2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

BSIK-SVD: A DICTIONARY-LEARNING ALGORITHM FOR BLOCK-SPARSE REPRESENTATIONS

Yongqin Zhang, Jiaying Liu**, Mading Li, Zongming Guo*

Institute of Computer Science and Technology, Peking University, Beijing 100871, China

ABSTRACT

Sparse dictionary learning has attracted enormous interest in image processing and data representation in recent years. To improve the performance of dictionary learning, we propose an efficient block-structured incoherent K-SVD algorithm for the sparse representation of signals. Without relying on any prior knowledge of the group structure for the input data, we develop a two-stage agglomerative hierarchical clustering method for block sparse representations. This clustering method adaptively identifies the underlying block structure of the dictionary under the restricted conditions of both a maximal block size and a minimal distance between the blocks. Furthermore, to meet the constraints of both the upper bound and the lower bound of the mutual coherence of dictionary atoms, we introduce a regularization term for the objective function to suppress the block coherence of the overcomplete dictionary. The experiments on synthetic data and real images demonstrate that the proposed algorithm has lower representation error, higher visual quality and better reconstructed results than other state-of-the-art methods.

***Index Terms*—** Dictionary learning, sparse coding, block sparsity, sparse representation

1. INTRODUCTION

In the past decade, the synthesis model has been a very popular approach for sparse representations [1, 2]. Such a model assumes that a signal $\mathbf{X} \in \mathbb{R}^d$ can be composed of a linear combination of a few atoms from a given dictionary $\mathbf{D}^{d\times K}$. The dictionary $\mathbf{D}$ satisfies $\|\mathbf{X} - \mathbf{DA}\|_p \leqslant \varepsilon$, where ε is an arbitrarily small positive number. The vector $\mathbf{A} \in \mathbb{R}^K$ denotes the sparse representation coefficients of the signal $\mathbf{X}$. Indeed, this general problem has been shown to be NP-hard. So far, the development of many sub-optimal solutions of the above model has concentrated on estimating the sparse representation $\mathbf{A}$ from a corrupted signal $\mathbf{X}$ (or sparse coding) and inferring the dictionary $\mathbf{D}$ from signal examples.

The group sparse coding [3] and block-sparse signals [4] have become the focus of research in the fields of sparse representations these years. Very recently, the optimized K-SVD method [5] has appeared, which adopts the sparse agglomerative clustering (SAC) approach to identify the block structure of atoms in the dictionary. However, the major drawbacks of the SAC approach are its susceptibility to errors in the initial steps that propagate all the way to its final output. Although the constraints of both the upper bound [6] and the lower bound [7] of the mutual coherence of dictionary atoms were pointed out [8], the coherence of the dictionary trained by K-SVD [1] or its variant [5, 9] has received no attention in the reported literatures. In this paper, we focus on the dictionary learning of the synthesis model and propose a block-structured incoherent K-SVD (BSIK-SVD) algorithm for the performance improvement of sparse representations of signals and images. The main contributions of our work can be summarized as follows: 1) the two-stage agglomerative hierarchical clustering technique is employed to identify the inherent block structures of the dictionary, and 2) the proposed BSIK-SVD framework incorporates a regularization constraint regulating the intra-block coherence of dictionary atoms. The purpose of this work is to find an efficient block-structured incoherent dictionary. The advantages of this optimized dictionary include compact representation, robust stability under the noise, and the rapid convergence.

The rest of this paper is organized as follows. Section 2 provides the detailed descriptions of the proposed algorithm. The numerical simulation and experiments are given in Section 3. The conclusions are drawn in Section 4.

2. DICTIONARY OPTIMIZATION

Many problems in signal and image processing can be cast as inverting the linear system:

$$\mathbf{X} = \mathbf{DA} + \upsilon, \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{d\times L}$ is the vectors of noisy observations (or measurements), $\mathbf{D} \in \mathbb{R}^{d\times K}$ with $d < K$ is a bounded linear operator, $\mathbf{A} \in \mathbb{R}^{K\times L}$ are the data to recover, and υ is an additive noise with bounded variance.

For a given set of the signals $\mathbf{X} \in \mathbb{R}^{d\times L}$, our goal is to find a dictionary $\mathbf{D} \in \mathbb{R}^{d\times K}$ whose atoms are block-

---

$*$ Corresponding author

This work was supported by National Natural Science Foundation of China under contract No.61201442, National Key Technology R&D Program of China under Grant 2012BAH18B03 and China Postdoctoral Science Foundation funded project under contract No.2013M530481.

structured and incoherent, which ensures that the signals can be accurately reconstructed through a computationally efficient algorithm. Then we formulate the following optimization problem:

$$\begin{aligned} &\arg\min_{\mathbf{D},\mathbf{b},\mathbf{A}} \left\{ \|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 + \lambda\Phi(\mathbf{D}) \right\}, \\ &\text{s.t.} \quad \|\alpha_i\|_{0,\mathbf{b}} \leqslant \kappa, \forall i, \quad |b_j| \leqslant s, j \in \mathbf{b}, \end{aligned} \tag{2}$$

where α_i is a column vector of sparse matrix $\mathbf{A}$, κ is the known sparsity level, $\mathbf{b}$ is a block structure with maximal block size s, $b_j = \{i \in 1, \cdots, K | b[i] = j\}$ is the set of indices for the block j, λ is a balance parameter and the regularization term on the mutual coherence of $\mathbf{D}$ is defined as:

$$\Phi(\mathbf{D}) = \sum_{j=1}^{B} \left(\sum_{p,q \in b_j, p \neq q} \left\| \varphi_p^T \varphi_q \right\|^2 \right), \tag{3}$$

where B is the number of the blocks, φ_p and φ_q denote the different atoms in each block b_j of the current dictionary, respectively.

To solve Problem (2), we employ the coordinate relaxation technique to minimize the objective function based on alternating $\mathbf{A}$ and $\mathbf{D}$. For the given $\mathbf{D}^{(0)}$ and $\mathbf{b}^{(0)}$, the sparse representation is initialized as the solution $\mathbf{A}^{(0)}$ to Problem (4) over $\mathbf{A}$, which is solved by using the OMP method with $\kappa \times s$ instead of κ non-zero entries. At each iteration t, the dictionary $\mathbf{D}^{(t-1)}$ is first fixed so that Problem (2) is reduced to Problem (4) as follows:

$$\begin{aligned} &\left[\mathbf{b}^{(t)}, \mathbf{A}^{(t)}\right] = \arg\min_{\mathbf{b},\mathbf{A}} \left\| \mathbf{X} - \mathbf{D}^{(t-1)}\mathbf{A} \right\|_F^2, \\ &\text{s.t.} \quad \|\alpha_i\|_{0,\mathbf{b}} \leqslant \kappa, \forall i. \end{aligned} \tag{4}$$

Then the BOMP method is used to calculate the sparse representation $\mathbf{A}^{(t-1)}$ by solving Problem (4).

Next, the two-stage clustering technique identifies the intrinsic block structure $\mathbf{b}^{(t)}$ of the block sparse representation $\mathbf{A}^{(t-1)}$, and then the BOMP method is used to update $\mathbf{A}^{(t)}$ again. That is, Problem (4) turns into finding a block structure $\mathbf{b}^{(t)}$ that is used for sparse coding of $\mathbf{A}$. To update the block structure $\mathbf{b}^{(t)}$, the cost function is minimized as follows:

$$\begin{aligned} &\mathbf{b}^{(t)} = \min_{\mathbf{b}} \sum_{j \in [1,B]} \left| \omega_j \left(\mathbf{A}^{(t-1)}, \mathbf{b} \right) \right|, \\ &\text{s.t.} \ |b_j| \leqslant s, j \in [1, B], \end{aligned} \tag{5}$$

where $|\omega_j|$ is the number of non-zero values in ω_j.

To solve Problem (5), the BK-SVD+SAC method [5] assumed that there is a maximal block size s for the block structure $\mathbf{b}$ in the block-sparsity constraint of $\mathbf{A}$. They adopted the agglomerative clustering algorithm [10] to extract the intrinsic block structure of the dictionary. However, in fact the number of most similar atoms may be larger than the predetermined fixed number s. The agglomerative clustering algorithm [10] has the disadvantage that the blocks may have been incorrectly grouped at an early stage for lack of a relocation provision of them. To overcome these drawbacks, we propose a two-stage clustering approach based on agglomerative hierarchical clustering. The two stages have the similar procedures except for the constraint. Since the initial block structure is estimated in the first stage, the agglomerative clustering accuracy will be much improved in the second stage so that the final clustering result is much better than the traditional agglomerative clustering algorithm [10]. For each loop in the first stage we compute the distances between blocks and find the closest pair $[j_1^*, j_2^*]$ such that

$$\begin{aligned} &[j_1^*, j_2^*] = \arg\min_{j_1 \neq j_2} F_{dist}(\omega_{j_1}, \omega_{j_2}), \\ &\text{s.t.} \quad F_{dist}(\omega_{j_1}, \omega_{j_2}) \leqslant V_{HT}, \end{aligned} \tag{6}$$

where V_{HT} is a threshold value and the distance $F_{dist}(\omega_{j_1}, \omega_{j_2}) = \sum_{i=1}^{L} |\omega_{j_1} - \omega_{j_2}|$ is the city block metric. Then the blocks j_1^* with j_2^* are combined by enforcing $\forall i \in b_{j_2} : b[i] \leftarrow j_1$, $\omega_{j_1} \leftarrow \{\omega_{j_1} \cup \omega_{j_2}\}$ and $\omega_{j_2} \leftarrow \phi$. This loop procedure is executed repeatedly until no blocks can be merged under the constraint on the preset threshold value. Subsequently, in the second stage the estimated block structure in the first stage is used as the starting values. Like the traditional agglomerative clustering algorithm [10], at each loop we find the closest pair $[j_1^*, j_2^*]$ of blocks satisfying the following formula:

$$\begin{aligned} &[j_1^*, j_2^*] = \arg\max_{j_1 \neq j_2} |\omega_{j_1} \cap \omega_{j_2}|, \\ &\text{s.t.} \quad |b_{j_1}| + |b_{j_2}| \leqslant s. \end{aligned} \tag{7}$$

Then we aggregate the closest pair of blocks j_1^* and j_2^* and update the block structure $\mathbf{b}^{(t)}$. This loop procedure is repeated until no blocks can be merged without breaking the constraint on the block size. After the block structure $\mathbf{b}^{(t)}$ and the sparse representation $\mathbf{A}^{(t)}$ are available, according to the definition of block coherence of the dictionary [4], we can reformulate Problem (2) in another form as follows:

$$\begin{aligned} &\left[\mathbf{D}^{(t)}, \mathbf{A}^{(t)}\right] = \arg\min_{\mathbf{D},\mathbf{A}} \left\{ \|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 + \lambda\Phi(\mathbf{D}) \right\} \\ &\text{s.t.} \quad \|\alpha_i\|_{0,\mathbf{b}^{(t)}} \leqslant \kappa, \forall i. \end{aligned} \tag{8}$$

This Problem (8) can be solved in two steps: first by updating dictionary, then adjusting the block coherence of the available dictionary. In the first step, by omitting the second term on the right-hand side of the equation defined in Problem (8), we can solve the reduced version of Problem (8) as follows:

$$\begin{aligned} &\left[\mathbf{D}^{(t)}, \mathbf{A}^{(t)}\right] = \arg\min_{\mathbf{D},\mathbf{A}} \|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 \\ &\text{s.t.} \quad \|\alpha_i\|_{0,\mathbf{b}^{(t)}} \leqslant \kappa, \forall i. \end{aligned} \tag{9}$$

Like BK-SVD [5], we sequentially update every block of atoms in the dictionary $\mathbf{D}^{(t)}$ and the corresponding nonzero values in $\mathbf{A}^{(t)}$ to minimize the representation error in Problem (9). For each block $j \in [1, B]$, the dictionary update step proceeds as follows. Let $\mathbf{R}_{\omega_j}$ denote the representation error of the signals $\mathbf{X}_{\omega_j}$ excluding the contribution of the j-th block, *i.e.* $\mathbf{R}_{\omega_j} = \mathbf{X}_{\omega_j} - \sum_{i \neq j} \mathbf{D}_{b_i} \mathbf{A}_{\omega_j}^{b_i}$. We deduce that $\left\| \mathbf{R}_{\omega_j} - \mathbf{D}_{b_j} \mathbf{A}_{\omega_j}^{b_j} \right\|_F$ is the representation error of the signals with the indices ω_j. In order to minimize this error, we take the matrix of maximum rank $|b_j|$ that best approximates of $\mathbf{R}_{\omega_j}$ as $\mathbf{D}_{b_j} \mathbf{A}_{\omega_j}^{b_j}$. According to the singular value decomposition (SVD) of the matrix $\mathbf{R}_{\omega_j}$, the dictionary update process is carried out as follows:

$$\begin{aligned} \mathbf{R}_{\omega_j} &= \mathbf{U} \boldsymbol{\Delta} \mathbf{V}^T, && (10) \\ \mathbf{D}_{b_j} &= \left[\mathbf{U}_1, \cdots, \mathbf{U}_{|b_j|} \right], && (11) \\ \mathbf{A}_{\omega_j}^{b_j} &= \left[\boldsymbol{\Delta}_1^1 \mathbf{V}_1, \cdots, \boldsymbol{\Delta}_{|b_j|}^{|b_j|} \mathbf{V}_{|b_j|} \right]^T, && (12) \end{aligned}$$

where the first $|b_j|$ principal components of $\mathbf{R}_{\omega_j}$ are truncated to update the block of atoms $\mathbf{D}_{b_j}$ and the group sparse coefficients $\mathbf{A}_{\omega_j}^{b_j}$.

Finally, the second step of the solution to Problem (8) is to adjust the block coherence of the updated dictionary $\mathbf{D}^{(t)}$. We compute the gradient of the objective function with respect to φ_r and equate it to zero. Thus the closed-form solution of Problem (8) on φ_r is given as follows:

$$\varphi_r = \left(\mathbf{I}_d \alpha_r \alpha_r^T + \lambda \sum_{j \in b_j, j \neq r} \varphi_j \varphi_j^T \right) \backslash \left(\mathbf{X} \alpha_r^T - \sum_{k \neq r} \varphi_k \alpha_k \alpha_r^T \right), \quad (13)$$

where $\mathbf{I}_d$ is the identity matrix of size $d \times d$, α_r is the r-th row group of the sparse representation $\mathbf{A}^{(t)}$, φ_r is in the block of atoms for b_j, and $\alpha_r \alpha_r^T$ indicates the weight of the atom φ_r used to encode $\mathbf{X}$.

3. RESULTS AND ANALYSIS

3.1. Simulation Evaluation

A random matrix of size $d \times K$ is generated as the initial dictionary $\mathbf{D} \in \mathbb{R}^{64 \times 96}$ with independent and identically normally distributed entries. Each column of the dictionary is normalized so that its Euclidean norm equals to 1. The initial block structure for $\mathbf{D}$ is chosen as $\mathbf{b} = [1, 1, 1, \cdots, 32, 32, 32]$. That is, the dictionary $\mathbf{D}$ is consists of 32 subspaces with 3 atoms in each one. $\mathbf{X}$ is a set of the $L = 2500$ test signals of dimension 64 with a 2-block sparse representation over $\mathbf{D}$. The active blocks in $\mathbf{A}$ are chosen randomly and the coefficients are uniformly distributed random entries again. According to Equation (1), the additive white Gaussian noise (AWGN) with variant noise levels is added to the data set $\mathbf{X}$ in order to generate the simulated data synthetically with the desired SNR values.

In this simulation, for the given signals $\mathbf{X}$, we evaluated the overall performance of our developed algorithm to recover the overcomplete dictionary $\mathbf{D}$ with the underlying block structure $\mathbf{b}$. The normalized representation error (NRE) was computed as a function of the signal-to-noise ratio (SNR) of the signals $\mathbf{X}$ corrupted by AWGN. For the synthetic signals $\mathbf{X}$ with the variant SNR, the results of the proposed algorithm are compared with those of K-SVD [1], and BK-SVD+SAC [5] shown in Table 1. Note that for SNR $\leqslant$ 25dB, K-SVD [1] reaches lower reconstruction error than BK-SVD+SAC [5] and our BSIK-SVD, which implies that the use of block-sparsifying dictionaries is unjustified. That is, the block structure no longer exists in the data when the SNR is low. Figure 1 gives the compared NRE results of these different algorithms evaluated as a function of the number of iterations in the noiseless setting. As can be seen from simulation results, the proposed algorithm has the smaller representation errors and the faster convergence rate than the existing state-of-the-art methods when the SNR is high enough.

Table 1: The compared NRE results of our BSIK-SVD algorithm, K-SVD [1], and BK-SVD+SAC [5] for simulated signals with the variant SNR values.

SNR	5	15	25	35	45
[1]	.4739	.2098	.0737	.0267	.0130
[5]	.5674	.2298	.0760	.0244	.0080
BSIK-SVD	.5668	.2292	.0747	.0236	.0075

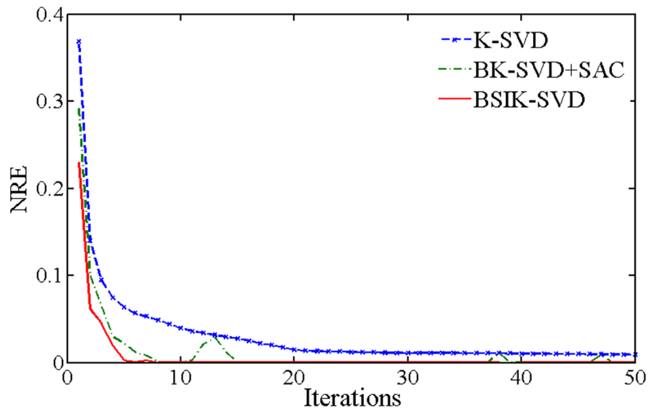


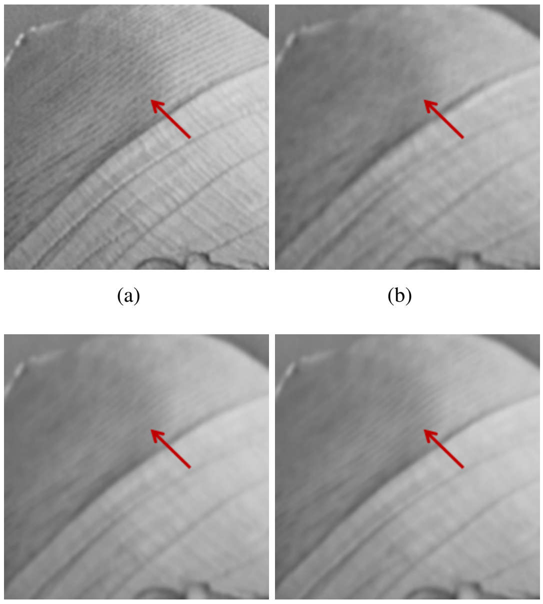
Fig. 1: The compared NRE results of our BSIK-SVD, K-SVD [1] and BK-SVD+SAC [5] evaluated as a function of the number of iterations in the noiseless setting.

3.2. Experiments on Real Images

Besides the simulations on synthetic signals, we have also implemented the qualitative and quantitative evaluation on numerous test images from standard image databases [11]. In this experiment, for a fair comparison of K-SVD [1], BK-SVD+SAC [5] and our proposed BSIK-SVD algorithm, we first initialized the dictionary $\mathbf{D}$ as a same random matrix of size 64×96 with normally distributed entries and its normalized columns. The maximal block size was $s = 3$ and the block sparsity is typically set to $\kappa = 2$. All the non-overlapping image patches of the same size 8×8 extracted from a training image were reshaped as the training signals $\mathbf{X}$. Then we adopted the proposed algorithm and other popular methods [1, 5] to optimize the dictionary where the number of iteration is 50. Next, both the proposed algorithm and BK-SVD+SAC were used to find a κ-block sparse solution, whereas K-SVD was used to find a $\kappa \times s$ sparse representation. Finally, we evaluated the results of these different algorithms by the NRE and their convergence behavior. When the training image was also used as the test image, Table 2 shows the comparison of the NRE/SSIM results obtained by the current popular methods [1, 5] and the proposed algorithm separately for a set of noiseless images. To further inspect the performance of dictionary learning, the detailed reconstructed results of these different methods for the fragments of the *Lena* image is shown in Figure 2, respectively. Table 2 presents that our proposed algorithm with less NRE results has higher SSIM results than K-SVD [1] and BK-SVD+SAC [5]. The visual comparisons in Figure 2 demonstrate that Figure 2(d) has much clearer textures recovered in the regions (pointed by the red arrow) than Figure 2(c), whereas Figure 2(b) almost loses all the fine structures. These experiments verify that the proposed algorithm has better recovery of real images than the baseline methods.

Table 2: NRE/SSIM results for a set of noiseless images.

Images	[1]	[5]	Proposed
House	.3440/.8277	.0828/.9163	.0739/.9204
Monarch	.3022/.8786	.1730/.8990	.1608/.9093
Baboon	.2509/.7399	.1996/.6860	.1921/.6872
Barbara	.1977/.8817	.1109/.9004	.1099/.9030
Lena	.1613/.9058	.0880/.9078	.0870/.9115



(a) (b) (c) (d)

Fig. 2: Visual comparisons of the reconstructed results for the *Lena* image: (a) Original subimage; (b) K-SVD [1]; (c) BK-SVD+SAC [5] and (d) the proposed algorithm.

4. CONCLUSIONS

In this paper, we addressed the problem of block-structured incoherent dictionary learning for sparse representations. In fact, the proposed algorithm, which can be seen as an extension of K-SVD [1] or BK-SVD+SAC [5], incorporates the block structure identification and the intra-block coherence regularization for block sparse representations of the signals. Our BSIK-SVD algorithm consists of two successive steps at each iteration: sparse coding and dictionary update. In sparse coding step, we employed the two-stage clustering technique to identify the intrinsic block structure existed in the dictionary for structured sparse coding. For dictionary update step, we solved the minimization problem that incorporated the desired constraints of the block coherence between the dictionary atoms after the block-sparsifying dictionary update.The experimental results of synthetic data and real images demonstrate that our BSIK-SVD algorithm has less representation error and fewer artifacts in most cases, which leads to accurate sparse representations for a given set of signals.

5. REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," IEEE Transactions on Signal Processing, vol.54, no.11, pp.4311-4322, 2006.

[2] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images," Siam Review, vol.51, no.1, pp.34-81, 2009.

[3] S. Bengio, F. Pereira, Y. Singer, and D. Strelow, "Group sparse coding," Advances in NIPS, vol.22, 2009, pp.82-89.

[4] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-Sparse Signals: Uncertainty Relations and Efficient Recovery," IEEE Transactions on Signal Processing, vol.58, no.6, pp.3042-3054, 2010.

[5] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Dictionary Optimization for Block-Sparse Representations," IEEE Transactions on Signal Processing, vol.60, no.5, pp.2386-2395, 2012.

[6] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," IEEE Transactions on Information Theory, vol.50, no.10, pp.2231-2242, 2004.

[7] T. Strohmer and R. W. Heath Jr., "Grassmannian frames with applications to coding and communication," Applied and Computational Harmonic Analysis, vol.14, no.3, pp.257-275, 2003.

[8] D. Barchiesi, and M. D. Plumbley, "Learning Incoherent Dictionaries for Sparse Approximation Using Iterative Projections and Rotations," IEEE Transactions on Signal Processing, vol.61, no.8, pp.2055-2065, 2013.

[9] Ender M. Eksioglu, "A Clustering Based Framework for Dictionary Block Structure Identification," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, May 22-27, 2011, pp.4044-4047.

[10] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol.2, no.1, pp.86-97, 2012.

[11] CVG, Test images, http://decsai.ugr.es/cvg/dbimagenes/, July 25, 2013.