

Обучение алгоритмов выделения кожи на цветных изображениях лиц с использованием самоорганизующихся нейронных сетей и морфологических классификаторов на разрезах графов

Ю.В. Визильтер, В.С. Горбачев, С.Л. Карамеев, Н.А. Костромов
viz@gosniias.ru

Рассмотрены два способа обучения алгоритмов выделения кожи на цветных изображениях лиц – на основе самоорганизующейся нейронной сети типа «растущий нейронный газ» и морфологической классификации путем построения минимальных разрезов графов соседства на обучающей выборке. В качестве рабочего цветового пространства использовалось пространство CIE Lab. Показана эффективность обоих использованных методов, исследованы различия полученных результатов обучения.

Работа выполнена при поддержке РФФИ, гранты №11-08-01114-а, №11-08-01039-а.

1. ВВЕДЕНИЕ

Задача выделения человеческих лиц на цифровых изображениях получила широкое распространение в связи с бурным развитием информационных сетей и охранных систем. Существует много алгоритмов и методов выделения человеческих лиц на изображениях, но наиболее широко применяемым является известный алгоритм Виолы-Джонса, основанный на использовании процедуры обучения типа AdaBoost и Хааро-подобных признаков [1]. Недостатком этого алгоритма является, необходимость практически попиксельного сканирования изображения окнами различных размеров, что приводит к заметной потере производительности при обработке изображений большого размера. Для преодоления описанных выше трудностей применяются алгоритмы предобработки позволяющие сузить область поиска и тем самым повысить производительность. При этом широкое распространение получили методы, основанные на цветовой сегментации изображений по признаку принадлежности человеческой кожи. В работе рассматривается два различных способа построения подобных классификаторов - на основе самоорганизующейся нейронной сети типа «растущий нейронный газ» и морфологическое обучение методом минимального разрезания графа соседства для обучающей выборки.

2. МОРФОЛОГИЧЕСКОЕ ОБУЧЕНИЕ МЕТОДОМ МИНИМАЛЬНОГО РАЗРЕЗАНИЯ ГРАФА СОСЕДСТВА ДЛЯ ОБУЧАЮЩЕЙ ВЫБОРКИ

Морфологический подход к синтезу классификаторов основан на рассмотрении задачи синтеза метрического классификатора как задачи оптимальной сегментации (optimal labeling) конечной выборки точек метрического пространства. При этом «форма» и «сложность» классификаторов трактуются в терминах «формы» и «сложности» изображений (образованных метками классов на точках выборки), то есть в терминах математической морфологии [3]-[7].

Для алгоритмической реализации процедур синтеза метрических классификаторов используется техника построения минимальных разрезов графов [8]-[12], применяемая к графам соседства элементов обучающей выборки.

Рассмотрим задачу обучения с учителем. Пусть даны пространство объектов \mathcal{A} , конечное множество классов $C = \{c_1, \dots, c_l\}$, и известно разбиение объектов по классам: $c_{\mathcal{A}}(a): a \in \mathcal{A} \mapsto c \in C$. Обозначение $c_{\mathcal{A}}$ указывает на то, что функция определена на \mathcal{A} .

Производится описание объектов из \mathcal{A} дескрипторами из пространства описаний (признаков) \mathcal{X} : $x_{\mathcal{A}}(a): a \in \mathcal{A} \mapsto x \in \mathcal{X}$.

Случайным образом формируется конечная выборка объектов $A \subseteq \mathcal{A}$, $\|A\| < +\infty$ и соответствующая выборка описаний $X \subseteq \mathcal{X}$, $\|X\| < +\infty$. Каждому значению x ставится в соответствие класс c породившего его объекта a :

$$c_X(x): x_{\mathcal{A}}(a) \in X \mapsto c_{\mathcal{A}}(a) \in C.$$

По обучающей выборке c_X требуется построить такой распознающий алгоритм или классификатор

$$f_X(x): x \in X \mapsto c \in C,$$

который обеспечивает наилучшее разбиение X на классы из C . «Наилучшее разбиение» формализуется при помощи тестовой выборки

$$c'_Y(x): x \in Y \mapsto c \in C, Y \subseteq X, Y \cap X = \emptyset, \|Y\| < +\infty,$$

и критерия эмпирического риска на выборке Y :

$$J_Y(f_X) = d_H(f_Y, c'_Y) / \|Y\|,$$

$$d_H(f_Y, c'_Y) = \sum_{x \in Y} 1(f(x) \neq c'_Y(x)),$$

где $1(\text{true})=1$, $1(\text{false})=0$, $\|Y\| = \sum_{x \in Y} 1$. Здесь расстояние Хэмминга d_H имеет смысл числа ошибок классификации на тестовой выборке Y .

Отсюда критерий среднего ожидаемого эмпирического риска имеет вид

$$J_X(f_X) = E_{Y \subseteq X} \{J_Y(f_X)\},$$

где $E_{Y \subseteq X} \{ \cdot \}$ – математическое ожидание по всем возможным выборкам $Y \subseteq X$.

Таким образом, может быть сформулирована задача построения оператора оптимального синтеза θ , доставляющего минимум критерию $J_X(f_X) = J_X(\theta c_X)$:

$$\theta: c_X \in \Omega_X \mapsto f_X \in \Omega_X,$$

$$\theta = \arg \min_{\theta'} \{J_X(\theta' c_X)\}. \quad (1)$$

Здесь Ω_X и Ω_x – множества всех возможных разбиений выборки X и пространства \mathcal{X} по классам из C .

В большинстве известных подходов от задачи синтеза (1) сразу переходят к задаче обучения классификаторов заданного класса при помощи обучающего правила известного типа:

$$\theta \in \Theta: c_X \in \Omega_X \mapsto f_X \in F_X \subseteq \Omega_X,$$

$$\theta = \arg \min_{\theta' \in \Theta} \{J_Y(\theta' c_X)\}, \quad (2)$$

где F_X – класс классификаторов, Θ – класс алгоритмов обучения классификаторов из F_X на выборках $X \subseteq \mathcal{X}$.

Кроме того, вместо недоступного критерия $J_X(f_X)$, на практике используется критерий наблюдаемого эмпирического риска $J_X(\theta c_X)$, который имеет глобальный минимум в точке $f_X \equiv c_X$, заведомо непригодный для неизвестной тестовой выборки Y . Этой проблеме посвящена теория оценки и контроля переобучения, созданная Вапником и Червоненкисом [2]. Здесь эмпирический риск оценивается по обучающей выборке, но сложность решающего правила искусственно ограничивается. Для этого вводится понятие сложности классификатора $Q(f_X)$, а точнее сложности класса

классификаторов $Q(F_X)$. Соответственно вместо задачи (2) решается задача минимизации наблюдаемого риска с регуляризацией по сложности класса обучаемого классификатора:

$$\theta \in \Theta: c_X \in \Omega_X \mapsto f_X \in F_X \subseteq \Omega_X, \\ \theta = \arg \min_{\theta \in \Theta} \{J_X(\theta' c_X) + \alpha Q(F_X)\}, \quad (3)$$

где $\alpha \geq 0$ – параметр регуляризации, определяющий компромисс между точностью на обучающей выборке X и сложностью решающего правила, от которой зависит поведение f_X на тестовой выборке Y из (2).

Морфологический подход к машинному обучению направлен непосредственно на решение задачи (1) и позволяет исключить из рассмотрения априорно заданные классы классификаторов и алгоритмов обучения. При этом решение задачи (1) отыскивается в виде композиции решений подзадач:

$$\theta_\alpha = \delta_\alpha \Psi_\alpha \quad (4)$$

где Ψ_α – оператор (процедура) синтеза оптимального отклика классификатора на обучающей выборке X с учетом его сложности (локальной некомпактности)

$$\Psi_\alpha: c_X \in \Omega_X \mapsto f_X \in \Omega_X, \\ \Psi_\alpha = \arg \min_{\Psi'} \{J_X(\Psi' c_X) + \alpha Q_X(\Psi' c_X)\}; \quad (5)$$

δ_α – оператор (процедура) оптимальной корректной интерполяции (расширения) классификатора f_X на X с учетом сложности получаемого классификатора f_X :

$$\delta_\alpha: f_X \in \Omega_X \mapsto f_X \in \Omega_X, \\ \delta_\alpha = \arg \min_{\delta'} \{J_{NM}(\delta' f_X) + \beta Q(\delta' f_X)\}. \quad (6)$$

Здесь

$$J_{NM}(\delta f_X) = \{+\infty, \text{ если } \exists x \in X: \delta f_X(x) \neq f_X(x); \\ d_H(\delta f_X(x), \delta^{NM} f_X(x)) - \text{ в противном случае};$$

d_H – расстояние Хэмминга; δ^{NM} – простейший оператор интерполяции классификатора, соответствующий правилу ближайшего соседа (Nearest Neighbor).

Поскольку в задаче (1) функционал J имеет вид расстояния Хэмминга, из Утверждения 1 следует, что оператор θ_α (4) является алгебраическим проектором:

$$\theta_\alpha^2 = \theta_\alpha \Rightarrow \forall x \in X: \theta_\alpha f_X(x) = f_X(x). \quad (7)$$

Кроме того, на основе θ_α образуется система вложенных классов решающих правил, монотонная относительно α :

$$\forall \alpha \geq \beta \Rightarrow F_\alpha^X \subseteq F_\beta^X: Q(F_\alpha^X) \leq Q(F_\beta^X), \quad (8)$$

где $F_\alpha^X = \{f_X: \theta_\alpha f_X = f_X\}$ – множество классификаторов (разбиений), стабильное относительно проектора θ_α . В морфологиях изображений такая система вложенных проективных классов рассматривается как множество Пытьевских «форм» нарастающей сложности. В задаче синтеза классификаторов последовательность «форм» используется для решения проблемы переобучения методом минимизации структурного риска.

Определим такой критерий $Q_X(f_X)$, который отдает предпочтение решающим правилам f_X , более компактным на выборке X . Для этого определим систему вложенных окрестностей $O_k(x) \subseteq X$, $x \in X \subseteq X$, $k=1, \dots, \|X\|-1$, состоящих из k ближайших соседей. Введем локальную меру некомпактности f_X в окрестности $O_k(x)$:

$$Q_k(x, f_X) = q_H(O_k(x)) / \|O_k(x)\|, \\ q_H(O_k(x)) = \sum_{y \in O_k(x)} 1(f_X(x) \neq f_X(y)), \\ \|O_k(x)\| = \sum_{y \in O_k(x)} 1. \quad (9)$$

Тогда глобальная мера k -некомпактности имеет вид:

$$Q_X^k(f_X) = Q_H(X, f_X) / \|X\|, \\ Q_H(X, f_X) = \sum_{x \in X} Q_k(x, f_X). \quad (10)$$

Значение $Q_X^k(f_X)$ (11) характеризует эмпирическую оценку вероятности того, что один из k ближайших соседей в разбиении $f_X(x)$ будет отнесен к другому классу. При любых фиксированных k и X усложнению классификатора f_X соответствует нарастание меры k -некомпактности $Q_X^k(f_X)$. С

другой стороны, при увеличении параметра k в (10) преимущество получают более простые и «гладкие» разделяющие поверхности (см. [10]).

С учетом критерия (10) задача (5) сводится к хорошо известной задаче оптимальной разметки графа на основе скрытой Марковской модели [13], для которой существует эффективное приближенное решение методом минимального разреза графа, вычислимого за низко полиномиальное время относительно числа узлов графа (объектов в выборке) при любом конечном числе классов. Более того, для случая двух классов метод разреза графов может давать точное глобально оптимальное решение.

Алгоритм нахождения минимального разреза на графе с двумя терминальными вершинами позволяет находить минимум функционала энергии вида:

$$E(T) = E_0 + \sum_{i=1..N} E_i(t_i) + \sum_{(i,j) \in V} E_{ij}(t_i, t_j), \quad (11)$$

где N – число нетерминальных вершин графа; $T = \langle t_1, \dots, t_N \rangle$, $t_1, \dots, t_N \in \{0, 1\}$ – метки ассоциирования каждой нетерминальной вершины с одной из терминальных; $E_i(0)$, $E_i(1) \in \{0, 1\}$ – унарные потенциалы; $E_{ij}(t_i, t_j)$ – парные потенциалы, задаваемые четверкой действительных коэффициентов $E_{ij}(0,0)$, $E_{ij}(0,1)$, $E_{ij}(1,0)$, $E_{ij}(1,1)$; V – подмножество пар индексов переменных, задающее систему соседства на T .

Энергия (12) считается субмодулярной [12], если:

$$\forall (i,j) \in V: E_{ij}(0,0) + E_{ij}(1,1) \leq E_{ij}(0,1) + E_{ij}(1,0). \quad (12)$$

Для субмодулярной энергии (11)-(12) метод построения минимального разреза графа [8], [11] гарантирует нахождение точного минимума [9], [12].

Для реализации задачи синтеза двухклассового классификатора (5),(10) примем: $C = \{0, 1\}$, $N = \|X\|$, $T = \langle t_1, \dots, t_N \rangle$, $t_1 = f_X(x_1), \dots, t_N = f_X(x_N)$; $E_i(x) = 1(f_X(x) \neq c_X(x))$, $E_{ij}(t_i, t_j) = 1(f_X(x_i) \neq f_X(x_j))$; $V = \{(i,j): j \in O_k(x_i)\}$. Легко убедиться, что соответствующая энергия (11) будет субмодулярной, а значит, метод минимального разреза графа k -соседства для выборки c_X действительно оптимален и порождает α -семейства проекторов.

3. САМООРГАНИЗУЮЩАЯСЯ НЕЙРОННАЯ СЕТЬ – РАСТУЩИЙ НЕЙРОННЫЙ ГАЗ

В качестве второго способа цветовой сегментации рассматривается алгоритм кластеризации цветового пространства, основанный на аппроксимации цветового пространства изображения самоорганизующейся сетью, обучаемой по алгоритму растущего нейронного газа [15],[16]. Главным преимуществом данного алгоритма является осуществление так называемой «адаптивной» кластеризации входных данных. Т.е пространство не только разделяется на кластеры, но и определяется необходимое их количество исходя из топологии распределения самих данных. Начиная всего с двух нейронов, алгоритм последовательно изменяет (по большей части, увеличивает) их число, одновременно создавая набор связей между нейронами, наилучшим образом отвечающую распределению входных векторов, используя подход соревновательного хеббовского обучения. Каждый нейрон характеризуется т.н. «локальной ошибкой». Соединения между узлами характеризуются «возрастом». Пример такой структуры показан на рис.1.

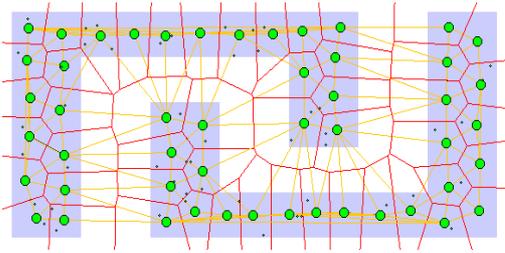


Рис 1. Структура нейронного газа: распределение кластеров (зеленый), связей (оранжевый) и топология данных (синий), конкретные сигналы показаны в виде отдельных точек.

Алгоритм работы растущего нейронного газа кратко можно описать следующим образом:

1. Инициализация: создать два узла с векторами весов, разрешенными распределением входных векторов, и нулевыми значениями локальных ошибок; соединить узлы связью, установив ее возраст равным 0.
2. Подать на вход нейросети вектор x .
3. Найти два нейрона s и t , ближайших к x , т.е. узлы с векторами весов w_s и w_t , такими, что $\|w_s - x\|^2$ минимальное, а $\|w_t - x\|^2$ второе минимальное значение расстояния среди всех узлов.
4. Обновить локальную ошибку нейрона-победителя s путем добавления к ней квадрата расстояния между векторами w_s и x : $E_s \leftarrow E_s + \|w_s - x\|^2$
5. Сместить нейрон-победитель s и всех его топологических соседей (т.е. все нейроны, имеющие соединение с победителем) в сторону входного вектора x на расстояния, равные долям ϵ_w и ϵ_n от полного.
 $w_s \leftarrow w_s + \epsilon_w \cdot (w_s - x)$
 $w_n \leftarrow w_n + \epsilon_n \cdot (w_n - x)$
6. Увеличить на 1 возраст всех соединений, исходящих от победителя s .
7. Если два лучших нейрона s и t соединены, обнулить возраст их связи. В противном случае создать связь между ними.
8. Удалить все соединения, возраст которых превышает age_{max} . Если после этого имеются нейроны, не имеющие связей с другими узлами, удалить эти нейроны.
9. Если номер текущей итерации кратен λ , и предельный размер сети не достигнут, создать новый нейрон r по следующим правилам:
 - Найти нейрон u с наибольшей локальной ошибкой.
 - Среди соседей u найти нейрон v с максимальной ошибкой.
 - Создать узел r "посредине" между u и v :
 $w_r = (w_u + w_v) / 2$
 - Заменить связь между u и v на связи u и r , v и r .
 - Уменьшить ошибки нейронов u и v , установить значение ошибки нейрона r .
 $E_u \leftarrow E_u \cdot \alpha$
 $E_v \leftarrow E_v \cdot \alpha$
 $E_r \leftarrow E_u$
 - Уменьшить ошибки всех нейронов j на долю β .
 $E_j \leftarrow E_j - E_j \cdot \beta$
10. Если критерий останова не выполнен, перейти к шагу 2.

4. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНОГО ИССЛЕДОВАНИЯ

Тестирование проводилось на базе изображений людей снятых при различных условиях съемки. Изображения были предварительно размечены вручную на области принадлежности пикселей классу кожи. Выборка точек для

обработки получена путем перевода тестовых изображений в цветное пространство CIE Lab с целью отделения цветowych компонент от яркости. Это повышает компактность представления, т.к. кожа имеет характерный цвет, а не яркостную составляющую. Обучение производилось на 10% точек от общего объема выборки в 800 000 точек.

При построении графа соседства использовался алгоритм триангуляции Делоне с динамическим кэшированием треугольников [17]. Нахождение оптимальных разрезов графов осуществлялось с использованием библиотеки [18]. Полученная вероятность правильной классификации цвета пикселя (кожа/не кожа) на тестовой выборке – 0,937.

Для самоорганизующейся нейронной сети были рассмотрены результаты при 32, 128 и 256 кластерах, полученных после кластеризации обучающей выборки. Полученная вероятность правильной классификации цвета пикселя (кожа/не кожа) на тестовой выборке – 0,925.

Как видно, численные значения результатов обучения, полученных двумя описанными методами в задаче цветовой сегментации кожи на изображениях лиц, оказались достаточно близки. Однако более подробное рассмотрение выделенных кластеров демонстрирует существенные различия в их форме. На рис.2 показана обучающая выборка в плоскости ab цветового пространства CIE Lab. Красными точками помечены пиксели кожи, зелеными – других классов.

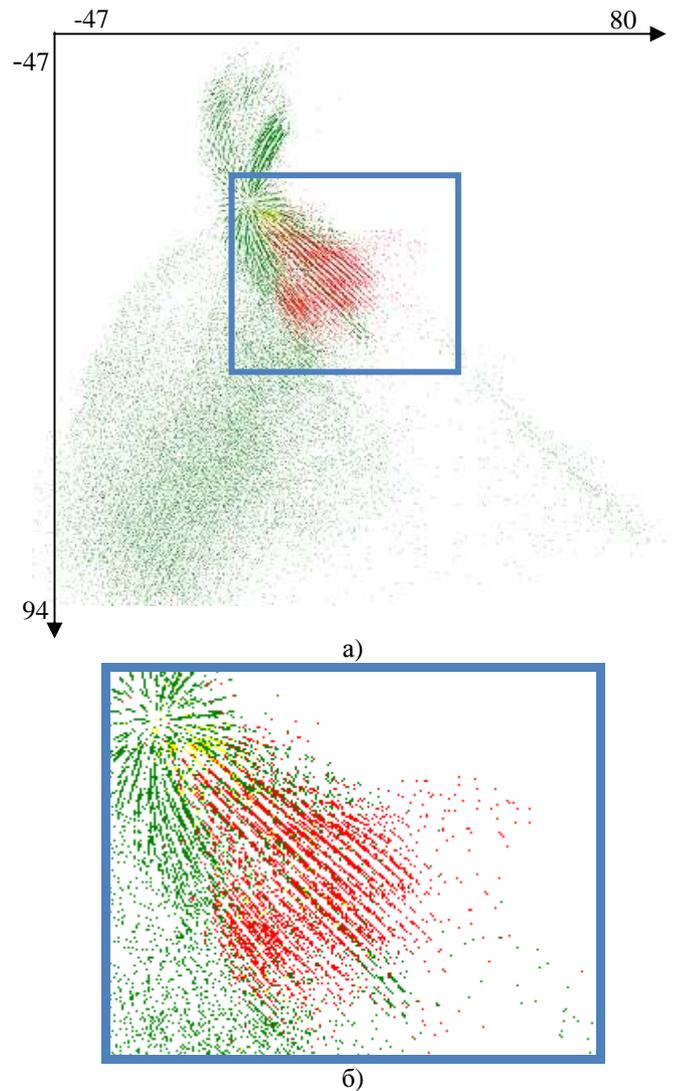


Рис2. а) Обучающая выборка в цветовой плоскости ab . б) Увеличенный фрагмент обучающей выборки содержащий точки "кожи"(красные) и "не кожи" (зеленые). Желтым показаны точки, имеющие обе метки.

На рис.3 приводится результат переразметки обучающей выборки после применения процедуры обучения на основе разреза графа соседства, а на рис.4 – после обучения на основе «растущего нейронного газа». Рис.5 демонстрирует различия в форме кластеров, полученных двумя способами обучения.

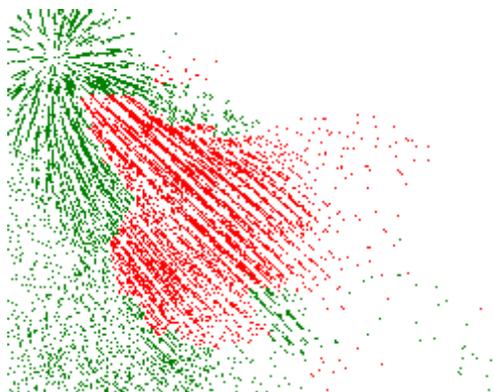


Рис3. Результат переразметки обучающей выборки после морфологического обучения на основе разреза графа.

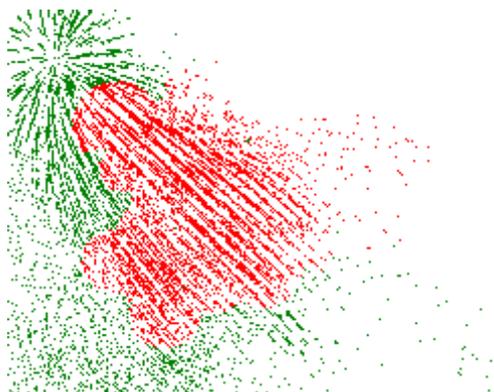


Рис4. Результат переразметки обучающей выборки после обучения на основе «растущего нейронного газа».

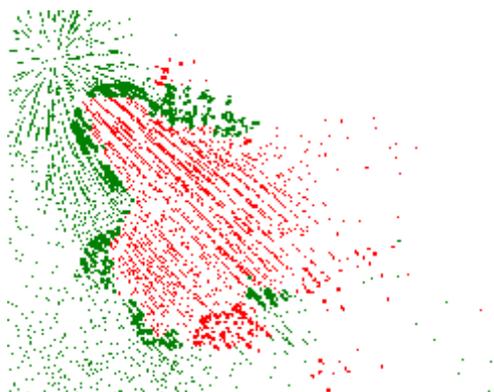


Рис5. Различия в результатах обучения (рис 3, 4). Жирным выделены точки, классифицируемые по-разному.

Как видно, значительные отличия в форме полученных кластеров цвета кожи указывают на существенно различную природу этих процедур обучения, что позволяет в дальнейшем рассматривать возможность их комплексирования с целью повышения вероятности правильной классификации.

Литература

- [1] Paul Viola, Michael Jones “Robust Real-Time Object Detection” Second International Workshop On Statistical And Computational Theories Of Vision – Modeling, Learning, Computing, And Sampling Vancouver, Canada, July 13, 2001.
- [2] Вапник В. Н. Восстановление зависимостей по [3] Pavel M. Fundamentals of Pattern Recognition. Marcel Dekker. Inc., New York, 1989.
- [4] Serra J. Image Analysis and Mathematical Morphology, Academic Press, London, 1982.
- [5] Пытьев Ю.П. Морфологический анализ изображений. Доклады АН СССР, 1983. Т. 269. № 5. С. 1061-1064.
- [6] Пытьев Ю.П., Чуличков А.И. Методы морфологического анализа изображений // М.: ФИЗМАТЛИТ, 2010. 336с.
- [7] Визильгер Ю.В. Обобщенная проективная морфология. «Компьютерная оптика», Том 32, №4. 2008, С.384-399.
- [8] L. Ford and D. Fulkerson. Flows in Networks. Princeton University Press, 1962.
- [10] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. Journal of the Royal Statistical Society, 51(2):271–279, 1989.
- [11] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In Proc. IEEE International Conf. Computer Vision (ICCV), pages 26–33, 2003.
- [12] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), 26(9):1124–1137, 2004.
- [13] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts?. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), 26(2):147–159, 2004.
- [14] Geman S., Geman D. Stochastic relaxation, Gibbs distributions, the Bayesian restoration of images. – IEEE Trans. Pattern Analysis, Machine Intelligence, 1984, № 6, pp.721-741.
- [15] B. Fritzke. A growing neural gas network learns topologies. In G. Tesauro, D. S. Touretzky, and T. K. Leen, editors, *Advances in Neural Information Processing Systems 7*, pages 625-632. MIT Press, Cambridge MA, 1995a.
- [16] B. Fritzke. Fast learning with incremental RBF networks. *Neural Processing Letters*, 1(1):-5, 1994b.
- [17] Скворцов А.В. Обзор алгоритмов построения триангуляции Делоне. Вычислительные методы и программирование. 2002. Т3. с.14-39.
- [18] Yuri Boykov, Vladimir Kolmogorov. MAXFLOW - software for computing mincut/maxflow in a graph. V.3.01: <http://www.cs.ucl.ac.uk/staff/V.Kolmogorov/software.html>