*ICIC Express Letters*
*Part B: Applications*
Volume **7**, Number **9**, September **2016**

 ISSN 2185-2766
pp. 1885–1890

# ONLINE VISUAL TRACKING WITH IMPROVED SPARSE PROTOTYPES

Hongli Yan[1], Chenhua Liu[2] and Yonghua Shi[1]

[1]School of Electronic and Electrical Engineering (Ministry of Education)
Chuzhou University
No. 1528, Fengle Ave., Chuzhou 239000, P. R. China
Yhli81@163.com

[2]Key Laboratory of Advanced Process Control for Light Industry
Jiangnan University
No. 1800, Lihu Ave., Wuxi 214122, P. R. China
liuchenhua1991@163.com

Received February 2016; accepted May 2016

Abstract. *In this paper, a robust tracking method based on improved sparse prototypes is proposed by solving a Bayesian inference problem. Different from the traditional sparse prototypes based tracking method, we employ the sparse representation for the observation instead of the collaborative representation, which can effectively reject the redundant features in target subspace. Moreover, an effective numerical method based on Accelerated Proximal Gradient (APG) is applied within the process of object representation minimization. Both qualitative and quantitative experiments show that the proposed method achieves more favorable performance than several competitive methods.*
**Keywords:** Visual tracking, Bayesian inference, Improved sparse prototypes

1. **Introduction.** As one of the underlying issues in computer vision, object tracking is of great importance for its multitudinous potential applications including image compression, video surveillance, activity analysis and so on. While much work has been done in the past decades, it is still a challenging task in numerous aspects including pose variation, shape deformation, varying illumination, camera motion, and occlusions.

Recently, sparse representation has been extensively studied for object tracking [1-5]. In [1], Mei and Ling present an L1 tracker based on sparse representation. The candidate target region is constructed with sparse representation of target templates and the error term is handled with trivial temples. However, the L1 tracker needs to solve a series of L1-minimization problems with rather expensive computational cost. In [2], an efficient gradient descent approach is applied to accelerating the solving process of the L1 minimization problem. Some other sparse based methods have been proposed from different views. In [3], Jia et al. construct an alignment pooling that integrates both the local and global information of target region. Zhong et al. [4] develop a collaborative model based on two independent sparsity-based trackers and evaluate the candidates by integrating these information.

Inspired by the study of subspace learning, some collaborative representation based methods have also been proposed to effectively employ all the feature bases in target subspace [5-7]. In [5], Xiao et al. apply the L2-norm to regulating the target coefficient and trivial coefficient. Although the appearance model in [5] has a much lower computational complexity for the convex and differentiable property of L2-norm compared to the L1-tracker, the weak sparse projection coefficient will cause the redundant features, which may deteriorate the ambiguity of square templates. To reject the outliers in the process of tracking, Wang et al. develop sparse prototypes model which integrates the subspace

collaborative representation and trivial templates sparse representation [6]. However, we empirically find that the information in subspace is not all from the target. As the collaborative representation needs to use all these feature bases in subspace, the constructed samples can be interfered by these redundant features (e.g., background), which may further affect the measure to candidates.

Motivated by the above-mentioned work, in this paper, we present a robust tracking method based on the improved sparse prototypes. There are two main differences between [6] and our work. Firstly, we use the L1-norm to regulate both the target coefficient and the trivial coefficient but not only for the trivial coefficient, which ensures our appearance model can effectively reject redundant features in target subspace. Secondly, an effective numerical method based on Accelerated Proximal Gradient (APG) [8] is introduced to solve the target projection coefficient within the process of object representation minimization.

In the rest of our paper, Section 2 describes the object representation based on the improved sparse prototypes. After that, the tracking framework is proposed in Section 3. The experiments are then given in Section 4. Finally, the conclusion is introduced in Section 5.

## 2. Object Representation.

In this section, we introduce the object representation with improved sparse prototypes and an effective numerical algorithm based on APG for the solution of the minimization problem to object function.

### 2.1. Improved sparse prototypes.

Given an orthogonal PCA subspace $\mathbf{U} \in R^{d\times m}$, where $d$ and $m$ are the feature dimension and the number of PCA basis, respectively. The target region $\mathbf{y} \in R^{d\times 1}$ can be represented by the subspace with projection coefficient $\mathbf{z} \in R^{m\times 1}$ and a Laplacian error term $\mathbf{e} \in R^{d\times 1}$, and thus we have the following minimization problem in [6]

$$\min_{z,e} \frac{1}{2}\|\mathbf{y} - \mathbf{U}\mathbf{z} - \mathbf{e}\|_2^2 + \lambda\|\mathbf{e}\|_1 \tag{1}$$

where $\lambda$ is a penalty parameter. However, we note that the image subspace still includes the redundant features. To remove redundant features while preserving the useful parts in the subspace, we use L1-norm to select useful features, and we have the improved sparse prototypes:

$$\min_{z,e} \frac{1}{2}\|\mathbf{y} - \mathbf{U}\mathbf{z} - \mathbf{e}\|_2^2 + \mu\|\mathbf{z}\|_1 + \lambda\|\mathbf{e}\|_1 \tag{2}$$

where $\mu$ is a penalty parameter. The $\|\mathbf{z}\|_1$ is employed to select the useful features in target coefficient, while the $\|\mathbf{e}\|_1$ is used to reject outliers. Figure 1 shows the object representation with improved sparse prototypes.

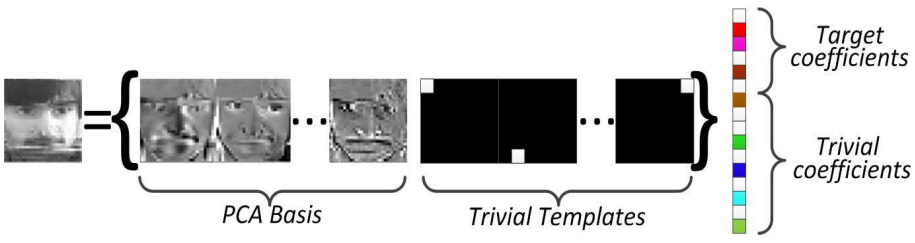


FIGURE 1. The object representation

2.2. **Effective numerical method for solving (2).** It can be seen that there is no close-form solution for Equation (2), and thus we propose a special method for the minimization of Equation (2). We extract the convex and differentiable part of Equation (2) as:

$$F(\mathbf{z}, \mathbf{e}) = \frac{1}{2}\|\mathbf{y} - \mathbf{U}\mathbf{z} - \mathbf{e}\|_2^2 \tag{3}$$

Then, we can iteratively estimate $\mathbf{z}$ and $\mathbf{e}$. When we fix projection term $\mathbf{z}$, the error term $\mathbf{e}$ can be directly estimated with Soft-threshold operation in [6]. When we fix error term $\mathbf{e}$, the target projection term $\mathbf{z}$ can be estimated via the APG method. The whole iterative method is summarized in Algorithm 1. It can be seen that there are two subproblems in Algorithm 1:

$$\mathbf{e}_{k+1} = \arg\min_{e} \lambda\|\mathbf{e}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{U}\mathbf{z}_k - \mathbf{e}_k\|_2^2 \tag{4}$$

$$\mathbf{z}_{k+1} = \arg\min_{z} \mu\|\mathbf{z}\|_1 + \frac{\xi}{2}\left\|\mathbf{z}_k - \mathbf{g}_{k+1}^z + \frac{1}{\xi}\nabla_z F\left(\mathbf{g}_{k+1}^z, \mathbf{e}_{k+1}\right)\right\|_2^2 \tag{5}$$

**Algorithm 1** Effective numerical method for solving (2)

1: set $e_0 = e_{-1} = 0$, $z_0 = z_{-1} = 0$, and $t_0 = t_{-1} = 1$
**Input:** The PCA subspace $\mathbf{U}$, the candidate sample $\mathbf{y}$, the Lipschitz constant $\xi$
2: **for** $k = 0, 1, \ldots$, until both the $\boldsymbol{z}$ and $\boldsymbol{e}$ are convergent to optimal state **do**
3: $\mathbf{e}_{k+1} = \arg\min_{e} \lambda\|\mathbf{e}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{U}\mathbf{z}_k - \mathbf{e}_k\|_2^2$
4: $\mathbf{g}_{k+1}^z = \mathbf{z}_k + \frac{t_{k-1}}{t_k}(\mathbf{z}_k - \mathbf{z}_{k-1})$
5: $\mathbf{z}_{k+1} = \arg\min_{z} \mu\|\mathbf{z}\|_1 + \frac{\xi}{2}\left\|\mathbf{z}_k - \mathbf{g}_{k+1}^z + \frac{1}{\xi}\nabla_z F\left(\mathbf{g}_{k+1}^z, \mathbf{e}_{k+1}\right)\right\|_2^2$
6: $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$
7: **end for**
**Output:** The optimal $\boldsymbol{z}^*$ and $\boldsymbol{e}^*$

We refine $\boldsymbol{\alpha}_k^e = \mathbf{y} - \mathbf{U}\mathbf{z}_k$, $\boldsymbol{\alpha}_k^z = \mathbf{g}_{k+1}^z - \frac{1}{\xi}\nabla_z\left(\mathbf{g}_{k+1}^z, \mathbf{e}_{k+1}\right)$, $\varphi = \frac{\mu}{\xi}$. As $\nabla_z F(\mathbf{z}, \mathbf{e}) = \mathbf{U}^{\mathrm{T}}(\mathbf{U}\mathbf{z} + \mathbf{e} - \mathbf{y})$, we can easily get solutions:

$$\mathbf{z}_{k+1}^* = S_\lambda\left(\boldsymbol{\alpha}_k^e\right) \tag{6}$$

$$\mathbf{e}_{k+1}^* = S_\varphi\left(\boldsymbol{\alpha}_k^z\right) \tag{7}$$

where $S_\tau(x)$ is the soft-threshold operation defined as $S_\tau(x) = sgn(x)\left(\mid x \mid -\tau\right)$.

3. **Tracking Framework.** The object tracking task can be cast as a Bayesian inference problem in the hidden Markov model. Given a series of observed samples $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t\}$, the purpose is to estimate the hidden state variable $\mathbf{x}_t$ recursively:

$$p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right) \propto p\left(\mathbf{y}_t|\mathbf{x}_t\right) \int p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right) p\left(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right) d\mathbf{x}_{t-1} \tag{8}$$

where $\mathbf{x}_t$ is the object state, and $\mathbf{y}_t$ is the observation at time $t$. $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is called the motion model that describes the state transition between two continuous object states while $p(\mathbf{y}_t|\mathbf{x}_t)$ indicates the observation model which is applied to computing the likelihood of candidates.

**Motion Model:** Let $\boldsymbol{x_t} = \{l_x, l_y, \theta, s, \alpha, \phi\}$, where $l_x$, $l_y$, $\theta$, $s$, $\alpha$, and $\phi$ indicate $x$, $y$ translations, rotation angle, scale, aspect ratio, and skew respectively. These affine parameters are supposed to be independent and modeled by six scalar Gaussian distributions. We use the random walk to formulate the state transition, i.e., $p(\mathbf{x}_t|\mathbf{x}_{t-1}) = N(\mathbf{x}_t; \mathbf{x}_{t-1}, \boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ is a diagonal covariance matrix.

**Observation Model:** It is necessary to take the occlusion into consideration for the measure to candidates, and we note the precise location can benefit from the error term

**e**. In this paper, we take the mask in [6] to distinguish non-occluding and occluding parts for different operations in the likelihood function:

$$p\left(\mathbf{y}^i|\mathbf{x}^i\right) = \exp\left(-\|\boldsymbol{\rho}^i \odot \left(\mathbf{y}^i - \mathbf{U}\mathbf{z}^i\right)\|_2^2 - \omega\|1 - \boldsymbol{\rho}^i\|_1\right) \tag{9}$$

where $\mathbf{x}^i$ is the $i$th sample of candidates, $\mathbf{y}^i$ denotes the image patch predicated by $\mathbf{x}^i$, and $\boldsymbol{\rho}^i = [\rho_1^i, \rho_2^i, \ldots, \rho_d^i]^{\mathrm{T}}$ indicates the zero elements vector of error term $\mathbf{e}^i$. If the $j$th element of $\mathbf{e}^i$ is 0, then $\rho_j^i = 1$; otherwise $\rho_j^i = 0$. $\odot$ is the Hadamard product, and $\omega$ denotes a penalty term. The former part of Equation (9) accounts for the reconstruction error of unoccluded proportion of the target image, and the latter part aims to handle the occluded pixel.

**Online Update:** The online update of target subspace is important to the changes of object in the process of tracking. As the error term **e** can identify the outliers, the samples used to update the subspace can be collected as:

$$y_j^i = \begin{cases} y_j^i & |e_j^i| = 0 \\ \mu_j & \text{otherwise} \end{cases} \tag{10}$$

where $y_j^i$ is the $j$th element of the $i$th candidate sample, and $\mu_j$ is the $j$th element of mean vector of subspace. Then, we can use the collected samples to update the subspace with the incremental principal component method in [7].

4. **Experiments.** The proposed tracker is implemented in MATLAB and runs at 4 frames per second on a 3.06 GHz i7 core PC with 4GB memory. We empirically set $\lambda = 0.024$, $\mu = 0.2$, and the Lipschitz constant $\xi = 6$. The location of the target is manually denoted in the first frame. 16 PCA bases are used for the subspace in all the sequences. Our proposed tracker is incrementally updated when 5 usable image patches are accumulated. To prove the effectiveness of the proposed algorithm, we use six challenge image sequences which contain different challenging factors (e.g., severe occlusion, motion blur) and compare our method with four competitive methods: SCM [4], IVT [7], L2-RLS [5], and OTSP [6].

TABLE 1. Average overlap rate. The best result is shown in **bold** font.

| **Sequence** | IVT | SCM | L2-RLS | OTSP | Ours |
|---|---|---|---|---|---|
| *Occlusion2* | 0.73 | 0.82 | 0.78 | 0.74 | **0.85** |
| *DavidOutdoor* | 0.52 | 0.38 | **0.75** | 0.74 | **0.75** |
| *DavidIndoor* | 0.44 | 0.51 | 0.23 | 0.45 | **0.77** |
| *Singer1* | 0.47 | **0.84** | 0.24 | 0.80 | **0.84** |
| *Face* | 0.71 | 0.56 | 0.73 | 0.63 | **0.78** |
| *Deer* | 0.24 | 0.61 | 0.60 | 0.58 | **0.68** |
| **Average** | 0.52 | 0.62 | 0.56 | 0.66 | **0.78** |

TABLE 2. Average center location error. The best result is shown in **bold** font.

| **Sequence** | IVT | SCM | L2-RLS | OTSP | Ours |
|---|---|---|---|---|---|
| *Occlusion2* | 7.8 | 4.4 | 5.5 | 9.0 | **3.7** |
| *DavidOutdoor* | 52.4 | 67.1 | 6.0 | 8.5 | **5.7** |
| *DavidIndoor* | 35.9 | 17.7 | 132.6 | 26.1 | **3.3** |
| *Singer1* | 11.9 | 3.3 | 72.8 | 3.0 | **2.7** |
| *Face* | 15.0 | 46.9 | 13.8 | 48.4 | **11.9** |
| *Deer* | 135.2 | 10.1 | 9.4 | 11.3 | **6.2** |
| **Average** | 43.0 | 24.9 | 40.0 | 17.7 | **5.6** |

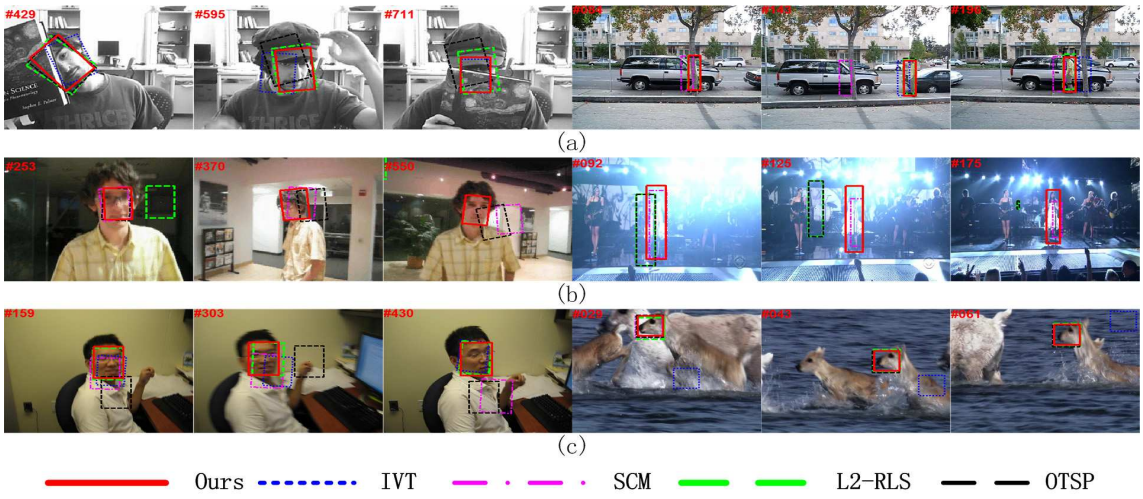


FIGURE 2. Sample tracking results on six challenging sequences. (a) Occlusion2 and DavidOutdoor with severe occlusion. (b) DavidIndoor and Singer1 with illumination variation. (c) Face and Deer with motion blur.

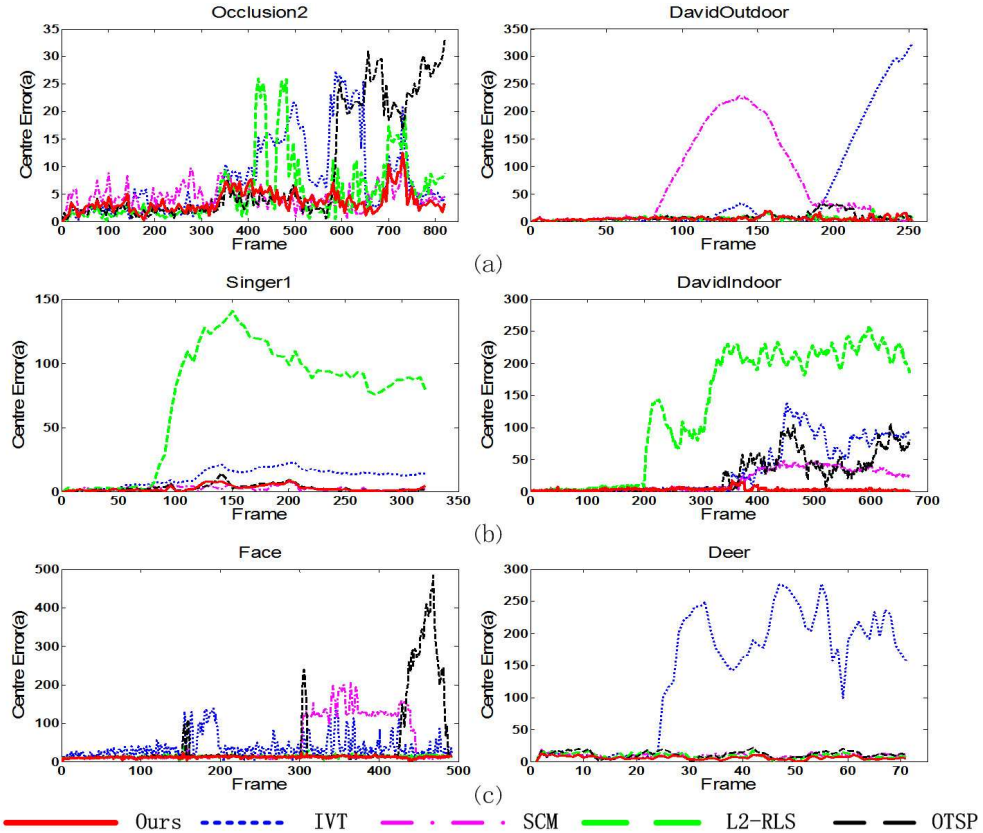


FIGURE 3. Quantitative evaluation of the trackers in terms of position errors (in pixels). (a) Occlusion2, DavidOutdoor. (b) Singer1, DavidIndoor. (c) Face, Deer.

4.1. **Quantitative evaluation.** Quantitative evaluation is applied to evaluating the effectiveness of tracking algorithms. We evaluate the aforementioned algorithms by computing their average overlap rates and center errors. A bigger overlap rate and a smaller center error mean a more proper result. Given the result of each frame, and corresponding ground truth, we can get the overlap rate and the center error by the PASSCAL VOC [9] criterion and the Euclidian distance, respectively. The results of overlap rate are listed in Table 1, and the results of center error are listed in Table 2 and Figure 3.

4.2. **Qualitative evaluation.**

**Severe Occlusion:** We test two sequences (Occlusion2, DavidOutdoor) characterizing in having either long-time severe occlusion or partial occlusion. The IVT tracker does not take the occlusion into consideration, and it is less effective in both two sequences. Although the OTSP considers the occlusion in object representation, the redundant features of subspace can worsen the results. Overall, our tracker can perform well in both the two sequences.

**Illumination Change:** Figure 2(b) presents the tracking results in the sequences with drastic illumination change. Moreover, the target scale and rotation also change rapidly. Although the IVT, L2-RLS, and OTSP tracker adopt the incremental PCA, it is difficult to handle the changes of object for the interference of redundant features. Compared with these trackers, our improved appearance model can reject the redundant features, and obtain more effective results.

**Motion Blur:** Figure 2(c) shows results from two challenging sequences with abrupt motion. The motion blur is a challenging problem which can increase the difficulty to predict the location of target. We note that our tracker can track well in sequences of Face and Deer, which can be attributed to the good balance between sparse representation and collaborative representation. Besides, we also employ the Laplacian error term to reject the outliers. Overall, our tracker can perform well in terms of motion blur.

5. **Conclusion.** In summary, based on the framework of sparse prototypes in [6], we adopt L1-norm to regulate the projection coefficient for the rejection of redundant features. Furthermore, we also introduce an effective numerical method to solve the minimization of improved object representation. Thus, our tracker can effectively reject the redundant features while keeping enough useful feature information. Extensive experiments show that our method performs better than several competitive methods. In the future, we plan to introduce discriminate information in the object representation for more effective tracking results.

**Acknowledgment.** This work is partially supported by Anhui Provincial Natural Science Foundation (1408085QF134), Natural Science Foundation of Higher Education Institutions (KJ2015A252), Anhui Provincial Leading Talents Introduction and Cultivation Project in Universities and Colleges (gxfxZD2016252).

## REFERENCES

[1] X. Mei and H. Ling, Robust visual tracking using L1 minimization, *IEEE International Conference on Computer Vision*, pp.1436-1448, 2009.

[2] C. Bao, Y. Wu, H. Ling and H. Ji, Real time robust L1 tracker using accelerated proximal gradient approach, *IEEE International Conference on Computer Vision and Pattern Recognition*, pp.1830-1837, 2012.

[3] X. Jia, H. Lu and M. Yang, Visual tracking via adaptive structural local sparse appearance model, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1822-1829, 2012.

[4] W. Zhong, H. Lu and M. Yang, Robust object tracking via sparsity based collaborative model, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1838-1845, 2012.

[5] Z. Xiao, H. Lu and D. Wang, L2-RLS based object tracking, *IEEE Trans. Circuits and Systems for Video Technology*, vol.24, no.8, pp.1301-1308, 2014.

[6] D. Wang, H. Lu and M. Yang, Online object tracking with sparse prototypes, *IEEE Trans. Image Processing*, vol.22, no.1, pp.314-325, 2013.

[7] D. Ross, J. Lim, R. Lin and M. Yang, Incremental learning for robust visual tracking, *International Journal of Computer Vision*, vol.77, no.1, pp.125-141, 2008.

[8] D. P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, *Siam Journal on Optimization*, 2008.

[9] M. Everingham, L. Van Gool, C. K. Williams, J. Winn and A. Zisserman, The pascal visual object classes (VOC) challenge, *International Journal of Computer Vision*, vol.88, no.2, pp.303-338, 2010.