρ_1, \dots, ρ_m из совокупности значений метрики.

Рассмотрим конкретный вариант принятия решений в задаче распознавания, основанный на выражении $j_\Lambda = \arg \max_{j \in J} \mathcal{G}(\Lambda, \Lambda_0^j)$, где мера близости $\mathcal{G}(\Lambda, \Lambda_0^j)$ есть мерой сходства, которая растет с увеличением числа соответствующих элементов. Вычислим $\mathcal{G}(\cdot)$ в виде $\mathcal{G}(\Lambda, \Lambda_0^j) = \zeta^j / \mu^j$, где ζ^j — число однозначных соответствий элементов $\lambda \in \Lambda$, сформированных при сопоставлении с эталоном Λ_0^j [1]. Однозначное соответствие $\theta(v)$ будем понимать как принятие локального решения на основе ρ_1 из (3) при выполнении условия $\rho_1 \leq 0,5\rho_2$, что обеспечивает высокую достоверность однозначных решений [9]. Выберем среди множества соответствий $\theta = \{\theta(1), \dots, \theta(\zeta^j)\}$ m наиболее значимых с точки зрения наименьших величин характеристики ρ_1 и построим на их основе решение для сжатых описа-

ний. Значение ζ^j / m характеризует степень сжатия для j -го эталона. В данной модификации отбор m наименьших значений осуществляется на глобальном уровне. Локальный уровень обрабатывает только значения ρ_1, ρ_2 , что не требует значительных вычислительных затрат.

В описанной постановке ключевым параметром, от которого зависят характеристики распознавания и, соответственно, решение задачи (2), является значение m . При этом важным является также то, какие именно точки представляют объект в сжатом описании.

Конкретной реализацией рассмотренного подхода есть вариант, когда на первом шаге путем применения метода m ближайших соседей формируется список наиболее схожих элементов, после чего на втором этапе реализуется перебор сочетаний из полученного набора соответствий [6]. Наиболее приемлемые результаты с точки зрения вероятности распознавания обеспечивает применение голосования пар и триад элементов [5].

Возможные варианты выбора значения m сводятся к следующему:

а) m равно минимуму среди числа описаний эталонов, т.е. $m = \mu_{\min}^j$, где $\mu_{\min}^j = \min_j \mu^j$;

б) m определяется как функция $m = f(\mu_{\min}^j)$, например, в виде $m = 0,5\mu_{\min}^j$;

в) m определяется как функция от описания $\mu(\Lambda)$ распознаваемого объекта;

г) m определяется как функция от описания эталона и принимает различные значения при сопоставлении с разными эталонами.

Из практического опыта для вариантов в), г) предложим следующее эмпирическое соотношение, на основе которого будем выбирать значение m : $m(m-1) = N$, где N — число элементов в полном описании объекта или эталона. Это соотношение в некоторой мере отражает процесс сравнения для пар признаков.

Результаты моделирования

Исследование предложенного метода и решение задачи оптимизации (2) осуществлено путем компьютерного моделирования для реальных изображений. За основу взята база данных из 30 аквариумных рыбок, для которой сформировано множество описаний на основе детектора SIFT [1]. В целях анализа вероятности распознавания в условиях шума осуществлялось искажение дескрипторов действием аддитивного гауссового шума, оцениваемого соотношением сигнал-шум $\eta = 1/\sigma$, где единица соответствует максимальному значению элемента дескриптора, σ — СКО шума. В качестве ρ использовано евклидово расстояние на множестве векторов.

Порог δ_ρ выбирался из следующих соображений. Характерный признак принимает участие в голосовании (значимый признак), если он отклоняется от 0 в пределах 10 % своего максимального значения $\rho_{\max} = \sqrt{128}$, 128 — размер дескриптора. Конкретно $\delta_\rho = 1,13$. Учитывая, что методы с установлением однознач-

ных соответствий дескрипторов из-за большей вероятности ложного совпадения имеют более низкий уровень достоверности, что сказывается на величине вероятности правильного распознавания [5], основное внимание уделено исследованию группы пороговых методов, при которых фиксируется множественное соответствие пар и триад ХП. В связи с этим рассмотрен набор методов, основанный на следующих типах голосования: 1) пары ХП; 2) триады ХП; 3) пары на сжатом множестве ХП; 4) триады на сжатом множестве ХП.

Рассмотрены разные варианты а)–г) выбора параметра m . Экспериментальная таблица содержит предельные значения соотношения сигнал–шум η , при котором вероятность правильного распознавания находится на высоком уровне ($P > 0,95$), а также среднее время распознавания в сек. Чем ниже значение η , тем лучше помехозащищенность метода. Индексы в номере метода рядом с цифрой соответствуют выбранному варианту формирования значения m .

Значения отношения сигнал–шум и времени распознавания

Метод	1	2	3а	4а	3б	4б	3в	4в	3г	4г
η	1,4	0,025	2,566	0,063	3,21	0,127	2,52	0,047	2,52	0,047
Время	0,085	5,19	0,068	0,148	0,026	0,059	0,033	0,081	0,029	0,076

Как видим из таблицы, за счет применения сжатых описаний удается значительно снизить время распознавания (при сравнении методов 2 и 4г — более чем в 60 раз, для методов 1 и 3г — в три раза). При этом помехозащищенность несколько снижается (для тех же методов на основе характеристики η — примерно в два раза). При этом помехозащищенность для сжатых описаний остается на высоком уровне — для пар признаков — в пределах $\eta = 3,2$, для триад — в пределах $\eta = 0,06$. Эти выводы можно распространить на все описанные виды сжатия.

Наши исследования показали, что среди рассмотренных способов выбора параметра m наиболее практичным является вариант г), т.к. он опирается на эталонную информацию, не подверженную действию помех, и имеет приемлемые показатели быстродействия и помехозащищенности (см. таблицу).

Результаты пробных экспериментов для другой традиционной базы coil-20 оказались близкими, что подтверждает универсальность рассмотренного подхода.

Выводы

Применение иерархического анализа множеств ХП при построении мер сходства для описаний визуальных объектов позволяет улучшить быстродействие распознавания в ситуациях, связанных с влиянием фона и локальных помех. Основой такого усовершенствования является сжатие имеющегося структурного описания.

Впервые показано, что сжатие множества соответствий характерных признаков не только является эффективным способом повышения быстродействия, но и обеспечивает достаточно высокую помехозащищенность распознавания. Изучены особенности и свойства иерархических мер сходства на основе метода ближайших соседей.

Практически важным является получение характеристик распознавания в зависимости от способа формирования параметра m для реальных баз видеоданных и экспериментальное подтверждение эффективности предложенных иерархических мер, что говорит о целесообразности применения описанных модификаций сжатия в задачах компьютерного зрения.

Перспективы исследования состоят в разработке подходов к выбору оптимальных значений параметров описанных мер при их применении в системах распознавания, а также в совместном использовании геометрической и негеометрической информации.

1. *Гороховатский В.А.* Структурное распознавание изображений на основе моделей голосования признаков характерных точек / В.А. Гороховатский, Е.П. Путятин // Реєстрація, зберігання і обробка даних. — 2008. — Т. 10, № 4. — С. 75–85.

2. *Баклицкий В.К.* Методы фильтрации сигналов в корреляционно-экстремальных системах навигации / В.К. Баклицкий, А.М. Бочкарев, М.П. Мусьяков. — М.: Радио и связь, 1986. — 216 с.

3. Методы компьютерной обработки изображений; [под ред. В.А. Сойфера]. — М.: Физматлит, 2003. — 784 с.

4. *Гороховатский В.А.* Иерархия пространственных отношений структурных признаков в задачах сопоставления визуальных объектов / В.А. Гороховатский // Системи управління, навігації та зв'язку: зб. наук. пр. — К.: ЦНДІ навігації і управління, 2008. — Вип. 3(7). — С. 85–89.

5. *Гороховатський В.О.* Дослідження завадостійкості методів зіставлення множин структурних елементів / В.О. Гороховатський, Р.В. Гребінник // Штучний інтелект. Інтелектуальні системи (ШІ-2008): 9-та Міжн. наук.-техн. конф. — Донецьк–Кацивелі, 22–27 верес. 2008 р. — Донецьк: ИПШІ «Наука і освіта». — Т. 1. — С. 301–305.

6. *Гребинник Р.В.* Двухэтапный подход к оценке сходства множеств визуальных компонент / О.В. Гороховатський, Р.В. Гребинник // Радиоелектроника и молодежь в XXI веке: 13-й международный молодежный форум, 30 мар. – 1 апр. 2009 г.: Сб. материалов форума. — Харьков: ХНУРЭ. — 2009. — С. 338.

7. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин; под ред. С.А. Айвазяна. — М.: Финансы и статистика, 1989. — 607 с.

8. *Srisuk S.* Face authentication using the trace transform / S. Srisuk, M. Petrou, W. Kurutach, A. Kadyrov // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). — 2003. — Vol. 1. — P. 305–312.

9. *Lowe D.G.* Distinctive Image Features from Scale-Invariant Keypoints / D.G. Lowe // International Journal of Computer Vision. — 2004. — 60, N 2. — P. 91–110.

Поступила в редакцию 05.07.2009