



# Statistical Gaussian Model of Image Regions in Stochastic Watershed Segmentation

Jesus Angulo

## ► To cite this version:

Jesus Angulo. Statistical Gaussian Model of Image Regions in Stochastic Watershed Segmentation. Second International Conference on Geometric Science of Information, Oct 2015, Palaiseau, France. pp.396-405, 10.1007/978-3-319-25040-3\_43 . hal-01134047v2

**HAL Id: hal-01134047**

**<https://hal.science/hal-01134047v2>**

Submitted on 17 Jan 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Statistical Gaussian Model of Image Regions in Stochastic Watershed Segmentation

Jesús Angulo

MINES ParisTech, PSL-Research University,  
CMM-Centre de Morphologie Mathématique, France  
`jesus.angulo@mines-paristech.fr`

**Abstract.** Stochastic watershed is an image segmentation technique based on mathematical morphology which produces a probability density function of image contours. Estimated probabilities depend mainly on local distances between pixels. This paper introduces a variant of stochastic watershed where the probabilities of contours are computed from a gaussian model of image regions. In this framework, the basic ingredient is the distance between pairs of regions, hence a distance between normal distributions. Hence several alternatives of statistical distances for normal distributions are compared, namely Bhattacharyya distance, Hellinger metric distance and Wasserstein metric distance.

## 1 Introduction

Image segmentation is one of the most studied and relevant problems in low level computer vision. Indeed, the state-of-the-art is vast and rich in multiple paradigms. We are interested here on approaches based on statistical modeling of pixels and regions. Examples of methods fitting in such a paradigm and having excellent performance are mean shift [9] and statistical region merging [15]. Hierarchical contour detection is another successful paradigm with approaches based for instance on machine learned edge detection [4] or on watershed transform [6].

Instead of dealing with a determinist set of contours, the idea of the stochastic watershed [2] (SW) is to estimate a probability density function (pdf) of contours by MonteCarlo simulations. Some variants included multiscale framework [3], bagging framework [11], robust framework [5], etc. It was shown in [16] that the corresponding pdf obtained by SW can be calculated in closed form without simulation by using graph algorithms, for more recent results see also [17]. Nevertheless, we focuss here on an approach working on simulations. In particular, our contribution is in the line of [12], where the estimation of the probability of each contour is based on a regional model of each region, the model being in [12] the mean color. In the present work, the approach is pushed forward such that each region should be modeled as a multivariate normal distribution. The basic ingredient will be a distance between pairs of regions, hence a distance between normal distributions. In this context, several alternatives of statistical distances for normal distributions are compared.

The rest of the paper is organized as follows. In Section 2, we remind the MonteCarlo framework of stochastic watershed and in particular the simulation of regionalized random germs. Section 3 introduces the contribution of the paper: first, distances between multivariate normal distributions are reviewed; second, computation of probability density of contours based on a normal region model is formulated. Results are discussed and compared to statistical region merging segmentation [15].

## 2 Remind on Stochastic Watershed

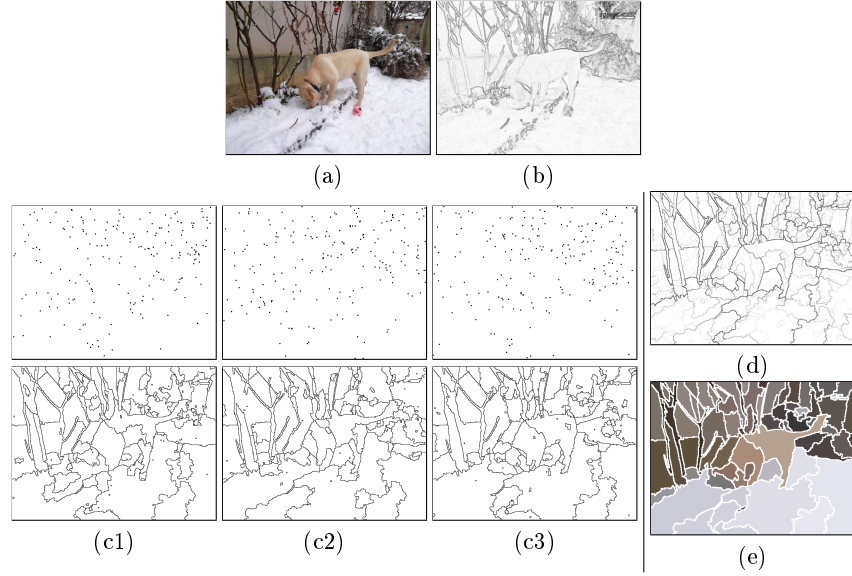
**Regionalized Poisson points.** We first consider the notion regionalized random points as well as the algorithm used to simulate a realization of  $N$  random germs associated to a spatial density.

A rather natural way to introduce uniform random germs is to generate realizations of a Poisson point process with a constant intensity  $\theta$  (i.e., average number of points per unit area). It is well known that the random number of points  $N(D)$  falling in a domain  $D$ , which is considered a bounded Borel set, with area  $|D|$ , follows a Poisson distribution with parameter  $\theta|D|$ , i.e.,  $\Pr\{N(D) = n\} = e^{-\theta|D|} \frac{(\theta|D|)^n}{n!}$ . In addition, conditionally to the fact that  $N(D) = n$ , the  $n$  points are independently and uniformly distributed over  $D$ , and the average number of points in  $D$  is  $\theta|D|$  (i.e., the mean and variance of a Poisson distribution is its parameter).

More generally, we can suppose that the density  $\theta$  is not constant; but considered as measurable function, defined in  $\mathbb{R}^d$ , with positive values. For simplicity, let us write  $\theta(D) = \int \theta(\mathbf{x})d\mathbf{x}$ . It is also known [13] that the number of points falling in a Borel set  $B$  according to a regionalized density function  $\theta$  follows a Poisson distribution of parameter  $\theta(D)$ , i.e.,  $\Pr\{N(D) = n\} = e^{-\theta(D)} \frac{(\theta(D))^n}{n!}$ . In such a case, if  $N(D) = n$ , the  $n$  are independently distributed over  $D$  with the probability density function  $\hat{\theta}(\mathbf{x}) = \theta(\mathbf{x})/\theta(D)$ . In practice, in order to simulate a realization of  $N$  independent random germs distributed over the image with the pdf  $\pi_k(\mathbf{x})$  we propose to use an inverse transform sampling method. More precisely, the algorithm to generate  $N$  random germs in an image  $m : E \rightarrow \{0, 1\}$  according to density  $\hat{\theta}(x)$  is as follows:

1. Initialization:  $m(x_i) = 0 \ \forall x_i \in E$ ;  $P = \text{Card}(E)$
2. Compute cumulative distribution function:  $cdf(x_i) = \frac{\sum_{k \leq i} \hat{\theta}(x_k)}{\sum_{k=1}^P \hat{\theta}(x_k)}$
3. for  $j = 1$  to  $N$
4.      $r_j \sim \mathcal{U}(1, P)$
5.     Find the value  $s_j$  such that  $r_j \leq cdf(x_{s_j})$ .
6.      $m(x_{s_j}) = 1$

**Marker-driven watershed transform.** Let  $g(x)$  and  $mrk(x)$  be respectively a (norm of) gradient image and a marker image. Intuitively, the associated watershed transformation [6],  $WS(g, mrk)(x)$ , produces a binary image with the contours of regions “marked” by the image  $mrk$  according to the strength of contour given by the gradient image  $g$ . The classical paradigm of watershed



**Fig. 1.** Stochastic watershed segmentation of “Custard” image: (a) color image  $f$ , (b) its color gradient  $g$ , (c) top, three realizations of random markers  $mrk_n$  regionalized from  $g$ , bottom, corresponding watershed lines  $WS(g, mrk_n)$ , (d) estimated density of contours  $pdf$ , (e) segmentation from  $d$ . In (c) and (d) images in negative for better visualization.

segmentation lays on the appropriate choice of markers, which are the seeds to initiate the flooding procedure.

**Probability density of contours using MonteCarlo simulations of watershed.** In the stochastic watershed (SW) approach [2], an opposite direction is followed, by spreading random germs for markers on the watershed segmentation. This arbitrary choice will be balanced by the use of a given number  $M$  of realizations, in order to filter out non significant fluctuations. Each piece of contour may then be assigned the number of times it appears during the various simulations in order to estimate a probability density function (pdf) of contours. In the case of uniformly distributed random germs, large regions will be sampled more frequently than smaller regions and will be selected more often. Using  $g(x)$  as density for regionalization of random germs involves sampling high contrasted image areas and it has been proved to be an appropriate choice [2]. In this case, the probability of selecting a contour will offer a trade-off between strength of the contours and size of the adjacent regions.

More precisely, given a color image  $f$  the associated SW pdf of contours is obtained as follows. Let  $\{mrk_n(x)\}_{n=1}^M$  be a series of  $M$  realizations of  $N$  spatially distributed random markers according to its gradient image  $g$ . Each realization of random germs is considered as the marker image for a watershed

segmentation of gradient image  $g$  in order to obtain the binary image

$$WS(g, mrk_n)(x) = \begin{cases} 1 & \text{if } x \in \text{Watershed lines} \\ 0 & \text{if } x \notin \text{Watershed lines} \end{cases}$$

Consequently, a set of  $M$  realizations of segmentation is computed, i.e.,  $\{WS(g, mrk_n)\}_{1 \leq n \leq M}$ . Note that in each realization the number of points determines the number of regions obtained (i.e., essential property of watershed transformation). Then, the probability density function of contours is computed by the kernel density estimation method as follows:

$$pdf(x) = \frac{1}{M} \sum_{n=1}^M WS(g, mrk_n)(x) * K_\sigma(\mathbf{x}). \quad (1)$$

Typically, the kernel  $K_\sigma(\mathbf{x})$  is a spatial Gaussian function of width  $\sigma$ , which determines the smoothing effect.

Then, the image  $pdf(x)$  can be segmented by selecting the contours of probability higher than a given contrast [2]. Fig. 1 depicts an example of color image segmentation using SW. As we can note from this example, which includes large homogenous areas, well contrasted objects as well as textured zones, the pdf and corresponding segmentation produces relatively satisfactory results. However, large homogeneous areas are oversegmented and textured zones are not always well contoured. Obviously, low contrasted areas (e.g., boundary between dog head and wall) are not properly segmented. All those are well known drawbacks of SW which have been addressed by the robust stochastic watershed (RSW) [5] or by the regional regularized stochastic watershed [12]. We adopt here an approach related to the latter one, based on a statistical model of regions.

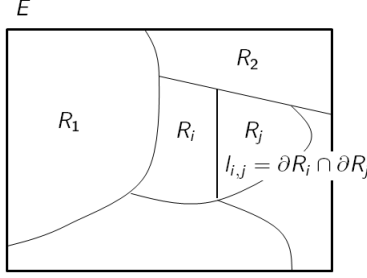
Let us remind the principle of the RSW [5] since it will be also used in the results of next section. The fundamental property of watershed is the insensitivity to the placement of seed points, which usually enables the SW segmentation to find reliably relevant boundaries, but, in the case of “false boundaries” it works to our disadvantage. The idea of RSW is to introduce a perturbation  $\epsilon_n$  (i.e., small amount of noise) into the flooding function  $g$  (i.e., gradient) at each realization  $n$ , in order to reduce the number of times that a “false boundary” will appear. More precisely, the kernel density estimator (1) becomes now

$$pdf(x) = \frac{1}{M} \sum_{n=1}^M WS(g + \epsilon_n, mrk_n)(x) * K_\sigma(\mathbf{x}), \quad (2)$$

where  $\epsilon_n(x)$  is in our experiments a zero-mean Gaussian white noise with an intensity-dependent variance from  $g(x)$ .

### 3 Multivariate Gaussian Model of Regions in SW

As we just discussed, the watershed segmentation of  $g$  from  $N$  markers (imposed minima) produces a set of thin lines dividing the image domain  $E$  into  $N$  disjoint



**Fig. 2.** Tessellation  $\tau$  of  $E$  from watershed  $WS(x)$ .

regions, denoted  $\{R_r\}_{1 \leq r \leq N}$ . This structure is called tessellation  $\tau$  of  $E$ , see Fig. 2, defined as a (finite) family of disjoint open sets (or classes, or regions)

$$\tau = \{R_r\}_{1 \leq r \leq N}, \quad \text{with } i \neq j \Rightarrow R_i \cap R_j = \emptyset,$$

such that

$$E = \cup_r R_r \bigcup WS(x) \Leftrightarrow WS(x) = E \setminus \cup_r R_r = \cup l_{i,j}.$$

The watershed lines  $WS(x)$  can be decomposed into the curves that separates the regions. More precisely, let us denote by  $l_{i,j}$  the curve (or irregular arc segment) defined as the boundary between regions  $R_i$  and  $R_j$ , i.e.,

$$l_{i,j} = \partial R_i \cap \partial R_j.$$

We obviously have  $WS = \cup l_{i,j}$ , but we also note that in the case of three (or more adjacent) regions, their boundary segments intersect at their junctions (or triple points).

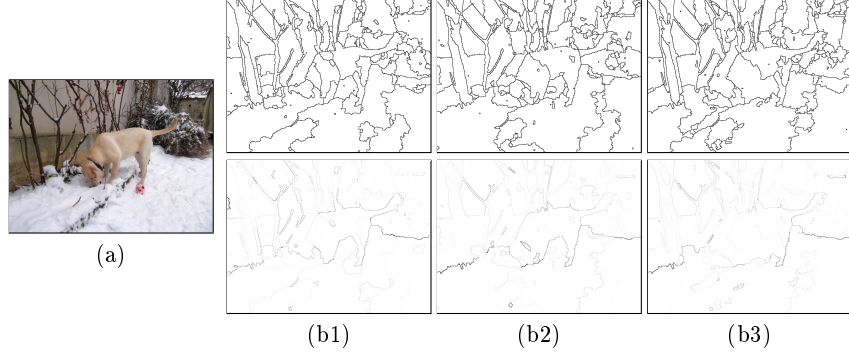
The color image values restricted to each region of the partition,  $P_i = f(R_i)$ , can be modeled by different statistical distributions. Here we focuss on a multivariate normal model.

### 3.1 Distances for multivariate normal distributions

Let us consider a family of multivariate normal distributions  $P_i$  of mean  $\mu_i$  and covariance matrix  $\Sigma_i$ , i.e.,  $P_i \sim \mathcal{N}(\mu_i, \Sigma_i)$ . Different distances are defined in the space of  $P_i$ .

**Bhattacharyya distance and Hellinger metric distance.** The Bhattacharyya distance  $D_B(P_1, P_2)$  measures the similarity of two discrete or continuous probability distributions  $P_1$  and  $P_2$ . More precisely, it computes the amount of overlap between the two statistical populations, i.e.,  

$$D_B(P_1, P_2) = -\log \int \sqrt{P_1(x)P_2(x)} dx.$$



**Fig. 3.** Given the “Custard” color image (a), (b1)-(b3) are, on the top, three realizations of SW lines  $WS(x, n)$ , at the bottom, the corresponding probability maps  $Pr(x, n)$  using the Bhattacharyya distance.

For multivariate normal distributions the Bhattacharyya distance  $D_B(P_1, P_2)$  is given by

$$D_B(P_1, P_2) = \frac{1}{8}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \frac{1}{2} \log \left( \frac{\det \boldsymbol{\Sigma}}{\sqrt{\det \boldsymbol{\Sigma}_1 \det \boldsymbol{\Sigma}_2}} \right), \quad (3)$$

where

$$\boldsymbol{\Sigma} = \frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2}.$$

Note that the first term in the Bhattacharyya distance is related to the Mahalanobis distance, both are the same when the covariance of both distributions is the same.

We have  $0 \leq D_B \leq \infty$  and it is symmetric  $D_B(P_1, P_2)$ , it is not a metric. But  $D_B$  does not obey the triangle inequality and therefore it is not a metric. Nevertheless, it can be metrized by transforming it into the following Hellinger metric distance  $D_H(P_1, P_2)$  given by  $D_H(P_1, P_2) = \sqrt{1 - \exp(-D_B(P_1, P_2))}$ , such that

$$D_H(P_1, P_2) = \sqrt{1 - \left( \frac{\det \boldsymbol{\Sigma}}{\sqrt{\det \boldsymbol{\Sigma}_1 \det \boldsymbol{\Sigma}_2}} \right)^{-1/2} e^{(-\frac{1}{4}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))}}. \quad (4)$$

Hellinger distance is an  $\alpha$ -divergence [1], which corresponds to the case  $\alpha = 0$  and it is the solely being a metric distance. Hellinger distance can be related to measure theory and asymptotic statistics. For more details on Bhattacharyya and Hellinger distances, see for instance [14].

**Wasserstein metric distance.** The Wasserstein metric is a distance function defined between probability measures on a given metric space based on the notion optimal transport [20]. Namely, the  $W_2$  Wasserstein distance between probability measures  $\mu$  and  $\nu$  on  $\mathbb{R}^n$  is  $W_2(\mu, \nu) = \inf \mathbb{E}(\|X - Y\|^2)^{1/2}$ , where

the infimum runs over all random vectors  $(X, Y) \in \mathbb{R}^n \times \mathbb{R}^n$  with  $X \sim \mu$  and  $Y \sim \nu$ . For the case of discrete distributions, it corresponds to the well-known earth mover's distance [18].

In the case of two multivariate normal distributions, the Wasserstein metric distance is obtained in a closed form as [10,19]:

$$D_W(P_1, P_2) = \sqrt{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 + \text{Tr}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 - 2\boldsymbol{\Sigma}_{1,2})}, \quad (5)$$

where

$$\boldsymbol{\Sigma}_{1,2} = \left( \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{1/2} \right)^{1/2}.$$

We note in particular that in the commutative case  $\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1$  we have

$$D_W(P_1, P_2)^2 = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 + \|\boldsymbol{\Sigma}_1^{1/2} - \boldsymbol{\Sigma}_2^{1/2}\|_F^2.$$

### 3.2 Probability density function estimation

We have now the ingredients to compute by MonteCarlo simulations the probability density function from a color image. The idea is to assign to each piece of contour  $l_{i,j}$  between regions  $R_i$  and  $R_j$  the statistical distance between the color gaussian distributions  $P_i$  and  $P_j$ :

$$\pi_{i,j} = \frac{D(P_i, P_j)}{\sum_{l_{k,l} \in WS} D(P_k, P_l)}, \quad (6)$$

where  $D(P_i, P_j)$  is any of the distances discussed above. Thus, for any realization  $n$  of SW, denoted  $WS(x, n)$ , one can compute the following image of weighted contours:

$$Pr(x, n) = \begin{cases} \pi_{i,j} & \text{if } x \in l_{i,j}^n \\ 0 & \text{if } x \notin l_{i,j}^n \end{cases}$$

where  $l_{i,j}^n$  is the boundary between regions  $R_i$  and  $R_j$  from  $WS(x, n)$ . Fig. 3 gives an example of three realizations of SW lines  $WS(x, n)$  and the corresponding probability maps  $Pr(x, n)$  using the Bhattacharyya distance.

Finally, integrating across the  $M$  realizations, the MonteCarlo estimate of the probability density function of contours is given by

$$pdf(x) = \frac{1}{M} \sum_{n=1}^M Pr(x, n) * K_\sigma(x). \quad (7)$$

Obviously, this approach is compatible with the robust stochastic watershed (RSW) variant discussed in previous section, since each realization  $n$  of the RSW produces also a tessellation of  $E$  since the image of weighted contours  $Pr(x, n)$  can be computed.

In Fig. 4 is depicted a comparison of image segmentation of ‘‘Custard’’ using gaussian region model SW and gaussian model RSW. In particular, the  $pdf(x)$

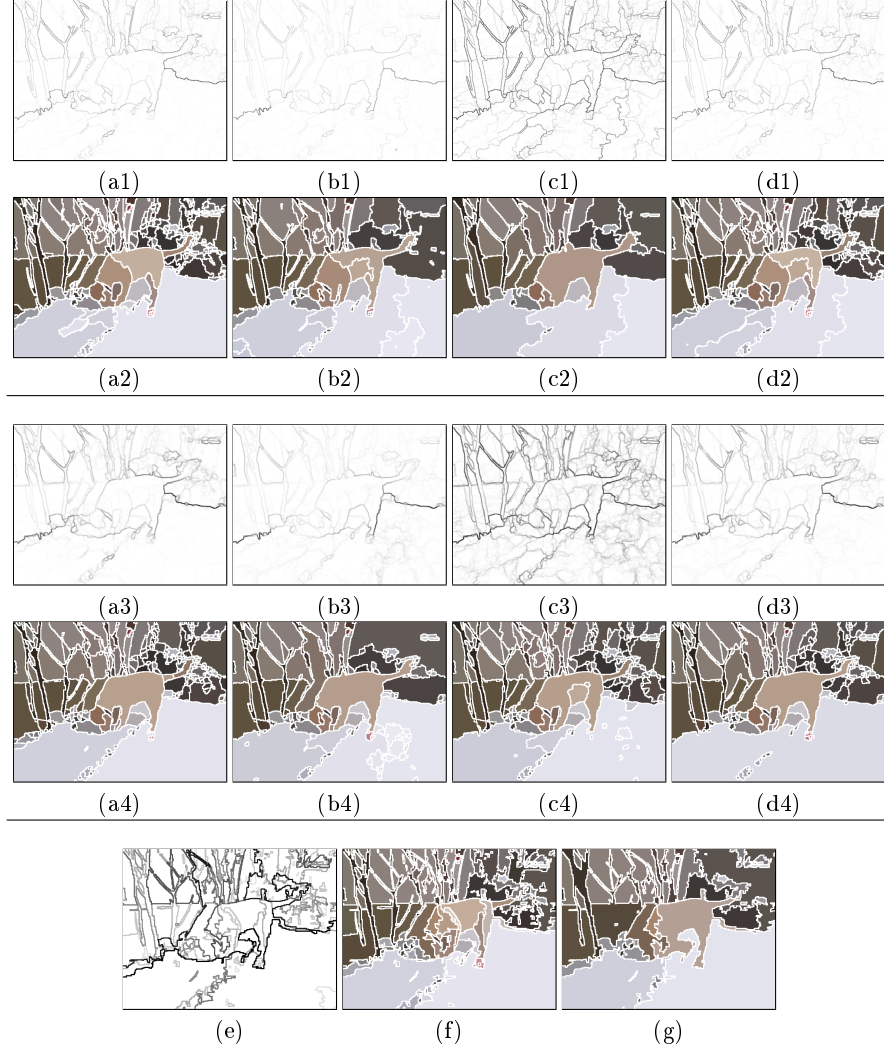


for the three considered distances (Bhattacharyya distance, Hellinger metric distance and Wasserstein metric distance) is provided, as well as the obtained segmentation by taking a probability contrast value which provides a similar degree of segmentation. It is also included the result obtained using only the color mean as the model proposed in [12]. From this example, and similar ones obtained from more experiments, Bhattacharyya distance produces good results and a better selectivity of contours than Hellinger metric distance. The results obtained by Wasserstein metric distance are also relevant but the influence of the covariance matrix is less significant, thus being closer to the results obtained using the distance between only the mean colors. Concerning the comparison between the standard SW and the RSW paradigm, it is visually observed that the gaussian model RSW produces improvements on the obtained segmentation with respect to SW.

We have also included in the figure the results obtained for this example by statistical region merging (SRM) [15], computed using MATLAB code provided in [8]. Using the scale parameter  $Q$ , we have computed nine segmentations, such that the sum of contours from the nine images can be viewed as a contours saliency function. Then, we have selected two values of  $Q$  giving segmentation similar to those of the gaussian region model SW. We observe that the detected regions are quite similar, however, the precision of contours in SW-approaches is qualitatively better for such kind of images.

## 4 Perspectives

Besides a more quantitative assessment of the performance of the proposed algorithms, other related perspectives can be considered in ongoing research. First, in addition to color information, texture at each pixel  $x$  can be described also by its structure tensor  $T(x) \in SPD(2)$ . Thus, texture at each region  $R_i$  can be described by a zero-mean gaussian distribution  $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_i)$ , where the covariance matrix is given by  $\mathbf{\Sigma}_i = |R_i|^{-1} \sum_{x \in R_i} T(x)$ . Hence, the approach presented in the paper can be also used to estimate pdf of contours from texture information or from color + texture. Second, the use of some available prior knowledge, typically represented by training images of annotated contours, could be considered in order to have a (semi-)supervised segmentation. In our framework, this goal can be formulated as a problem of distance learning in the space of multivariate normal distributions.



**Fig. 4.** Top, image segmentation of “Custard” using gaussian region model SW, where the first row is  $pdf(x)$  and second row the segmentation obtained at a given probability contrast; middle, gaussian region model RSW: (a) using only color mean [12], (b) Bhattacharyya distance, (c) Hellinger metric distance, (d) Wasserstein metric distance. The number of realization is  $M = 50$  and the number of random markers at each realization is  $N = 200$  germs. Bottom, image segmentation of “Custard” using statistical region merging (SRM) [15]: (e) sum of contours obtained from SRM using nine values of  $Q$  (256, 128, 64, 32, 16, 8, 4, 2, 1), (f) segmentation for  $Q = 128$  and (g) for  $Q = 32$ .

## References

1. S. Amari, H. Nagaoka. *Methods of information geometry*. Translations of mathematical monographs, Vol. 191, AMS, 2000.
2. J. Angulo, D. Jeulin. Stochastic watershed segmentation. In *Proc. of the ISMM'2007*, pp. 265–276, 2007.
3. J. Angulo, S. Velasco-Forero, J. Chanussot. Multiscale stochastic watershed for unsupervised hyperspectral image segmentation. In *Proc. of IEEE IGARSS'2009*, Vol. III, 93–96, 2009.
4. P. Arbelaez, M. Maire, C. Fowlkes, J. Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on PAMI*, 33(5): 898–916, 2011.
5. K.B. Bernander, K. Gustavsson, B. Selig, I.M. Sintorn, C.L. Luengo Hendriks. Improving the stochastic watershed. *Pattern Recognition Letters*, 34(9), 993–1000, 2013.
6. S. Beucher, F. Meyer. The Morphological Approach to Segmentation: The Watershed Transformation. In *Mathematical Morphology in Image Processing*, pp. 433–481, Marcel Dekker, 1992.
7. S. Boltz, F. Nielsen, S. Soatto. Earth Mover Distance on superpixels. In *Proc. of IEEE ICIP'10*, 4597–4600, 2010.
8. S. Boltz. Image segmentation using statistical region merging. <http://www.mathworks.com/matlabcentral/fileexchange/25619-image-segmentation-using-statistical-region-merging>, Retrieved, December 2014.
9. D. Comaniciu, P. Meer, P. (1999). Mean Shift Analysis and Applications. In *Proc. of ICCV'99*, Vol. 2, 1197–1203, 1999.
10. D.C. Dowson, B.V. Landau. The Fréchet distance between multivariate normal distributions. *J. Multivariate Anal.*, 12(3):450–455, 1982.
11. G. Franchi, J. Angulo. Bagging Stochastic Watershed on Natural Color Image Segmentation. In *Proc. of the ISMM'2015*, 2015.
12. F. López-Mir, V. Naranjo, S. Morales, J. Angulo. Probability Density Function of Object Contours Using Regional Regularized Stochastic Watershed. In *Proc. of IEEE ICIP'14*, 4762–4766 2014.
13. G. Matheron. *Random sets and integral geometry* Wiley, New York, 1975.
14. F. Nielsen, S. Boltz. The Burbea-Rao and Bhattacharyya centroids. *IEEE Trans. on Information Theory*, 57(8): 5455–5466, 2010.
15. R. Nock, F. Nielsen. Statistical Region Merging. *IEEE Trans. on PAMI*, 26(11):1452–1458, 2004.
16. F. Meyer, J. Stawiaski. A stochastic evaluation of the contour strength. In *Proc. of DAGM'10*, 513–522, 2010.
17. F. Meyer. Stochastic Watershed Hierarchies. In *Proc. of ICAPR'15*, 2015.
18. Y. Rubner, C. Tomasi, L.J. Guibas. A Metric for Distributions with Applications to Image Databases. In *Proc. of ICCV'98*, p. 59–66, 1998.
19. A. Takatsu. Wasserstein geometry of Gaussian measures. *Osaka J. Math.* 48(4):1005–1026, 2011.
20. C. Villani. *Optimal transport. Old and new*. Vol. 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.