

Исследование метода формирования набора информативных признаков для различных типов признаков

Н.С. Кравцова¹, Р.А. Парингер^{1,2}, А.В. Куприянов^{1,2}

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

²Институт систем обработки изображений РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

Аннотация. Данная работа посвящена формированию набора информативных признаков для классификации текстурных изображений. В качестве классифицируемых изображений используются изображения дендритных кристаллограмм, а в качестве признаков – текстурные признаки и локальные признаки пространственного спектра. Для отбора набора информативных признаков применялся дискриминантный анализ. Классификация проводилась на основе метода 3 ближайших соседей. Проведённые эксперименты показали, что наименьшая ошибка классификации для набора информативных текстурных признаков составила 7%. Для набора информативных локальных признаков пространственного спектра наименьшая ошибка составила 4%.

1. Введение

В настоящее время компьютерные методы обработки медико-диагностических изображений используются в качестве одного из важнейших инструментов исследований и повышения эффективности ранней диагностики различных заболеваний [1-2]. Одним из информативных показателей состояния здоровья человека является изменение состава биологических жидкостей [3-4]. Изменение метаболизма, происходящее при появлении патологического состояния, влияет на состав жидкости, происходят многочисленные изменения молекулярного состава тканей и биологических жидкостей. Одним из способов выявления этих изменений является перевод жидкостей из одного фазового состояния в другое. В лабораторной диагностике наиболее удобным способом изменения фазового состояния жидкости является её кристаллизация. Изменение свойств кристаллов является следствием изменения физико-химических свойств биологической жидкости. Медицинские кристаллограммы – это структуры, образованные при кристаллизации солей, вследствие высушивания биологических жидкостей.

Кристаллограммы характеризуются наличием квазипериодических структур. Квазипериодические структуры определяются наличием многоконтурной упорядоченной текстуры с выраженной ориентацией. В каждой малой области квазипериодической структуры функция яркости является периодической вдоль определённого направления, что визуально выражается в системе полос. Основными параметрами таких изображений являются преимущественное направление полос и их густота в каждой точке изображения. Кроме того,

на таких изображениях выделяются особые точки и линии, в которых нарушается периодическая структура.

Для определения характеристик исходных изображений кристаллограмм хорошо зарекомендовали себя текстурные признаки [4] и признаки пространственного спектра [5]. В данной работе будет производиться классификация кристаллограмм с применением этих признаков, а также для каждого типа признаков будет сформирован набор информативных признаков, достаточных для проведения классификации текстурных изображений.

2. Формирование признаков

2.1. Текстурные признаки

Для получения текстурных признаков в данной работе было использовано программное обеспечение MaZda. Данное программное обеспечение изначально использовалось для текстурного анализа результатов маммографии. В дальнейшем MaZda стала адаптироваться для работы с медицинскими изображениями различного типа.

При использовании MaZda в данной работе было получено более 270 признаков для каждого изображения, часть из которых представлены в таблице 1.

Таблица 1. Примеры текстурных признаков.

Название признака	256_256_1	256_256_2	256_256_3	256_256_4
Mean	20,0	21,8	21,2	27,9
Variance	2876,6	3075,2	2990,6	3866,7
Skewness	2,91	2,73	2,78	2,27
Kurtosis	7,4	6,3	6,6	3,8
Perc.99%	240	238	237	242
S(1,0)AngScMom	0,72	0,69	0,70	0,62
S(1,0)Contrast	104,3	112,9	107,0	139,8
S(1,0)Correlat	0,705	0,701	0,709	0,706
S(1,0)SumOfSqs	176,9	188,9	183,8	237,6
S(1,0)InvDfMom	0,858	0,843	0,849	0,803
S(1,0)SumAverg	11,4	12,2	12,0	15,3

2.2. Локальные признаки пространственного спектра

В данной работе используются признаки, полученные путём вычисления общей энергии на выбранных областях изображения спектра. Большая часть спектра содержит информацию не пригодную для определения характеристик исходного изображения. Для получения признаков нужно рассматривать функцию исходного изображения в пространственной области и её преобразование Фурье $F(u, v)$. Тогда величина $|F(u, v)|^2$ определяет энергетический спектр изображения [6].

В качестве признаков применяются значения, полученные путём вычисления общей энергии на выбранной области изображения спектра. Сегментация изображения спектра на области осуществляется по формуле:

$$C_{r_1 r_2 \theta_1 \theta_2} = \sum_{r=r_1}^{r_2} \sum_{\theta=\theta_1}^{\theta_2} |F(r, \theta)|^2,$$

где $r = \sqrt{u^2 + v^2}$ - радиусы разделения колец, θ_1 и θ_2 - углы, ограничивающие сектор [7].

Так как изображение спектра симметрично относительно центра, то для выделения признаков будет использоваться только половина изображения пространственного спектра.

2.3. Дискриминантный анализ

В данной работе решается задача выявления набора информативных признаков. Для этого из набора признаков каждого изображения необходимо выявить ряд признаков, которые имеют

большой вес, т.е. нужно выбрать только те, которые предполагаемо лучше всего классифицируют изображение к каждому классу. Основным инструментом для этого стал дискриминантный анализ, поскольку он позволяет не только уменьшить количество исходных признаков, но ещё и преобразовать несколько исходных признаков в один новый, что существенно снижает вычислительные затраты [8-9].

3. Исследование влияния выбора набора признаков на ошибку классификации

Для проведения экспериментальных исследований использовались изображения дендритных кристаллограмм. Изображения делились на 2 класса по 117 изображений в каждом. Вся выборка разбивалась на обучающую и тестовую в равных долях случайным образом. Эксперимент над каждым набором признаков повторялся 500 раз. Так как в данной работе авторы хотели сосредоточиться на повышении достоверности классификации, основная сложность состояла в отборе признаков. Поэтому, для упрощения процесса классификации изображений, в качестве метода классификации был выбран метод 3х ближайших соседей, который ранее хорошо зарекомендовал себя при классификации подобных изображений и при этом вычислительно простой. Ошибка классификации, определяющая, в скольких случаях классификатор принял неправильное решение, вычисляется по формуле: $\varepsilon = (m/n) \cdot 100\%$, – где m – число изображений, классифицированных неверно (при классификации был присвоен другой класс), n – общее число классифицируемых изображений.

3.1. Классификация на основе локальных признаков пространственного спектра

Сначала рассматривались локальные признаки пространственного спектра. На первом этапе была проведена классификация по разбиениям, площадь которых достаточна для вычисления признаков, вычисляемым по формуле из пункта 2.2, после чего был проведён отбор признаков для каждого предложенного разбиения. В таблицах 2 и 3 отображено значение ошибки классификации до и после отбора информативного набора признаков соответственно, округлённых до целого [10].

Таблица 2. Значение ошибки до отбора признаков, %.

		Количество колец							
		1	2	3	4	5	6	7	8
Количество секторов	1	10	12	11	11	11	11	12	12
	2	9	7	8	8	8	8	7	8
	3	9	7	7	7	7	7	7	7
	4	9	7	7	6	6	6	6	6
	5	7	7	7	7	7	7	7	7
	6	8	7	7	7	7	7	7	7
	7	8	7	7	7	7	7	7	7
	8	7	7	7	7	7	7	7	7

Таблица 3. Значение ошибки после отбора признаков, %.

		Количество колец							
		1	2	3	4	5	6	7	8
Количество секторов	1	9	10	9	9	9	9	10	10
	2	8	8	7	7	7	7	7	7
	3	8	7	6	6	7	7	7	7
	4	7	7	6	6	5	5	4	4
	5	7	7	7	6	6	6	7	7
	6	7	7	7	7	7	7	7	7
	7	7	7	7	7	7	7	7	7
	8	7	7	7	7	7	7	7	7

Результат классификации по исходным признакам и набору информативных признаков показал, что минимальная ошибка классификации для полного набора признаков составила 6%, для сформированного набора признаков 4% от общего числа классифицируемых изображений.

3.2. Классификация на основе текстурных признаков

Во второй части работы использовались текстурные признаки. На первом этапе для каждого из тестируемых изображений при помощи программного обеспечения MaZda было получено максимальное количество признаков. Таким образом было получено около 270 уникальных признаков. Затем произведено формирование набора информативных признаков, по которым будет возможно проклассифицировать данные изображения с уменьшением затрат ресурсов и без потери качества классификации. Для этого, используя метод дискриминантного анализа, рассчитаны значения индивидуального критерия разделимости для каждого из полученных признаков. Затем признаки, у которых это значение наибольшее и вошли в новую группу признаков (таблица 4).

Таблица 4. Значения индивидуального критерия разделимости текстурных признаков.

Название признака	Значение инд. критерия разделимости
Sigma	0,551
S(3,0)Correlat	0,514
S(5,0)Correlat	0,493
S(4,0)Correlat	0,484
S(2,0)Correlat	0,476
WavEnLL_s-4	0,442
WavEnLL_s-3	0,441
WavEnLL_s-2	0,440
WavEnLL_s-1	0,436

Результат классификации по исходным признакам и набору информативных признаков показал, что ошибка классификации по полному набору признаков составила 22%, для сформированного набора признаков 7% от общего числа тестируемых изображений.

4. Заключение

В работе описан способ формирования информативных наборов признаков. В качестве исходных признаков использовались текстурные признаки и локальные признаки пространственного спектра. Исследования проводились на изображениях кристаллограмм. В результате классификации наибольшую ошибку показали текстурные признаки. При классификации по всему набору текстурных признаков ошибка составила 22%, после был использован информативный набор признаков, и ошибка снизилась до 7%. Применение полного набора локальных признаков пространственного спектра дало минимальную ошибку 6%, использование набора информативных признаков позволило снизить минимальную ошибку до 4%. Анализируя полученные результаты можно сделать вывод, что наборы информативных признаков позволяют производить классификацию изображений с большей точностью, при этом уменьшая количество вычисляемых признаков.

5. Благодарности

Работа выполнена при поддержке Федерального агентства научных организаций (соглашение № 007-ГЗ/Ч3363/26), Министерства образования и науки РФ в рамках реализации мероприятий Программы повышения конкурентоспособности СГАУ среди ведущих мировых научно-образовательных центров на 2013–2020 годы; грантов РФФИ № 15-29-03823, № 15-29-07077, № 16-41-630761, № 16-29-11698; № 17-01-00972; программы № 6 фундаментальных исследований ОНИТ РАН «Биоинформатика, современные информационные технологии и математические методы в медицине» 2018 г.

6. Литература

- [1] Камышев, И.А. Анализ планарных кристаллограмм лазерно облученных растворов биожидкостей / И.А. Камышев, А.Н. Малов, А.В. Неупокоева // Конференция. – Иркутск, 2012.

- [2] Пьянкова, О.Б. Применение кристаллографии биологических жидкостей в диагностике генеза механической желтухи / О.Б. Пьянкова, Т.И. Карпунина, Н.С. Карпунина // Гастроэнтерология Санкт-Петербурга. – 2010. – №2-3. – С. 259.
- [3] Булкина, Н.В. Сравнительная характеристика кристаллографической картины ротовой жидкости и жидкости десневой борозды или пародонтальных карманов при диагностике воспалительных заболеваний пародонта / Н.В. Булкина, Г.Е. Бриль, Д.Э. Постнов, В.Т. Поделинская, О.В. Еремин // Российский стоматологический журнал. – 2012. – №4. – С. 12-16.
- [4] Ильясова, Н.Ю. Информационные технологии анализа изображений в задачах медицинской диагностики / Н.Ю. Ильясова, А.В. Куприянов, А.Г. Храмов. – М.: Радио и связь, 2012. – 424 с.
- [5] Кравцова, Н.С. Исследование алгоритмов классификации текстурных изображений на основе анализа пространственного спектра / Н.С. Кравцова, Р.А. Парингер, А.В. Куприянов // Методы компьютерной диагностики в биологии и медицине, 2015. – С. 190-193.
- [6] Fukunaga, K. Introduction to statistical pattern recognition / K. Fukunaga. – San Diego: Academic Press, 1990. – 592 p.
- [7] Kravtsova, N. Development of methods for crystallogramms images classification based on technique of detection informative areas in the spectral space / N. Kravtsova, R. Paringer, A. Kupriyanov // CEUR Workshop Proceedings. – 2016. – Vol. 1638 – P. 357-363.
- [8] Ilyasova, N.Yu. Formation features for improving the quality of medical diagnosis based on the discriminant analysis methods / N.Yu. Ilyasova, A.V. Kupriyanov, R.A. Paringer // Computer Optics. – 2014. – Vol. 38(4) – P. 851-855.
- [9] Biryukova, E. Development of the effective set of features construction technology for texture image classes discrimination / E. Biryukova, R. Paringer, A. Kupriyanov // CEUR Workshop Proceedings. – 2016. – Vol. 1638. – P. 263-269.
- [10] Kravtsova, N.S. Parallel implementation of the informative areas generation method in the spatial spectrum domain / N.S. Kravtsova, R.A. Paringer, A.V. Kupriyanov // Computer Optics. – 2017. – Vol. 41 (4). – P. 585-587.

Research of the informative features generation method for various types of features

N.S. Kravtsova¹, R.A. Paringer^{1,2}, A.V. Kupriyanov^{1,2}

¹Samara National Research University, Moskovskoye shosse, 34, Samara, Russia, 443086

²Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

Abstract. This work is devoted to the formation of a set of informative features for the classification of textural images. As the classified images are used images of dendritic crystallograms, and as features of textural features and local features of the spatial spectrum. The conducted experiments showed that the smallest error in the set of informative texture attributes was 7%. For the collection of informative local features of the spatial spectrum, the smallest error was 4%.

Keywords: image analysis, spatial spectrum, discriminant analysis.