

## ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНОЙ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЯ ТОЛПЫ МЕТОДОМ ВОДОРАЗДЕЛОВ

**С.В. Шолтанюк**

*Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск,  
Беларусь, [ssholtanyuk@bsu.by](mailto:ssholtanyuk@bsu.by)*

Описана методика сегментации изображения толпы методом водоразделов с использованием показателя минимальной видимости сегментов, что позволяет решать задачу избыточной сегментации. Показано, что результаты сегментации не зависят от числа людей в толпе или разрешения изображения.

**Ключевые слова:** сегментация изображения; сегментация водоразделами; изображения толпы.

## FINDING THE OPTIMAL SEGMENTATION OF A CROWD IMAGE WITH WATERSHED METHOD

**S.V. Sholtanyuk**

*Belarusian State University, 4 Niezalieznasci Avenue, Minsk 220030, Belarus,  
[ssholtanyuk@bsu.by](mailto:ssholtanyuk@bsu.by)*

In this paper, we described a method for segmentation of a crowd image with watershed method which uses minimal saliency of segments. This approach addresses the oversegmentation problem. We showed segmentation results to depend neither on number of people in a crowd nor on an image resolution.

**Keywords:** image segmentation; watershed segmentation; crowd images.

### **Введение и постановка задачи**

Определение числовых характеристик толпы как в реальном времени, так и на готовых снимках и видеопоследовательностях является задачей, представляющей на сегодняшний день обширный научный интерес. Наиболее распространённым и действенным подходом оказалось использование глубоких нейронных сетей [1, 2]. Однако такие нейронные сети зачастую обладают достаточно сложной архитектурой, поэтому требуется большое количество временных и вычислительных ресурсов для их обучения. Альтернативными подходами, лишёнными этого недостатка, являются использование легковесных нейронных сетей [3], а также использование подходов, основанных на моделировании текстур скоплений людей [4].

Многие текстурные методы подсчёта количества людей в толпе, изображённой на снимке, опираются на сегментацию изображения [5–7]. В данной статье рассматривается сегментация при помощи водоразделов, который является быстрым и достаточно простым в реализации методом, нашедшим широкое применение во многих приложениях обработки изображений [8–9].

## 1. Методология

Сегментация изображения  $I$  методом водоразделов заключается в следующем. Каждому пикселю ставится в соответствие значение его яркости, тем самым задаётся функция  $B: \mathbb{N}^2 \rightarrow [0,1]$ . Затем происходит инициализация сегментов изображения, задаваемая либо явным перечислением исходных пикселей или областей (маркерный алгоритм), либо множеством локальных минимумов функции  $B$ , которые находятся исходя из градиентов (градиентный алгоритм). Сегменты расширяются в направлениях увеличения градиента до тех пор, пока не происходит соприкосновение с другими сегментами. Построение сегментов по такому алгоритму можно схематично описать через наводнение на поверхности, каждая точка которой имеет высоту  $B(i, j)$ , где  $(i, j)$  – её координаты, при источниках в заданных областях, причём по мере наполнения низменностей и впадин будет происходить слияние бассейнов, которые подлежат разделению водоразделами (Рисунок 1).

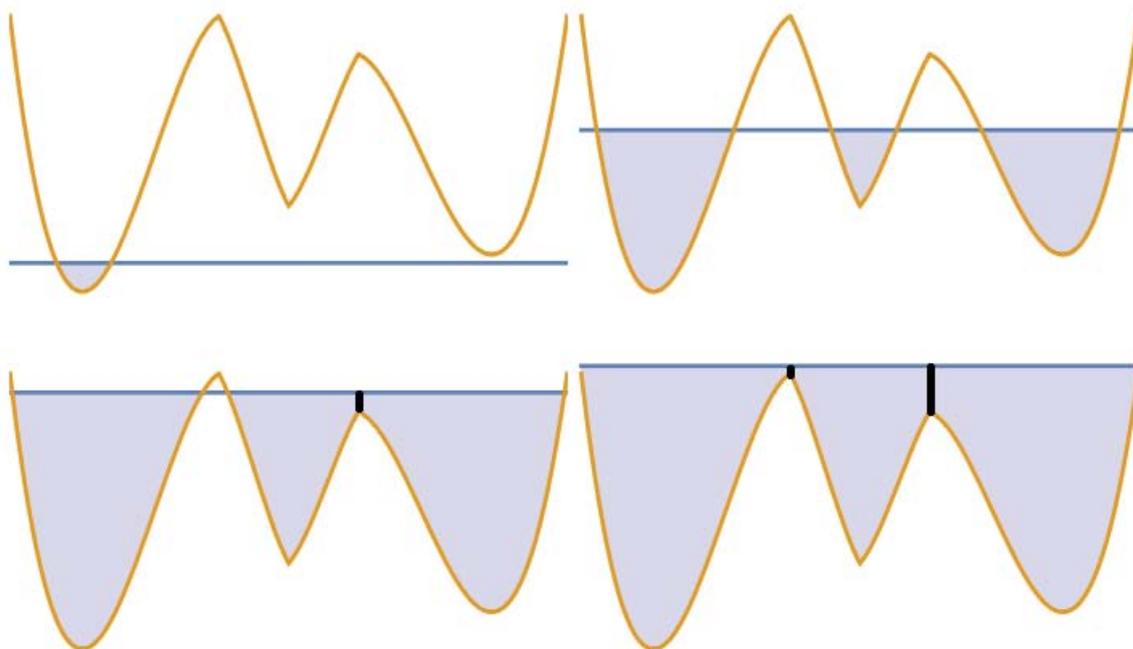


Рисунок 1 – Построение водоразделов по мере наводнения поверхности

В данной статье исследуется градиентный алгоритм сегментации изображения водоразделами. В общем случае он даёт избыточное количество сегментов (Рисунок 2, а). Поэтому следует ввести в рассмотрение понятие видимости сегмента  $S_i$  – величину, равную отношению «глубины бассейна», соответствующего данному сегменту, к «глубине» всего изображения:

$$\mu_i = \frac{\max_{(i,j) \in S_i} B(i,j) - \min_{(i,j) \in S_i} B(i,j)}{\max_{(i,j) \in I} B(i,j) - \min_{(i,j) \in I} B(i,j)}.$$

Нетрудно видеть, что  $\mu_i \in [0,1]$ . Тогда при «заполнении водой» водораздел между двумя сегментами ставится тогда и только тогда, когда у них обоих видимость не меньше некоторой величины  $\mu$ , а иначе они сливаются в единый сегмент. Таким образом, при увеличении  $\mu$  число сегментов уменьшается (Рисунок 2).

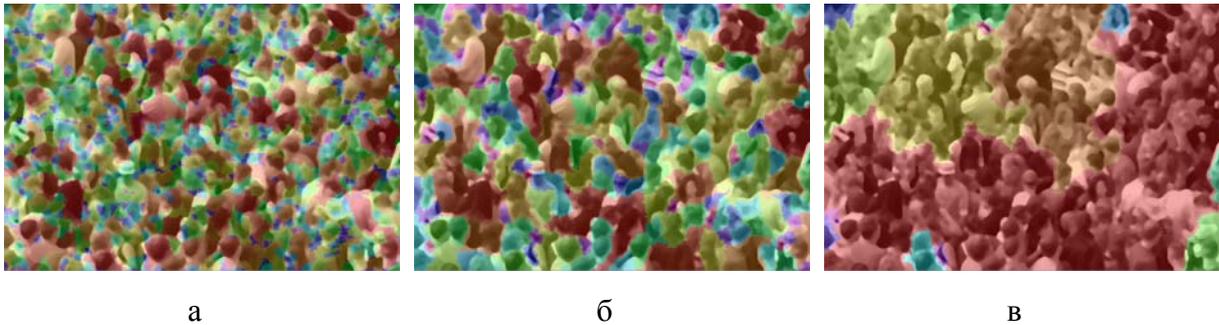


Рисунок 2 – Сегментация изображения при использовании минимальной видимости, равной 0 (а), 0.27 (б), 0.7 (в)

Минимальная видимость  $\mu$  для изображения подбирается таким образом, чтобы все сегменты имели площадь в пикселях, не бóльшую некоторого значения  $s_{\max}$ . Для вычисления  $s_{\max}$  выбирается уровень значимости  $\alpha$ , с которым проверяется гипотеза о том, что в некотором сегменте находится ровно одна метка. Если предположить, что расположение той или иной метки равновероятно во всяком пикселе исходного изображения, то вероятность того, что в сегмент  $S_i$  попало ровно  $k$  меток, равна

$$p_{\lambda_i}(k) = \frac{e^{-\lambda_i} \lambda_i^k}{k!}, \text{ где } \lambda_i \text{ – математическое ожидание количества меток, попавших в сегмент } S_i, \text{ которое пропорционально его площади и равно}$$

$$\lambda_i = \frac{s_i N}{S},$$

где  $s_i$  – площадь сегмента  $S_i$ ,  $N$  – число людей на изображении,  $S$  – разрешение изображения. В частности, при  $k = 1$  получается вероятность, равная  $p_{\lambda_i}(1) = e^{-\lambda_i} \lambda_i$ . Если рассматривать  $\lambda_i$  как случайную величину с функцией распределения  $F(\lambda) = p_{\lambda}(1)$ , то гипотеза о попадании в сегмент ровно одной метки отклоняется на уровне значимости  $\alpha$ , если  $\lambda$  попадает в двустороннюю критическую область  $[0, \lambda_{\min}(\alpha)) \cup (\lambda_{\max}(\alpha), +\infty)$ , где значения  $\lambda_{\min}(\alpha)$  и  $\lambda_{\max}(\alpha)$  можно найти как положительные решения следующих уравнений относительно  $x$ :

$$\int_x^1 e^{-\lambda} \lambda d\lambda = (1 - \alpha) \int_0^1 e^{-\lambda} \lambda d\lambda, \tag{1}$$

$$\int_1^x e^{-\lambda} \lambda d\lambda = (1 - \alpha) \int_1^{+\infty} e^{-\lambda} \lambda d\lambda.$$

Границы  $(1 - \alpha)$ -доверительных интервалов, определяемых из уравнений (1), приведены в таблице 1.

Таблица 1 – Минимальные и максимальные критические значения математического ожидания числа меток, попавших в сегмент

Уровень значимости $\alpha$	Минимальное критическое значение $\lambda_{\min}(\alpha)$	Максимальное критическое значение $\lambda_{\max}(\alpha)$
5%	0.172	5.113
1%	0.075	6.990
0.5%	0.052	7.777
0.1%	0.023	9.573

Таким образом, максимальная площадь сегмента вычисляется по следующей формуле:

$$s_{\max}(\alpha) = \frac{\lambda_{\max}(\alpha)S}{N}. \tag{2}$$

Значение минимальной видимости  $\mu$  для изображения  $I$  при заданном уровне значимости  $\alpha$  определяется по следующему алгоритму:

1. Инициализируется отрезок  $[\mu_{\min}, \mu_{\max}]$ , которому может принадлежать  $\mu$ :  $\mu_{\min} := 0$ ,  $\mu_{\max} := 1$ . Задаётся также точность  $\varepsilon$  определения искомой видимости.
2. Проводится сегментация изображения  $I$  водоразделами при наименьшей видимости  $\mu_{\text{med}} = \frac{\mu_{\min} + \mu_{\max}}{2}$ . Если максимальная пло-

щадь сегмента оказалась больше порога, вычисляемого по формуле (2), то переопределяется правый конец отрезка:  $\mu_{\max} := \mu_{\text{med}}$ , в противном случае – левый:  $\mu_{\min} := \mu_{\text{med}}$ .

3. Алгоритм завершает свою работу либо когда величина отрезка  $[\mu_{\min}, \mu_{\max}]$  стала не больше  $\varepsilon$ , либо когда  $\mu_{\text{med}}$  оказывается меньше некоторого изначально заданного порога  $\mu_0$ . Последнее условие вводится во избежание избыточной сегментации.

## 2. Вычислительный эксперимент и его результаты

Для эксперимента использовались изображения скоплений людей из набора данных ShanghaiTech [10], который для каждого изображения содержит также метки с их расположением. Перед сегментацией изображений водоразделами проводилась предобработка: каждый снимок переводился в чёрно-белый формат, после чего к нему применялась медианная фильтрация. Также сегментации подвергались негативные изображения (Рисунок 3). Таким образом, каждому исходному снимку соответствуют два чёрно-белых изображения, непосредственно подлежащих сегментации.



Рисунок 3 – Пример первоначального снимка (а) и два изображения, полученных после предобработки (б, в)

После предобработки для каждого полученного чёрно-белого изображения при уровнях значимости из таблицы 1 определялась минимальная видимость по описанному выше алгоритму при  $\varepsilon = 0.005$ ,  $\mu_0 = 0.25$ , причём если алгоритм заканчивался по условию  $\mu_{\text{med}} < \mu_0$ , то изображению при заданном уровне значимости пропускалось. Для каждого из исходных изображений в итоге выбиралась сегментация, соответствующая максимальному из двух полученных значений минимальной видимости.

Из полученной сегментации исключались сегменты площадью меньше

$$s_{\min}(\alpha) = \frac{\lambda_{\min}(\alpha)S}{N}.$$

Доля пропущенных пар «изображение – уровень значимости» оказалась равна 69.6%, доля оставшихся пар, где большее значение минимальной видимости получилось при сегментации негативного чёрно-белого изображения, равна 80.8%. Остальные результаты приведены в таблице 2.

Таблица 2 – Результаты эксперимента

Число людей на изображениях	Минимальное	33
	Медианное	259
	Максимальное	2348
Минимальная видимость сегментов	Минимальное	0.26
	Медианное	0.43
	Максимальное	0.75
Число кластеров	Минимальное	18
	Медианное	246
	Максимальное	1551
Среднее число меток в кластере	Минимальное	0.392
	Медианное	1.131
	Максимальное	2.962
Средняя доля исключённых меток		1.6%

Эти результаты не показали зависимости ни от числа людей на изображении, ни от его разрешения.

### Заключение

Сегментация водоразделами снимков с изображением толпы в некоторых случаях может быть улучшена при помощи задания минимальной видимости сегмента, которая может быть оценена при помощи описанной в данной статье методики. При этом, с одной стороны, получается число сегментов, достаточное для дальнейшего локального анализа изображения, а с другой – такая сегментация позволяет исключить сегменты малой площади без значительного сокращения общего числа людей на снимке. Однако во многих случаях имеет место избыточная сегментация в силу того, что остаются кластеры больших размеров. Также показано, что в большинстве случаев результаты сегментации можно улучшить, используя негативное изображение.

## Библиографические ссылки

1. Li Y., Zhang X., Chen D. CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018. P. 1091–1100.
2. hi-Qi Cheng, Qi Dai, Hong Li, Jingkuan Song, Xiao Wu, Alexander G. Hauptmann. Rethinking Spatial Invariance of Convolutional Networks for Object Counting // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022. P. 19638–19648.
3. Sholtanyuk S., Leunikau A. Lightweight deep neural networks for dense crowd counting estimation // *Pattern Recognition and Information Processing: proceedings of the 15th international conference*. 2021. P. 61–64.
4. Шолтанюк С.В. Текстурные методы подсчёта количества людей в плотной толпе // *Веб-программирование и интернет-технологии WebConf2021 : материалы 5-й Междунар. науч.-практ. конференции*. 2021. С. 173–176.
5. Wang Q., Breckon T.P. Crowd Counting via Segmentation Guided Attention Networks and Curriculum Loss // *IEEE Transactions on Intelligent Transportation Systems*. 2022. P. 1–11.
6. Rahmun M., Deb T., Bijoy S.A., Raha M.H. UAV-CROWD: Violent and non-violent crowd activity simulator from the perspective of UAV. arXiv:2208.06702.
7. Bruno M., Ferjani Z., Sabeur B., Arbab-Zavar D., Cetinkaya L., Johnstone M., Sallal D., Benaouda. High-Level Feature Extraction for Crowd Behaviour Analysis: A Computer Vision Approach // *International Conference on Image Analysis and Processing: revised selected papers of 21st international conference, part II*. 2022. P. 59–70.
8. Vincent L., Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations // *IEEE Transactions On Pattern Analysis And Machine Intelligence*. 1991. № 13(6). P. 583–598.
9. Jackway P. Gradient Watersheds in Morphological Scale-Space // *IEEE Transactions on Image Processing*. 1996. № 5(6). P. 913–921. DOI: 10.1109/83.503908.
10. Zhang Y., Zhou D., Chen S., Gao S., Ma Y. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. P. 589–597.