MRF Energy Minimization Approach with Epitomic Textural Global Term for Image Segmentation Problems*

Elshin D., Kropotov D.
elshin.den@yandex.ru, dmitry.kropotov@gmail.com
Russia, Moscow, Lomonosov Moscow State University, Dorodnicyn Computing Centre of RAS

The epitomic model by Jojic et al. is the state-of-the-art image generative model using image textural information. In the paper we adapt this model for solving image segmentation problems. The resulting epitomic textural potential is used then for image segmentation within Markov random field energy minimization approach. We propose an efficient iterative scheme for solving this energy minimization problem with the global epitomic potential. The experimental results on toy texture segmentation problem and on a real problem of mouse brain slices segmentation show the applicability of our approach. We also compare the proposed potential with the state-of-the-art texton boost descriptor.

KEYWORDS: image segmentation, texture segmentation, Markov random field, probabilistic models, texture descriptors, epitomes, gene expression.

Introduction

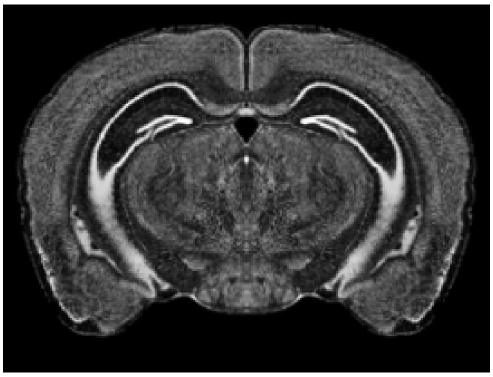
Gene expression analysis is a promising technique in modern neurobiology for understanding cognitive processes in animal brains. The standard procedure here is brain decapitation, freezing, slicing and staining the obtained thin slices using Nissl method (resulting in histological images) and ISH technique for recovering particular gene expression [3]. The consequent statistical analysis of gene expression in different brain anatomical structures requires automatic segmentation of slice histological images into these anatomical structures (hippocampus, cortex, thalamus, etc.).

In the paper we consider this segmentation problem for mouse brain slice images from Allen Brain Atlas [7]. This atlas contains 132 coronal mouse brain slices both in histological (see Fig.1,a) and anatomical view (see Fig.1,b). The anatomical segmentation is made by experts. In these histological images there are almost no visible borders between different anatomical structures. That is why it is important to consider textural features of histological images for their successful segmentation into anatomical structures.

Markov random field energy minimization is the state-of-the-art approach for image segmentation [2]. In this approach the energy function combines a set of potentials (usually only unary and pairwise), each corresponding to different image pixel features such as color, position, texture, etc. as well as similarity of segment labels for neighbouring pixels. The energy function containing only unary potentials and pairwise Potts potentials can be effectively minimized using α-expansion algorithm [5].

In this paper we focus on textural potentials. In the literature there were proposed a lot of different textural descriptors, for example, SIFT descriptors [4], texton boost descriptors [2], local binary patterns [8] and many others. The main drawback of all these descriptors is necessity to tune a set of parameters by hand for appropriate segmentation results. For example, in texton boost approach it is important to choose a good set of image filters as well as different parameters of textons. For LBP descriptors we need to choose a distance from central pixel to

The work is supported by the Russian Foundation for Basic Research (projects Nos. 10-01-90015, 10-01-90419, 09-01-00409).



(a)

(b)

Fig. 1. Examples of mouse brain slice images from Allen Brain Atlas. (a) – histological image, (b) – its expert segmentation into anatomical structures.

a set of patches, their number and sizes, angles between them, etc. In this paper we propose an epitomic textural potential, which requires almost no parameters for tuning by user. As a result, this potential can be quickly applied for different image segmentation problems.

For construction of epitomic textural potential we adapt the epitomic probabilistic model, proposed in [1]. The resulting textural potential depends on the whole image and its segmentation, and so this potential is global. In the general case minimizing MRF energy function with global potentials could be a very hard problem. However, for this particular global potential it is possible to propose an effective iterative optimization technique, where on each iteration the energy function is bounded by a function with unary and pairwise potentials, which can be easily minimized by α-expansion algorithm [5].

The paper is structured in the following way. First we briefly introduce the standard epitomic model and present our modification. Then we discuss acceleration of training procedure for the new probabilistic model. After that we present a segmentation algorithm for test image using our textural potential within MRF energy minimization approach. Here we also give an iterative procedure for such minimization. In experimental section we first show the results of our approach for toy texture segmentation problem and then compare our approach with the state-of-the-art texton boost method for mouse brain slice segmentation problem for data from Allen Brain Atlas. In the last section we conclude and point out directions for future research.

Epitomic Probabilistic Model

Consider the classic epitomic approach for image generation [1]. Suppose we have a training grayscale image Z of size $N\times M$. The epitome e for the image Z is its condensed version of size $N_e\times M_e$, $N_e < N, M_e < M$, that reflects textural peculiarities of the original image (see Fig. 2). Every pixel t of the epitome e contains two components: intensity level μ_t and its variance φ_t.

Hereinafter we use the following notation: image Z is provided as a set of P patches $\mathcal{Z} = \{Z_k\}_{k=1}^P$ (which possibly overlap); for every patch Z_k we define a set of its pixel coordinates as S_k, so that $Z_k = \{z_{k,i} \mid i \in S_k\}$. For each patch Z_k a hidden mapping T_k is defined that maps coordinates of Z_k (S_k) to the coordinates of same size patch in e. According to [1], it is possible to consider mappings of different complexity (matching an image patch to a number of smaller epitome patches, etc), but we use only patch-to-patch mappings.

Given the epitome $e = (\mu, \varphi)$ and the mapping T_k, Z_k is generated by copying the appropriate pixels μ from the epitome and adding Gaussian noise with variance φ:

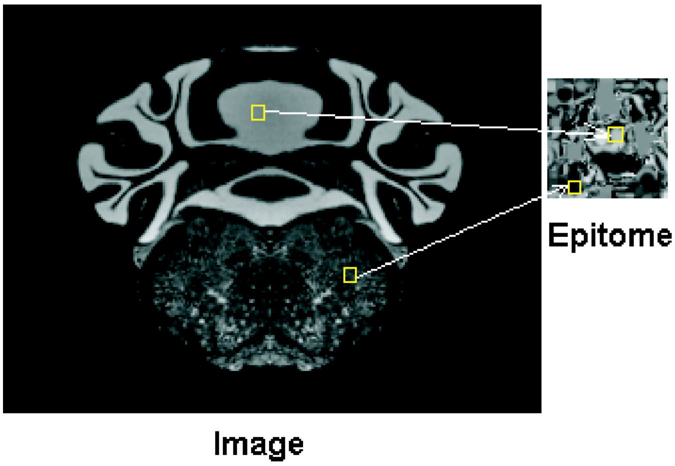


Fig. 2. Example of grayscale image and its epitome. Arrows show the correspondence between two patches of the image and the epitome.

$$\mathsf{P}(Z_k \,|\, T_k, e) = \prod_{i \in S_k} \mathcal{N}(z_{k,i} \,|\, \mu_{T_k(i)}, \varphi_{T_k(i)}). \tag{1}$$

It's assumed that the patches are generated independently, so the joint distribution is:

$$\mathsf{P}(\{Z_k, T_k\}_{k=1}^P, e) = \mathsf{P}(e) \prod_{k=1}^{P} \mathsf{P}(T_k)\, \mathsf{P}(Z_k \,|\, T_k, e). \tag{2}$$

In the following we assume both priors $\mathsf{P}(T_k)$ and $\mathsf{P}(e)$ to be flat.

The model (1)-(2) is a generative model for new images and hence can't be applied directly for solving image segmentation problem. Suppose we are given an image Z split into a set of P patches $\{Z_k\}_{k=1}^P$ and its segmentation Q into L classes that cause agreed segmentation of patches $\{Q_k\}_{k=1}^P$, $Q_k(i) \in \{1, \ldots, L\}$. We adapt the model (1)-(2) for this case in the following way. First we add a new component ψ into epitome $(e, \psi) = (\mu, \varphi, \psi)$, where $\psi_{t,l}$ is probability of class l of the t-th element in the epitome.

Then, the modified model becomes the following:

$$\mathsf{P}(\{Z_k, Q_k, T_k\}_{k=1}^P, e, \psi) = \mathsf{P}(e)\, \mathsf{P}(\psi \,|\, \alpha) \prod_{k=1}^{P} \mathsf{P}(T_k)\, \mathsf{P}(Z_k, Q_k \,|\, T_k, e), \tag{3}$$

$$\mathsf{P}(Z_k, Q_k | T_k, e, \psi) = \prod_{i \in S_k} \mathcal{N}(z_{k,i} | \mu_{T(i)}, \varphi_{T(i)}) \psi_{T(i), Q_k(i)}, \tag{4}$$

$$\mathsf{P}(\psi \,|\, \alpha) = \prod_{t} \mathrm{Dir}(\psi_t \,|\, \alpha). \tag{5}$$

Here $\mathrm{Dir}(\psi_t \,|\, \alpha) \propto \psi_{t,1}^{\alpha-1} \psi_{t,2}^{\alpha-1} \ldots \psi_{t,L}^{\alpha-1}$ is Dirichlet prior with scalar parameter α. This prior prevents zero probability of some classes in epitome pixels.

Learning epitome parameters (e, ψ) in the model (3)-(5) by maximizing posterior distribution

$$\mathsf{P}(e, \psi | Z, Q, \alpha) \to \max_{e, \psi}$$

can be done using EM-algorithm with hidden variables T_k:

E-step:

$$q(T_k) = \mathsf{P}(T_k \mid Z_k, Q_k, e, \psi, \alpha) \propto \mathsf{P}(Z_k, Q_k \mid T_k, e, \psi) = \prod_{i \in S_k} \mathcal{N}(z_{k,i} \mid \mu_{T_k(i)}, \varphi_{T_k(i)}) \psi_{T_k(i), Q_k(i)}. \quad (6)$$

M-step:

$$\mu_t = \frac{\sum_{k=1}^P \sum_{i \in S_k} \sum_{T_k : T_k(i) = t} q(T_k) z_{k,i}}{\sum_{k=1}^P \sum_{i \in S_k} \sum_{T_k : T_k(i) = t} q(T_k)}, \quad (7)$$

$$\varphi_t = \frac{\sum_{k=1}^P \sum_{i \in S_k} \sum_{T_k : T_k(i) = t} q(T_k)(z_{k,i} - \mu_t)^2}{\sum_{k=1}^P \sum_{i \in S_k} \sum_{T_k : T_k(i) = t} q(T_k)}, \quad (8)$$

$$\psi_{t,l} = \frac{\alpha_l - 1 + \sum_{k=1}^P \sum_{i : Q_k(i) = l} \sum_{T_k : T_k(i) = t} q(T_k)}{\sum_j \alpha_j - K + \sum_{k=1}^P \sum_{i \in S_k} \sum_{T_k : T_k(i) = t} q(T_k)}. \quad (9)$$

For learned epitome parameters (e, ψ) segmentation Q of a test image Z can be obtained by maximizing the posterior distribution

$$\mathsf{P}(Q|Z, e, \psi) \to \max_Q \ \Leftrightarrow \ \prod_k \mathsf{P}(Z_k, Q_k | e, \psi) \to \max_{\{Q_k\}}. \quad (10)$$

Here maximization is performed w.r.t. agreed segmentations $\{Q_k\}_{k=1}^P$. However, in the following we do not solve the problem (10) for image segmentation, but rather combine textural potential $-\sum_{k=1}^P \log \mathsf{P}(Z_k, Q_k | e, \psi)$ with other potentials in MRF energy minimization framework.

Epitome Training Acceleration

The proposed EM-algorithm (6)-(9) for learning epitome parameters (μ, φ, ψ) requires a considerable calculation time. For example, for training image of size 300×200, epitome size 60×60 and patch size 5×5, epitome learning with the direct implementation of formulas (6)-(9) leads to several hours of calculation. Hence for practical purposes the effective implementation of the EM-algorithm is required.

Consider an acceleration procedure for E-step (6). For this reason consider one patch Z and epitome parameters (μ, φ). E-step requires calculating the following matrix C of size $(N_e - D) \times (M_e - D)$:

$$C_{ts} = \sum_{p,q=1}^{D} \log \mathcal{N}(z_{pq} | \mu_{t+p, s+q}, \varphi_{t+p, s+q}).$$

Here (t, s) – position in the epitome, z_{pq} – pixel intensity in the patch Z in position (p, q) and D is patch size. Note that in this section we use two-dimensional indexation for pixels. Then calculation of the matrix C can be done using a set of convolution operations:

$$C_{ts} = -\frac{D^2}{2} \log 2\pi - \frac{1}{2} \sum_{p,q=1}^{D} \left[\frac{(z_{pq} - \mu_{t+p,s+q})^2}{\varphi_{t+p,s+q}} + \log \varphi_{t+p,s+q} \right] =$$

$$- \frac{D^2}{2} \log 2\pi - \frac{1}{2} \sum_{p,q=1}^{D} \left[z_{pq}^2 \frac{1}{\varphi_{t+p,s+q}^2} - 2 z_{pq} \frac{\mu_{t+p,s+q}}{\varphi_{t+p,s+q}^2} + \frac{\mu_{t+p,s+q}^2}{\varphi_{t+p,s+q}^2} + \log \varphi_{t+p,s+q} \right] =$$

$$- \frac{D^2}{2} \log 2\pi - \frac{1}{2} \Big[H_1 * \mathrm{rot180}(Z^2) - 2 H_2 * \mathrm{rot180}(Z) + H_3 * I \Big].$$

Here $*$ stands for two-dimensional convolution, $\text{rot180}(Z)$ – anticlockwise rotation of the matrix Z on 180°, $(Z^2)_{pq} = z_{pq}^2$, $(H_1)_{ts} = 1/\varphi_{ts}$, $(H_2)_{ts} = \mu_{ts}/\varphi_{ts}$, $(H_3)_{ts} = \mu_{ts}^2/\varphi_{ts} + \log\varphi_{ts}$, I – matrix with all unity elements. The convolution operation has effective implementation in the most of programming languages.

Test Image Segmentation

Consider a test image segmentation procedure. Suppose we have a test image with N pixels $Z = \{Z_i\}_{i=1}^N$ and we need to find its segmentation $Q = \{Q(i)\}_{i=1}^N$, into L classes, $Q(i) \in \{1,\ldots,L\}$.

For solving this problem we follow Markov random field energy minimization approach:

$$E(Q,Z) = \alpha_{col}\sum_i U_{col}(Z(i),Q(i)) + \alpha_{pos}\sum_i U_{pos}(Z(i),Q(i)) + \\ + \alpha_{Potts}\sum_{(i,j)\in\mathcal{E}} U_{Potts}(Q(i),Q(j)) + \alpha_{ep}U_{ep}(Z,Q) \to \min_Q. \quad (11)$$

Here U_{col} and U_{pos} are unary color and position potentials, U_{Potts} is pairwise Potts potential, U_{ep} is global epitomic potential, $\alpha_{col}, \alpha_{pos}, \alpha_{Potts}, \alpha_{ep}$ are some positive parameters. Hereinafter we use 4-pixel neighbourhood system $\mathcal{E}$.

Color potential U_{col} is determined using Gaussian mixture for each class:

$$\mathsf{P}(z|l) = \sum_{k=1}^{N(l)} w_k^l \mathcal{N}(z|\mu_k^l, \sigma_k^l),\ \sum_k w_k^l = 1,\ w_k^l \geqslant 0,$$
$$U_{col}(Z(i),Q(i)) = -\log\mathsf{P}(Z(i)|Q(i)).$$

These L Gaussian mixtures are restored using the standard EM-algorithm for training images.

Position potential U_{pos} is defined in the following way:

$$\mathsf{P}(l|i) = \frac{N_{i,l}}{N_i},$$
$$U_{pos}(Z(i),Q(i)) = -\log\mathsf{P}(Q(i)|i).$$

Here we consider some vicinity for each pixel i and calculate $N_{i,l}$ – the number of pixels from class l in the vicinity for training images and N_i – the total number of pixels in the vicinity for training images.

Potts potential U_{Potts} gives some fee for neighboring pixels of different classes:

$$U_{Potts}(Q(i),Q(j)) = \begin{cases} 1, & \text{if } Q(i) \neq Q(j); \\ 0, & \text{if } Q(i) = Q(j). \end{cases}$$

Epitomic potential U_{ep} is determined from the problem (10):

$$U_{ep}(Z,Q) = -\sum_k \log\mathsf{P}(Z_k, Q_k|e,\psi).$$

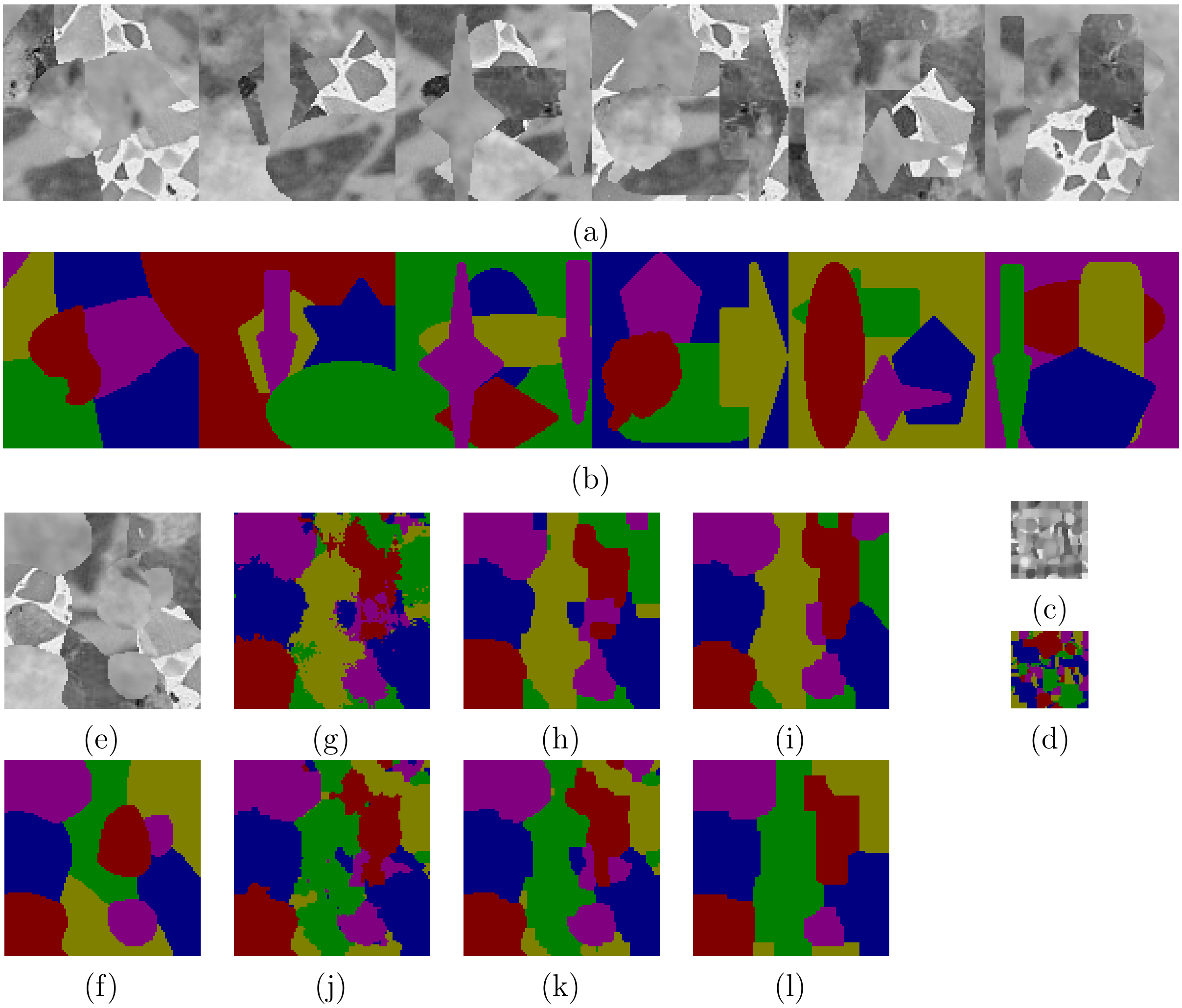


Fig. 3. Results for texture segmentation problem. (a) – train image, (b) – its true segmentation, (c) – learned epitome, (d) – the most probable classes for epitome, (e) – test image, (f) – its true segmentation, (g) – segmentation using textons only (accuracy 55%), (h) – segmentation using textons and Potts (accuracy 56.5%), (i) – segmentation using textons, color and Potts (accuracy 59%), (j) – segmentation using epitome only (accuracy 74%), (k) – segmentation using color only (accuracy 45%), (l) – segmentation using epitome, color and Potts (accuracy 78%).

The problem (11) can't be solved directly in an efficient way. Here we propose to bound global epitomic potential using Jensen inequality:

$$U_{ep}(Z,Q) = -\sum_k \log \mathsf{P}(Z_k, Q_k \,|\, e, \psi) \leqslant \sum_k \sum_{i \in S_k} \sum_{T_k} q(T_k)[\log \psi_{T_k(i),Q_k(i)} + \log \mathcal{N}(z_{k,i} \,|\, \mu_{T_k(i)}, \varphi_{T_k(i)})] +$$
$$+ \sum_k \mathbb{E}_{q(T_k)\log q(T_k)} - \sum_k \mathbb{E} \log \mathsf{P}(T_k) = \sum_i \sum_{k: i \in S_k} \sum_{T_k} q(T_k)[\log \psi_{T_k(i),Q_k(i)} + \log \mathcal{N}(z_{k,i} \,|\, \mu_{T_k(i)}, \varphi_{T_k(i)})] +$$
$$+ \sum_k \mathbb{E}_{q(T_k)\log q(T_k)} - \sum_k \mathbb{E} \log \mathsf{P}(T_k) = \sum_i U_{ep,b}(Z(i), Q(i)) + \sum_k \mathbb{E}_{q(T_k)\log q(T_k)} - \sum_k \mathbb{E} \log \mathsf{P}(T_k).$$

This upper bound combines unary potential $U_{ep,b}$ with an expression that doesn't depend on segmentation Q. Uniting this upper bound with other potentials gives upper bound for energy $E(Z,Q)$.

The resulting iterative procedure for solving the problem (11) is the following: on each iteration first we minimize the proposed upper bound w.r.t. distribution $q(T_k)$ with fixed segmentation Q (the solution is given by (6)) and then minimize the upper bound w.r.t. segmentation Q with fixed $q(T_k)$ using α-expansion algorithm [5]. The procedure is guarantied to converge since the upper bound for energy $E(Z, Q)$ is bounded from below and monotonically decreases.

Experiments

We consider two settings for experimental evaluation of the proposed textural potential. First we take a toy texture segmentation problem and then consider mouse brain slice segmentation problem.

In the first setting the images are compositions of different textures (see Fig. 3). All images have size 100×100, epitome size is 40×40, patch size is 5×5. We compare our epitomic potential with texton boost potential in different combinations of color, position and Potts potentials. The parameters α were tuned by brute force maximizing accuracy on the train image. The experimental results are shown on Fig. 3. It's easy to see that epitomic potential outperforms texton boost. Besides, the epitomic potential itself works almost as good as its combination with color and Potts.

In the second setting we compare epitomic potential with texton boost potential for mouse brain slice segmentation problem. In the experiment we take three neighboring slices from atlas, combine the first and the third slice in a train image and use the second slice as a test image. The obtained results are shown on Fig. 3. Similar to the first setting the epitomic potential outperforms texton boost.

Conclusion

We propose a new kind of textural potential in MRF energy minimization approach for image segmentation. Comparing to the state-of-the-art textural potentials such as texton boost and local binary pattern, a new one requires almost no parameters for tuning by hand. Hence this potential can be readily used for a wide class of image segmentation problems. Experimental results with epitomic and texton boost potentials show the superiority of the first one for two image segmentation settings.

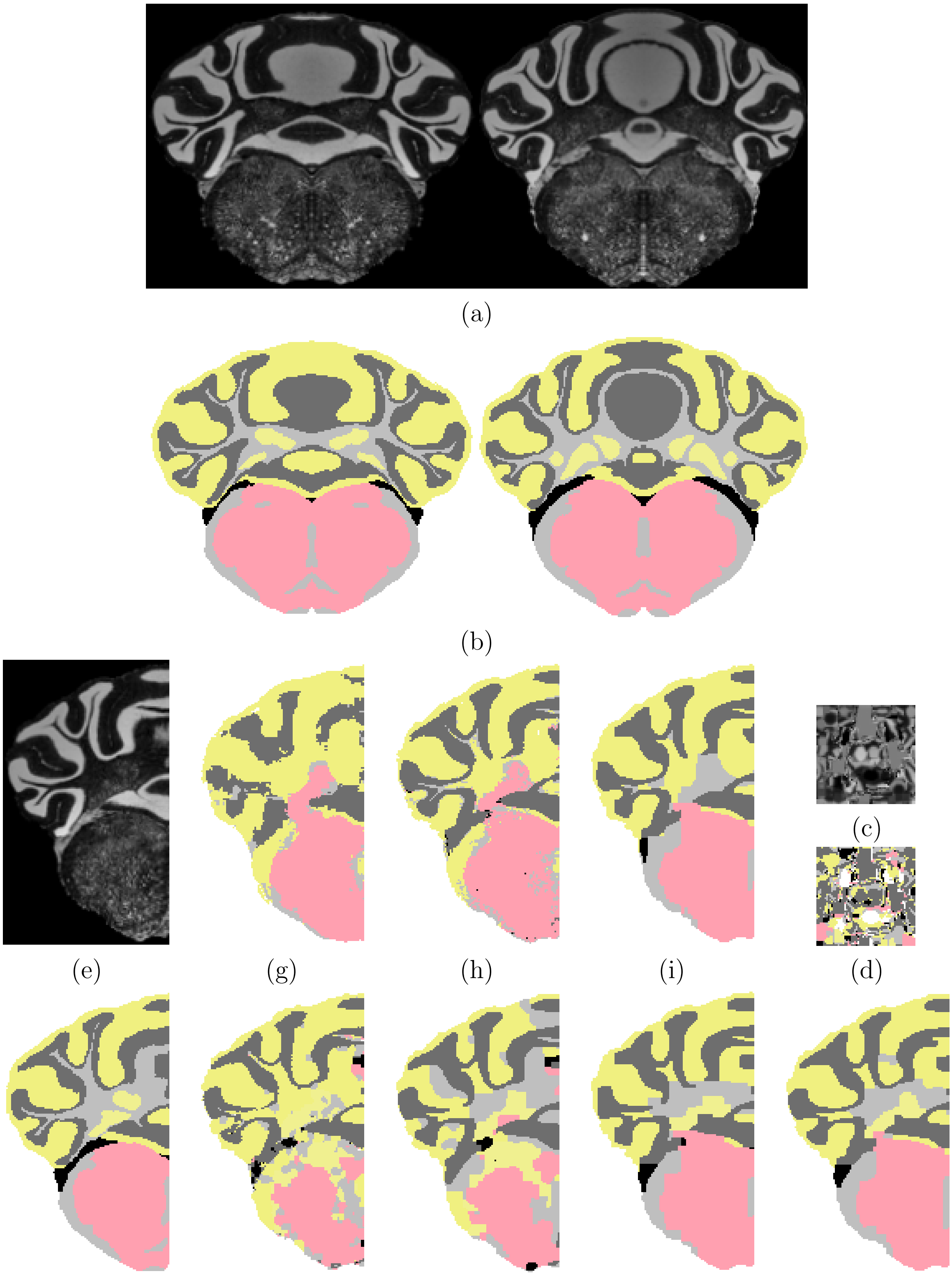
In the context of energy minimization approach the proposed epitomic potential is a global term. In the general case minimizing energy functions with global terms is a very hard optimization problem. However, here we proposed an effective iterative optimization procedure for minimizing energy with global epitomic potential. Possibly this approach could be expanded for a wider class of global potentials.

One of the important aspects in MRF energy minimization is choosing the coefficients α for combining potentials in the energy function. In this paper the coefficients were tuned by brute search, that requires significant time for training. We suppose to apply here structural learning methods, for example, structural support vector machine [6] as a direction for future research. We also plan to upgrade position potentials for the case of thin elongated classes in pixel vicinity.

References

[1] *Jojic. N., Frey. B., Kannan. A.* Epitomic analysis of appearance and shape // Int'l. Conf. on Computer Vision, 2003.

[2] *Shotton. J., Winn. J., Rother. C., Criminisi. A.* TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context // Int'l Journal on Computer Vision, 2009.



(a)

(b)

(e) (g) (h) (i) (c) (d)

(f) (j) (k) (l) (m)

Fig. 4. Results for mouse brain slice segmentation problem. (a) – train image, (b) – its true segmentation, (c) – learned epitome, (d) – the most probable classes for epitome, (e) – test image, (f) – its true segmentation, (g) – segmentation using textons only (accuracy 75%), (h) – segmentation using textons and color (accuracy 80%), (i) – segmentation using textons, color, position and Potts (accuracy 84%), (j) – segmentation using epitome only (accuracy 76%), (k) – segmentation using color and Potts (accuracy 74%), (l) – segmentation using color, position and Potts (accuracy 86%), (m) – segmentation using epitome, color, position and Potts (accuracy 88%).

[3] *Ng L., Pathak S.D., Cuan C., Lau C., Dong H., Sodt A., Dang C., Avants B., Yushkevich P., Gee J.C., Haynor D., Lein E., Jones A., Hawlyrycz M.* Neuroinformatics for Genome-Wide 3D Gene Expression Mapping in the Mouse Brain // IEEE Trans. Comp. Biol. and Bioinf. 2007. — Vol. 4, No. 3.

[4] *Lowe D.G.* Object recognition from local scale-invariant features // Int'l Conf. on Computer Vision, 1999.

[5] *Boykov. Y., Veksler. O., Zabih. R.* Fast Approximate Energy Minimization via Graph Cuts // IEEE Trans. Pat. Anal. Mach. Intell. 2001.

[6] *Tsochantaridis I., Joachims T., Hofmann T., Altun Y.* Large Margin Methods for Structured and Interdependent Output Variables // J. Mach. Learn. Res. 2005. — Vol. 6. — Pp. 1453–1484.

[7] Allen Brain Atlas (Internet). Allen Institute for Brain Science, Seattle, 2008. — http://www.brain-map.org

[8] *Ojala T., Pietikäinen M., Mäenpää T.* Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns // IEEE Trans. Pat. Anal. Mach. Intel., 2002. — Vol. 24, No. 7. — Pp. 971–987.