ICIC Express Letters
Part B: Applications
Volume 7, Number 12, December 2016

ICIC International ©2016 ISSN 2185-2766
pp. 2619–2624

ITERATIVELY REWEIGHTED SPARSE CODING FOR VISUAL TRACKING

Hongli Yan[1], Qibin Lin[1] and Yufeng Wang[2]

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

[2]College of Telecommunications and Information Engineering
Nanjing University of Post and Telecommunications
No. 66, Xin Mofan Road, Nanjing 210000, P. R. China
wfwang@njupt.edu.cn

Received June 2016; accepted September 2016

Abstract. *Recently the sparse representation has been successfully applied to visual tracking. However, the traditional sparse coding models always take the coding residual as Gaussian or Laplacian distribution, which may not describe the outliers accurately. In this paper, we model the sparse coding as a sparsity-constrained problem with a weighted matrix in the framework of Bayesian inference, which can be more robust to reject outliers. Moreover, we also propose an iterative numerical method to minimize the object representation function effectively. Extensive experimental results on several challenging sequences demonstrate that our method can be more effective compared with state-of-the-art methods in dealing with appearance occlusion, illumination, motion blur, etc.*

Keywords: Visual tracking, Bayesian inference, Weighted sparse coding

1. **Introduction.** Visual tracking plays a highly researched role in computer vision since it has been widely applied to video surveillance, human-computer and motion analysis. Although much progress has been obtained in recent years, designing a robust tracking system is still a difficult task due to large appearance changes caused by numerous interferential factors, including occlusion, illumination, abrupt motion, background clutter, etc.

Recently, sparse representation has been successfully used in object tracking [1-7]. In [1], Mei and Ling develop an L1 tracker that effectively uses both the target templates and trivial templates to reconstruct all the candidates with the sparse representation technique. However, heavy computational burden seriously hampers the speed of L1 tracker. In [2], an effective numerical operation based on APG has been applied to improving the speed and accuracy for the L1 tracker. To improve the effectiveness of L1 tracker, several methods have also been proposed from different views. In [3], Jia et al. propose a sparse appearance model based on structural local representation with a feature pooling method. In [4], Zhong et al. develop a collaborative sparse model that integrates the information from two independent sparse models.

Motivated by the study of subspace learning, the incremental subspace methods also have been widely introduced to visual tracking [8-10]. In [8], Ross et al. adopt online update method to update and learn the PCA subspace representation effectively. To handle the appearance change in the process of tracking, Xiao et al. [9] introduce square trivial templates in the subspace representation, and employ the L2-norm to regulate the error coefficient. However, the L2-norm is weak sparse, which may deteriorate the ambiguity of square trivial templates. To solve this problem, Wang et al. [10] further apply the L1-norm to regulating the error coefficient. Nevertheless, all these mentioned works take the residual as Gaussian distribution, which cannot tolerate the outliers effectively. Although the Laplacian distribution can be more robust compared with Gaussian distribution, it still does not hold well, especially when serious occlusion occurs in the process of tracking. To solve this problem, a weighted term is successfully introduced to face recognition [11,12]. The elements in this weighted term essentially reflect the weight to the outliers, and we can accurately measure the outliers after several iterations by minimizing the whole object function, which can be applied to rejecting the outliers more effectively.

Inspired by the above-mentioned work, in this paper, the weighted term is employed in our representation, and the resulted weights have clear meaning, i.e., outliers will have low weight values. Moreover, we take the minimization problem as an iteratively reweighted sparse coding problem. By iteratively computing the weights, the minimization solution of our proposed representation can be solved efficiently.

The rest parts of our paper are arranged as follows. Section 2 presents the proposed object representation model. Section 3 introduces the tracking framework. Section 4 conducts the experiments, and the conclusion is arranged in Section 5.

2. Object Representation.

This section introduces the proposed object representation with weighted sparse coding. Furthermore, an effective numerical method within the framework of APG [2] is also proposed for the minimization of object function.

2.1. Object representation with weighted sparse coding.

In [10], a d-dimensional candidate image $\boldsymbol{y} \in R^{d\times 1}$ is constructed as:

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{c} + \boldsymbol{n} + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{D} \in R^{d\times m}$ is a target subspace composed of m PCA basis vectors, $\boldsymbol{c} \in R^{d\times 1}$ means the target projection coefficient, $\boldsymbol{n} \in R^{d\times 1}$ denotes an additive error term, and $\boldsymbol{e} \in R^{d\times 1}$ is the residual vector. Then, we can have the following minimization problem in [8]:

$$\min_{c,n} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{n}\|_2^2 + \lambda \|\boldsymbol{n}\|_1 \tag{2}$$

where λ is a penalty parameter. In this case, the residual $\boldsymbol{e}$ is measured as Gaussian distribution, which is sensitive to outliers. Although the Laplacian distribution can be more accurate compared with Gaussian distribution, it is still not accurate enough to reject outliers. Thus we introduce a weight matrix $\boldsymbol{W}$ to restrict the residual term. Then we have:

$$\min_{c,n} \frac{1}{2} \|\boldsymbol{W}^{1/2}(\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{n})\|_2^2 + \lambda \|\boldsymbol{n}\|_1 \tag{3}$$

It should be noted that $\boldsymbol{W}$ is a diagonal matrix and the value of $W_{i,i}$ is the weight assigned to the corresponding pixel of candidate image $\boldsymbol{y}$. Thus the outlier pixels can have low weight value. We choose the logistic function as the weight function:

$$\boldsymbol{W} = \exp\left(\gamma\delta - \gamma\boldsymbol{e}^2\right) / \left(1 + \left(\exp\left(\gamma\delta - \gamma\boldsymbol{e}^2\right)\right)\right) \tag{4}$$

2.2. Effective numerical method for solving (3).

To the best of our knowledge, it is no close-form solution for the minimization of Equation (3). Thus an iterative numerical method is proposed, and we extract the convex and differentiable part of Equation (3) as:

$$F(\boldsymbol{c}, \boldsymbol{n}) = \frac{1}{2} \left\|\boldsymbol{W}^{1/2}(\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{n})\right\|_2^2 \tag{5}$$

Then, we can iteratively estimate $\boldsymbol{c}$ and $\boldsymbol{n}$ within the framework of APG. The whole iterative method is summarized in Algorithm 1.

Algorithm 1 Effective numerical method for solving (3)

```
set $\boldsymbol{n}_0 = \boldsymbol{n}_{-1} = 0$, $t_0 = t_{-1} = 1$, and $\boldsymbol{c}_0 = \left[\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right]^{\mathrm{T}}$
Input: The PCA subspace $\boldsymbol{D}$, the candidate sample $\boldsymbol{y}$, the Lipschitz constant $\xi$
for $k = 0, 1, \ldots$, until $\boldsymbol{c}$, $\boldsymbol{n}$, and $\boldsymbol{W}$ are convergent to optimal state do
    $\boldsymbol{e}_k = \boldsymbol{y} - \boldsymbol{D}\boldsymbol{c}_k - \boldsymbol{n}_k$
    $\boldsymbol{W}_k = \exp\left(\gamma\delta - \gamma\boldsymbol{e}_k^2\right) / \left(1 + \left(\exp\left(\gamma\delta - \gamma\boldsymbol{e}_k^2\right)\right)\right)$
    $\boldsymbol{g}_{k+1}^n = \boldsymbol{n}_k + \frac{t_{k-1}}{t_k}(\boldsymbol{n}_k - \boldsymbol{n}_{k-1})$
    $\boldsymbol{n}_{k+1} = \arg\min_n \lambda\|\boldsymbol{n}\|_1 + \frac{\xi}{2}\left\|\boldsymbol{n}_k - \boldsymbol{g}_{k+1}^n + \frac{1}{\xi}\nabla_n F(\boldsymbol{c}_k, \boldsymbol{g}_{k+1}^n)\right\|_2^2$
    $\boldsymbol{c}_{k+1} = (\boldsymbol{D}^{\mathrm{T}}\boldsymbol{W}_k\boldsymbol{D})^{-1}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{W}_k(\boldsymbol{y} - \boldsymbol{n}_{k+1})$
    $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$
end for
Output: The optimal $\boldsymbol{c}^*$, $\boldsymbol{n}^*$ and $\boldsymbol{W}^*$
```

In Algorithm 1, we need to solve a sub-problem:

$$\boldsymbol{n}_{k+1} = \arg\min_n \lambda\|\boldsymbol{n}\|_1 + \frac{\xi}{2}\left\|\boldsymbol{n}_k - \boldsymbol{g}_{k+1}^n + \frac{1}{\xi}\nabla_n F\left(\boldsymbol{c}_k, \boldsymbol{g}_{k+1}^n\right)\right\|_2^2 \tag{6}$$

It is easy to obtain the solution by soft-threshold operation in [10], which is defined as: $S_\tau(x) = sgn(x)\max(|x| - \tau)$:

$$\boldsymbol{n}_{k+1}^* = S_{\lambda/\xi}\left(\boldsymbol{g}_{k+1}^n - \frac{1}{\xi}\nabla_n F\left(\boldsymbol{c}_k, \boldsymbol{g}_{k+1}^n\right)\right) \tag{7}$$

3. **Tracking Framework.** The object tracking task can be taken as a Bayesian sequential importance sampling problem in the Markov model. Let $\boldsymbol{y}_t$ denote the observation in the t-frame, and $\boldsymbol{x}_t$ denotes the corresponding state variable of target. The purpose is to recursively estimate the posterior probability by the following two rules:

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

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) \propto \frac{p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})}{p(\mathbf{y}_t|\boldsymbol{y}_{1:t-1})} \tag{9}$$

where $\boldsymbol{x}_{1:t-1} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t\}$ are state vectors up to time t, and $\boldsymbol{y}_{1:t-1} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t\}$ denote corresponding observations. $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ means the motion model that describes the state transition between two continuous frames and $p(\boldsymbol{y}_t|\boldsymbol{x}_t)$ indicates the observation model that computes the likelihood of $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$.

Motion Model: Let $\boldsymbol{x}_{\boldsymbol{t}} = \{l_x, l_y, \theta, s, \alpha, \phi\}$, where l_x, l_y, θ, s, α, ϕ respectively indicates horizontal and vertical translations, rotation angle, scale, aspect ratio, and skew. A diagonalized Gaussian distribution is applied to formulating the state transition by random walk, i.e., $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = N(\boldsymbol{x}_t; \boldsymbol{x}_{t-1}, \boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ is a diagonal covariance matrix of these affine parameters.

Observation Model: After obtaining the target coefficient $\boldsymbol{c}^*$ and the additional error vector $\boldsymbol{n}^*$, we can measure the candidate by the following likelihood function in [10]:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \exp\left(-\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c}^* - \boldsymbol{n}^*\|_2^2 - \varphi\|\boldsymbol{n}^*\|_1\right) \tag{10}$$

Considering the fact that the weighted term $\boldsymbol{W}$ is calculated together with $\boldsymbol{c}$ and $\boldsymbol{n}$ in Algorithm 1, Equation (9) may not be accurate enough to measure the candidate when occlusion occurs. Thus we rewrite the likelihood function as:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \exp\left(-\|\boldsymbol{W}^{*1/2}(\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c}^* - \boldsymbol{n}^*)\|_2^2 - \varphi\|\boldsymbol{n}^*\|_1\right) \tag{11}$$

Online update: In this paper, the incremental PCA update method in [6] is adopted to update the subspace every 5 frames. The sample used to update the subspace can be collected as follows:

$$y_i = \begin{cases} y_i & |e_i| = 0 \\ \mu_i & \text{otherwise} \end{cases} \tag{12}$$

where y_i is the ith element of the target state in current frame, and μ_i is the ith element of mean vector in subspace. Then, the incremental principal component method can be applied to updating the subspace via these collected samples.

4. **Experiments.**

The proposed method is implemented in MATLAB and runs at 3 frames per second on a 3.50 GHz i7 core PC with 8GB memory. We empirically set $\lambda = 0.05$, $\varphi = 0.1$, $\gamma = 1.5$, $\delta = 2$, and the Lipschitz constant $\xi = 10$. There are 16 PCA basis vectors in the subspace for all sequences. 600 particles are sampled in each frame and the subspace is incrementally updated every 5 frames.

In this paper, ten challenging video sequences are employed to test our proposed method. These sequences contain some different challenging factors, including motion blur, severe occlusion, background clutter, etc. Moreover, five state-of-the-art methods are also used for comparison, including the Incremental Visual Tracking (IVT) [6], Sparsity based Collaborative Model (SCM) [4], Discriminative Sparse Similarity Tracking (DSST) [5], L2-regularized Least Square (L2-RLS) [9], and Least Soft-threshold Squares Tracking (LSST) [10].

4.1. **Quantitative evaluation.**

The center location error and overlap rate [13] are both used for quantitative comparison. It should be noted that a bigger overlap rate and a smaller center location error mean a more effective result. All the numerical results are shown in Table 1 and Table 2. Overall, our proposed method performs better than the other mentioned methods in terms of both center location error and overlap rate.

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

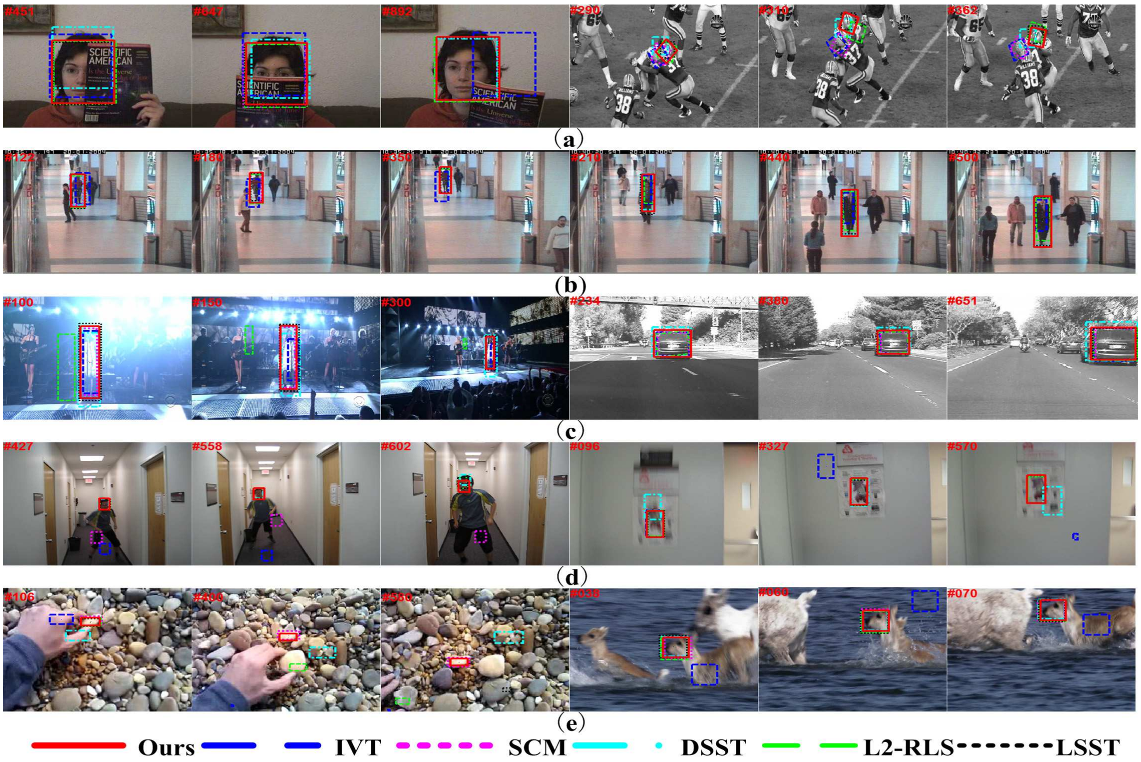
Sequence	IVT	SCM	DSST	L2-RLS	LSST	Ours
Occlusion1	0.80	**0.94**	0.82	0.91	0.89	0.90
Football	0.58	0.61	0.70	0.68	48	**0.78**
Caviar1	0.58	0.90	0.87	0.82	0.89	**0.91**
Caviar2	0.60	**0.83**	0.73	0.71	0.80	0.82
Singer1	0.47	0.84	0.70	0.24	0.81	**0.88**
Car4	**0.92**	0.90	0.77	0.91	0.85	0.91
Boy	0.19	0.53	0.78	**0.79**	0.31	**0.79**
Owl	0.21	0.77	0.78	0.78	**0.81**	**0.81**
Stone	0.12	0.62	0.10	0.37	0.17	**0.66**
Deer	0.24	0.61	0.63	0.60	0.59	**0.68**
Average	0.47	0.76	0.69	0.68	0.66	**0.81**

4.2. **Qualitative evaluation.**

Severe Occlusion: We test four sequences (Occlusion2, Football, Caviar1, and Caviar2) with partial occlusion or long-time severe occlusion. It can be seen that the SCM tracker can be much effective in Occlusion1 and Cavier2, which can be attributed to the combination of generative model and discriminative model. However, the whole performance of SCM tracker could be deteriorated when one or both of these two models catch the target inaccurately, which can be seen in sequence Football. As a whole, our method can perform well in terms of occlusion, which can be attributed to two factors: (1) The additional error term can be used to model error pixels; (2) We apply the weight matrix to restricting the coding residual, but not take the residual as Gaussian

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

Sequence	IVT	SCM	DSST	L2-RLS	LSST	Ours
Occlusion1	12.5	**3.2**	11.4	4.7	5.3	5.0
Football	17.3	15.2	8.4	9.2	43.7	**4.4**
Caviar1	11.0	**1.0**	1.9	4.2	1.4	1.2
Caviar2	3.1	2.1	4.8	**1.7**	2.3	2.2
Singer1	11.9	3.3	12.8	72.8	3.6	**1.8**
Car4	**2.9**	**2.9**	5.8	3.5	4.3	3.8
Boy	177.2	51.8	3.2	2.9	128.4	**2.8**
Owl	115.4	8.5	**6.0**	7.2	6.2	6.9
Stone	115.1	2.6	56.6	25.7	89.6	**1.9**
Deer	135.2	10.1	8.8	9.4	10.0	**6.6**
Average	60.2	10.1	12.0	14.1	29.5	**3.7**


(a)

(b)

(c)

(d)

(e)

FIGURE 1. Sample tracking results on ten challenging sequences: (a) Occlusion2 and Football with severe occlusion; (b) Caviar1 and Caviar2 with partial occlusion; (c) Singer1 and Car4 with illumination change; (d) Boy and Owl with motion blur; (e) Stone and Deer with background clutter

or Laplacian distribution, which can further improve the robustness to occlusion. Overall, our method can be more effective than other methods mentioned in this paper.

Illumination Change: Figure 1(c) presents the tracking results on the sequences Singer1, and Car4 with drastic illumination change. It can be seen that the IVT, L2-RLS and our method can perform well in Car4 when the target is just in terms of illumination, which can be attributed to the robustness of PCA to illumination. However, when the target undergoes additional scale and rotation changes but not just illumination change, the IVT and L2-RLS both get the target lost, which can be seen in Singer1.

Motion Blur: Figure 1(d) shows the tracking results on the sequences Boy and Owl with motion blur. It is a tough task to locate the position accurately when the target undergoes fast motion. The IVT tracker is sensitive to motion blur since it does not take outliers into the target representation. Compared with other methods, our method tracks well in both of the two sequences.

Background Clutter: Figure 1(e) demonstrates the results on the sequences Stone and Deer with background clutter. We can see that our proposed method can successfully track the target via effective subspace representation. Meanwhile, the additive residual constraint enables our method to measure the candidates more accurately.

5. **Conclusion.** In this paper, we present an effective tracking method with iteratively reweighted sparse coding. Different from traditional sparsity-based trackers that model the residual as Gaussian or Laplacian distribution, we introduce a weighted matrix to restrict the residual pixels in object function, which can be more accurate to describe the coding errors in practice. Furthermore, we also propose an iteratively numerical method within APG for the minimization of object function. We evaluated the proposed method on ten challenging sequences and five state-of-the-art methods, and the experimental results demonstrate the proposed method is effective compared with these mentioned methods. In our future work, we will pay more attention to hashing technologies for more efficient tracking system.

Acknowledgment. This work is partially supported by Anhui Provincial Natural Science Foundation (1408085QF134), Natural Science Foundation Higher Education Institutions (KJ2015A252), and 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 sparse collaborative appearance model, *IEEE Trans. Image Processing*, vol.23, no.5, pp.2356-2368, 2014.

[5] B. Zhuang, H. Lu, Z. Xiao and D. Wang, Visual tracking via discriminative sparse similarity map, *IEEE Trans. Image Processing*, vol.23, no.4, pp.1872-1881, 2014.

[6] D. Wang, H. Lu and C. Bo, Visual tracking via weighted local cosine similarity, *IEEE Trans. Systems, Man, and Cybernetics Part B*, vol.45, no.9, pp.1838-1850, 2015.

[7] L. Wang, H. Lu and D. Wang, Visual tracking via structure constrained grouping, *IEEE Signal Processing Letters*, vol.22, no.7, pp.794-798, 2015.

[8] 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.

[9] 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.

[10] D. Wang, H. Lu and M. Yang, Least soft-threshold squares tracking, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.2371-2378, 2013.

[11] M. Yang, L. Zhang and J. Yang, Regularized robust coding for face recognition, *IEEE Trans. Image Processing*, vol.22, no.5, pp.1753-1766, 2013.

[12] X. Li, D. Dai, X. Zhang and C. Ren, Face recognition with continuous occlusion using partially iteratively reweighted sparse coding, *IEEE First Asian Conference on Pattern Recognition*, pp.293-297, 2011.

[13] 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.