

УДК 681.518.3

**РАЗРАБОТКА СИСТЕМЫ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ С
ИСПОЛЬЗОВАНИЕМ САМООРГАНИЗУЮЩИХСЯ КАРТ КОХОНЕНА**

Парфентьев К.В., студент

*Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана
кафедра «Системы автоматического управления»*

Научный руководитель: Гаврилов А.И., к. т. н., доцент

Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана

pupkov@iul.bmstu.ru

В работе рассмотрены вопросы разработки систем распознавания изображений с использованием нейросетевых подходов. Для успешного решения задачи распознавания необходимо решить задачу предварительной обработки изображений. Для улучшения результатов распознавания можно использовать различные способы предварительной обработки, такие как синусное, косинусное и Фурье преобразования[1].

Методы предварительной обработки изображений

Предварительная обработка изображений играет важную роль в распознавании изображений. Без предварительной обработки трудно реализовать классификацию в силу того, что исходные изображения обладают свойствами высокого разрешения, низкого контраста, и т.д. Предварительная обработка может решить эти проблемы[2].

Преобразование цветного изображения в полутоновое

Изображение, полученное из фотоаппарата или видеокамеры обычно цветное. Большинство методов распознавания предполагает обработку полутоновых изображений, поэтому необходимо преобразовать цветное изображение в полутоновое, в котором останется только информация о яркости изображения[2]. Для получения значений яркости(S) берется взвешенная сумма по значениям красной(R), зеленой(G) и синей(B) составляющих (рис.1):

$$S = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B$$



Рис.1. Преобразование цветного изображения в полутоновое

Локализация и сегментация

Иногда целевой объект в изображении может быть небольшого размера, т.е. изображение содержит существенный объем фоновой составляющей. Локализация и сегментация могут удалить фоновую информацию[2].

Повышение контраста изображения

При слабом освещении изображение имеет низкий контраст, в результате чего распознавание может быть выполнено некорректно. В этом случае необходимо повысить контраст изображения (рис. 2). В работе используется метод эквализации гистограмм полутонового изображения [2] реализация которого выполняется в соответствии с соотношением:

$$P_k = \left[\sum_{i=0}^k \frac{255n_i}{n} \right] \quad k = 0,1,2 \dots, L - 1,$$

где n – количество пикселей; n_i – количество пикселей со значением яркости i ; L – максимальное значение пикселей; P_k – новое значение для пикселей, значение у которых равно k . Целью эквализации гистограмм полутонового изображения является уравнивание количества пикселей в каждом полутоновом значении.

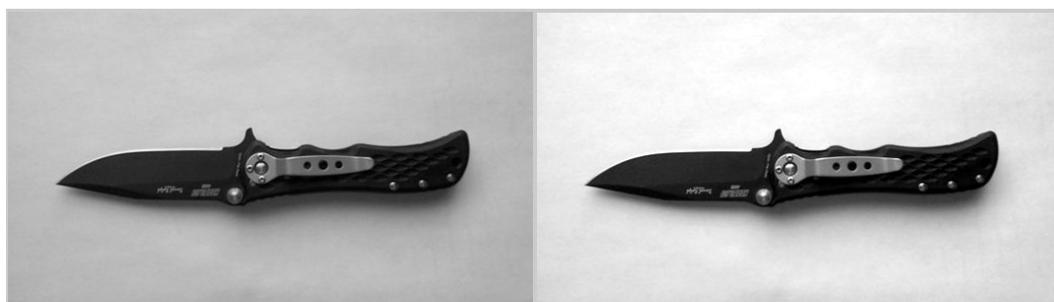


Рис. 2. Повышение контраста изображений

Масштабирование изображения

Изображение, полученное с фотоаппарата или видеокамеры, обычно обладает высоким разрешением (около мегабайта). Для глобальных или ядерных методов распознавания изображений непосредственное использование таких больших изображений будет значительно увеличивать время обработки. Поэтому необходимо масштабировать изображение к подходящей размерности. В данной работе при уменьшении размерности изображения значение яркости пикселя бралось как среднее геометрическое соседних пикселей.

$$S_0 = \frac{\sum_{i=1}^n \sum_{j=1}^m S_{ij}}{i \times j}$$

Структура самоорганизующихся карт SOM) и алгоритм обучения

Стратегия метода построения SOM заключается в представлении множества экспериментальных данных в виде топологической карты – решетки размерностью $n \times m$, состоящей из нейронов. Каждый нейрон представляет собой вектор, определяющий в d -мерном пространстве признаков некоторую характерную точку [3]. В результате реализации алгоритма обучения вектора-нейроны распределяются в пространстве признаков, аппроксимируя статистические взаимосвязи между данными обучающего множества.

Рассмотрим общую форму алгоритма обучения:

- Пусть задан набор векторов $M = \{m_i\}$ $i = \overline{1 \dots n}$, где m_i — “нейроны” карты (синаптические весовые вектора), он задан на некоторой регулярной сетке, задающей топологию, на которую будут отображаться данные выборки.
- Введем в пространстве векторов меру близости $d(x, y)$. Выбирая из обучающего множества вектора данных, найдем вектор-нейрон карты m_c , для которого выполняется условие $d(x_j, m_c) = \min(d(x_j, m_i))$. Этот нейрон и будет, пользуясь терминологией нейронных сетей, нейроном-“победителем” (в литературе по SOM, как правило, используется специальный термин “Bestmatchingunit” — BMU).
- Произведем настройку нейронов карты по рекурсивной формуле: $m_i(t + 1) = m_i + \alpha(t)h_{ci}(t)[x(t) - m_i(t)]$, где $h_{ci}(t)$ — симметричная невозрастающая функция, определенная вокруг m_c , называемая “функцией близости” (она определяет степень влияния нейрона—“победителя” на соседние нейроны), $\alpha(t)$ — коэффициент обучения (убывающая функция времени), причем $0 < \alpha(t) < 1$.

Таким образом, на каждом шаге алгоритма обучения определяется ближайший (в силу заданной метрики) к поступившим на вход сети данным нейрон, который «подтягивается» к входным данным вместе со своими соседями, определяемыми функцией близости $h_{ci}(t)$. Графическая интерпретация основной идеи алгоритма обучения представлена на рис.3.

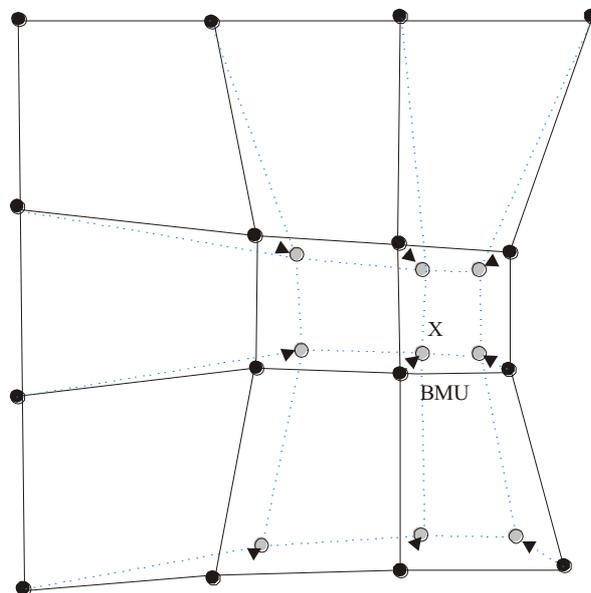


Рис. 3. Графическая интерпретация алгоритма обучения SOM

Разработка нейросетевой структуры для распознавания изображений.

Задача распознавания решалась следующим образом: в качестве векторов рассматривалась последовательность строк пикселей. Также добавлялась еще одна координата, идентификатор объекта (1,2,3 и т.д.). Определялся нейрон-победитель, считывалась метка принадлежности этого нейрона какому-либо объекту. Также вводилось ограничение на декартово расстояние. В случае, если нейрон-победитель находился слишком далеко от исходного вектора, принадлежность этого вектора объекту не подтверждалась.

Алгоритм обучения самоорганизующейся карты:

Инициализация

Наиболее распространены три способа задания первоначальных весов узлов:

1. Задание значений всех координат случайными числами.
2. Присваивание вектору веса значения случайного наблюдения из входных данных.
3. Выбор векторов веса из линейного пространства, натянутого на главные компоненты набора входных данных.

Основной цикл

Пусть t — номер итерации (инициализация соответствует номеру 0).

1. Выбрать произвольное наблюдение $x(t)$ из множества входных данных.

2. Найти расстояния от него до векторов веса всех узлов карты и определить ближайший по весу узел $M_c(t)$. Это — ВМУ. Условие на $M_c(t)$:

$$|x(t) - m_c(t)| \leq |x(t) - m_i(t)|,$$

для любого $m_i(t)$, где $m_i(t)$ — вектор веса узла $M_i(t)$. Если находится несколько узлов, удовлетворяющих условию, ВМУ выбирается случайным образом среди них.

3. Определить с помощью функции h (функции близости) соседей M_c и изменение их векторов веса.

Задание h

Функция определяет "меру близости" узлов M_i и M_c и изменение векторов веса. Она должна постепенно уточнять их значения, сначала у большего количества узлов и сильнее, потом у меньшего и слабее. Часто в качестве функции близости используется функция гаусса:

$$h_{ci}(t) = \alpha(t) e^{-\left(\frac{\|r_i - r_c\|^2}{2\sigma^2(t)}\right)}$$

где $0 < \alpha(t) < 1$ — коэффициент обучения, монотонно убывающий с каждой последующей итерацией (то есть определяющий приближение значения векторов веса ВМУ и его соседей к наблюдению; чем больше шаг, тем меньше уточнение);

r_i, r_c — координаты узлов M_i и M_c на карте;

$\sigma(t)$ — монотонно убывающая функция времени.

Параметры α, σ и их характер убывания задаются аналитиком.

Изменение векторов веса

Изменить вектор веса по формуле:

$$m_i(t) = m_i(t-1) + n_{ci}(t) \cdot (x(t) - m_i(t-1))$$

Таким образом, вектора весов всех узлов, являющихся соседями ВМУ, приближаются к рассматриваемому наблюдению.

Разметка карты

Для каждого вектора из обучающего множества находится ВМУ и каждому из этих ВМУ присваивается идентификатор (метка), означающая номер объекта (1,2,3). Этот номер может обозначать какому именно объекту принадлежит данная фотография.

Критерий обучения определяется, как среднее арифметическое расстояний между наблюдениями и векторами веса соответствующих им ВМУ:

$$\frac{1}{N} \sum_{i=1}^N |x_i - m_c|$$

где N - количество элементов набора входных данных.

Тестирование карты (распознавание)

1. Для тестирования также определялся ВМУ $M_c(t)$ в соответствии с соотношением. $|x(t) - m_c(t)| \leq |x(t) - m_i(t)|$

2. Проводилась проверка Евклидова расстояния тестируемого вектора и ВМУ $\sqrt{\sum (x_i(t) - m_{ci}(t))^2} \leq C$ Где C – заранее выбранное ограничение. x_i - координаты тестируемого вектора. m_{ci} - координаты вектора ВМУ.

3. Считывалась метка объекта и по ней определялась принадлежность тестируемой фотографии какому либо объекту.

Применение преобразования Фурье, синусных и косинусных преобразований с целью повышения эффективности системы.

Для повышения эффективности распознавания можно использовать преобразование Фурье, синусное и косинусное преобразование (рис. 4). Они позволяют перевести полученные координаты яркостей в частотную область. Также эти преобразования могут сократить количество существенной информации. По этим причинам подобные преобразования нашли себе применение в сжатии файлов, например, с расширением .jpeg.

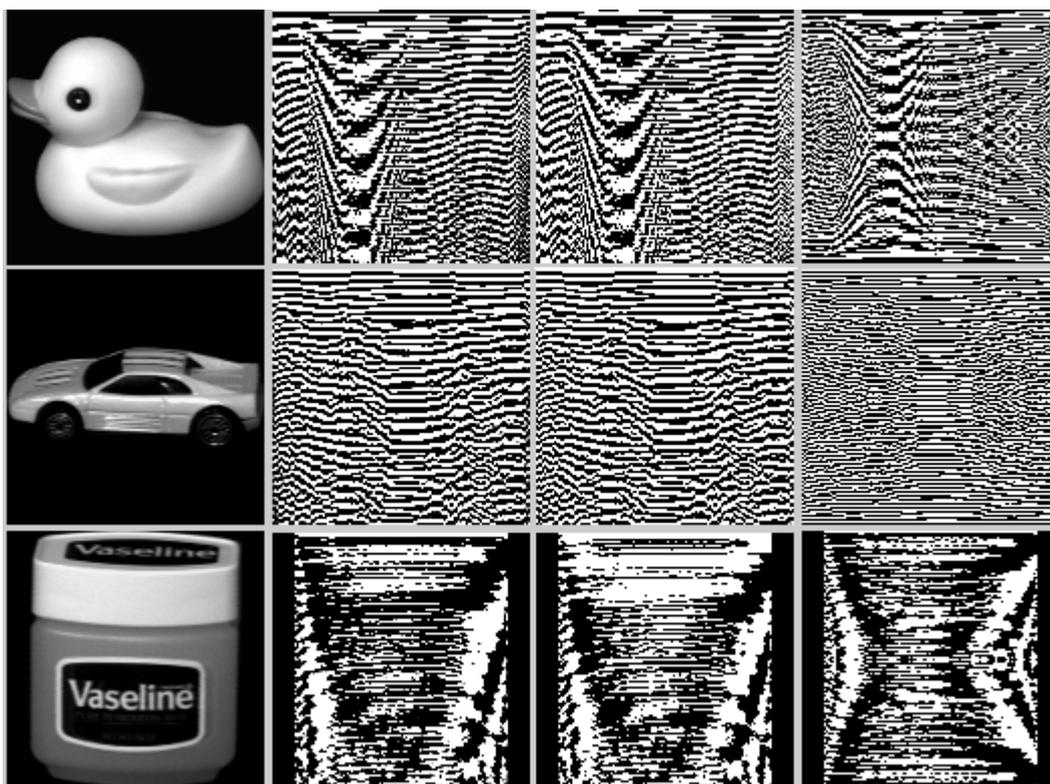


Рис. 4. Слева направо: оригинал, косинусное преобразование, синусное преобразование, преобразование Фурье

Распознавание лиц.

Для решения задачи распознавания использовалась база данных Orl[4]. Она содержит фотографии двадцати личностей, по 10 на каждого. Все фотографии в этой базе уже изначально полутоновые и обладают хорошим контрастом, поэтому из предварительной обработки изображения использовалась только операция масштабирования. Фотографии сжимались до разрешения 23x28 пикселей. Таким образом, каждый вектор будет иметь 644 координаты пикселей, плюс одну координату-идентификатор. Также производилось синусное преобразование для улучшения результатов.

Для обучения использовалось 80 фотографий 10-ти личностей (по 8 фотографий на каждую личность). Оставшиеся две фотографии использовались для тестирования. Также для тестирования будут использоваться 10 фотографий личностей, не входящих в обучающую выборку. Примеры фотографий из базы данных приведены на рисунке 5, красным отмечены фотографии, используемые для тестирования.

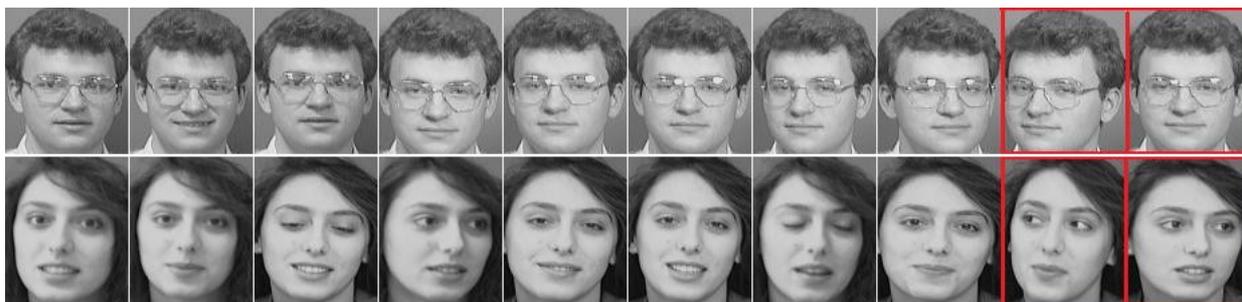


Рис. 5. Фотографии базы данных Ori

Результаты обучения и разметки приведены на рисунке 6.

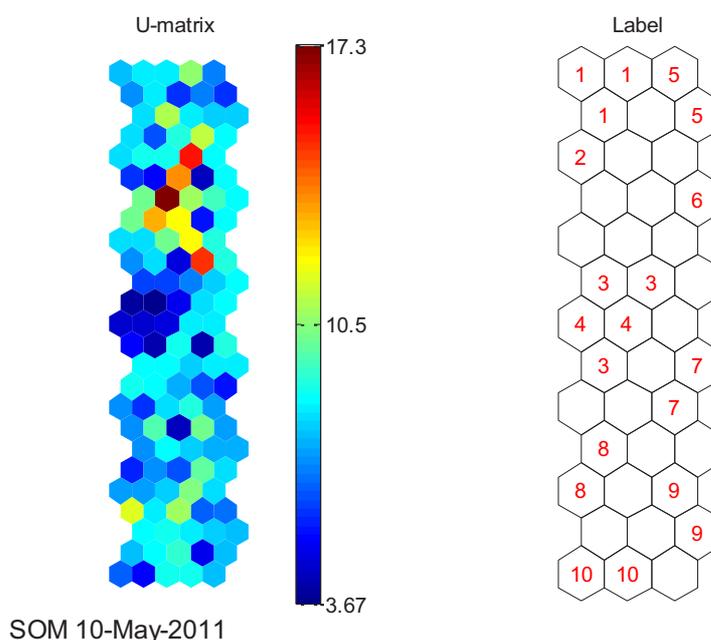


Рис. 6. Карта размером 13 на 3

Результаты тестирования данной карты: 5% ошибок первого рода (система не распознавала знакомый объект) и 0% ошибок второго рода (система распознавала ложный объект).

Распознавание моделей транспортных средств

В данной задаче использовались фотографии полученные вебкамерой. Снимались вырезанные модели транспортных средств. В качестве предварительной обработки использовались преобразование цветного изображения в полутоновое, улучшение контраста, локализация и сегментация, масштабирование. Все фотографии масштабировались до размера 48x32. Таким образом, каждый вектор будет иметь 1536 координаты пикселей, плюс одну координату идентификатор. Также производилось косинусное преобразование для улучшения результатов.

Для обучения использовалось 6 фотографий для каждого из 6 объектов. Для тестирования использовались 3 фотографии каждого из 6 объектов, входивших в обучающую выборку, а также 9 фотографий объектов, не входящих в нее. Примеры фотографий, использованных для обучения и тестирования, представлены на рис.7. Как видно из рисунка, они обладают низким качеством и разными размерами, что, естественно, затруднит распознавание.

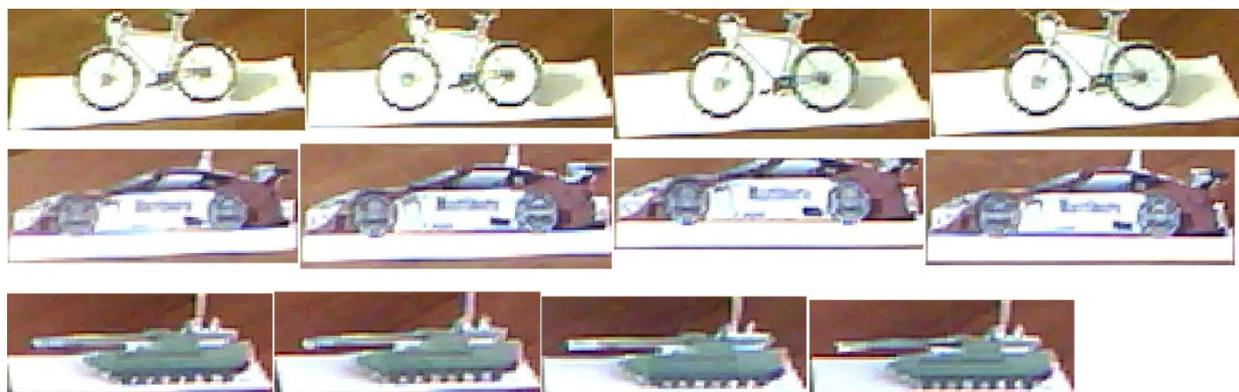
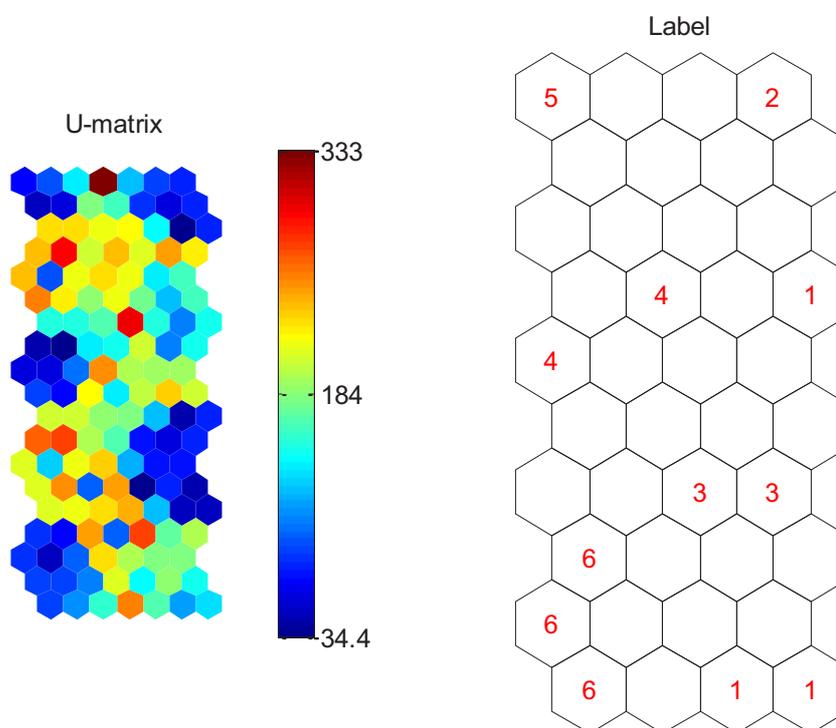


Рис. 7. Фотографии полученные с вебкамеры

Результаты обучения и разметки приведены на рисунке 8.



SOM 06-Dec-2012

Рис. 8. Карта размером 10 на 4

Результаты тестирования: 5.5% ошибок первого рода (система не распознавала знакомый объект) и 0% ошибок второго рода (система распознавала ложный объект).

Получение топографических карт по изображению подстилающей поверхности.

Задача заключалась в обнаружении на снимках карт характерных участков местности, таких как земля, поле, дорога, лес.

В качестве тестирования использовалась фотография подстилающей поверхности, размером 1392x697 пикселей (рис. 9).



Рис. 9. Фотография карты местности.

В качестве эталонов для обучения карты были взяты образцы конкретных участков местности размером 24x17 пикселей из других фотографий.

Примеры участков изображений, соответствующие четырем заданным классам (дорога, земля, лес, поле), представлены на рис. 10.



Рис. 10. Примеры участков фотографий, используемых для обучения.

Всего использовалось по 16 фотографий на каждый вид.

В этом случае для решения использовались самоорганизующиеся карты Кохонена.

В процессе тестирования карта разбивалась на области размером 24x17 пикселей и каждая область сравнивалась с эталонами.

Результат тестирования представлен на рис. 11.



Рис. 11. Результат тестирования.

Заключение

В статье рассмотрены основные принципы функционирования систем обработки цифровых изображений. Представлены структура, математическое описание и алгоритм обучения самоорганизующихся карт Кохонена. Разработан подход к решению задачи распознавания изображений, основанный на самоорганизующихся картах. Рассмотрены преобразования Фурье, синусное, косинусное преобразования и, по результатам исследования, подтверждена целесообразность их применения на этапе предварительной обработки изображений с целью повышения эффективности распознавания. В рамках проведенных работ разработана структура и проведена программная реализация системы распознавания изображений. Эффективность предложенных алгоритмов и программного обеспечения подтверждена решением задачи распознавания лиц, транспортных средств и построения топографических карт по изображению подстилающей поверхности с использованием карт Кохонена.

Список литературы

1. Дьяконов В.П. MATLAB 6.5 SP1/7.0 + Simulink 5/6. Обработка сигналов и проектирование фильтров. — М.: СОЛОН-Пресс, 2005, 576с.
2. Gonzalez R.C. Woods R.E. Digital Image Processing // Prentice Hall, Upper Saddle River, New Jersey, Second Edition, 2002. 813 p.
3. Kohonen T. Self-Organization and Associative Memory. - New-York: Springer - Verlag, 1988. 620 p.
4. Samaria F. Harter A. Parameterization of a stochastic model for human face identification // In Proc. IEEE Workshop on Applications of Computer Vision, pp. 138-142, 1994.