**Introduction & Motivation:** Visual tracking is a highly researched topic in the computer vision community since it has been widely applied in visual surveillance, driver assistant system, and many others. Although much progress has been made in the past decades, designing a practical visual tracking system is still a challenging problem due to numerous challenges in real world.

Very recent efforts have been made to improve this method in terms of both speed and accuracy by using APG algorithm [1] or modeling the similarity between different candidates [6]. The works in [4, 5] point out that the aforementioned methods do not exploit rich and redundant image properties which can be captured compactly with subspace representations. Thus, they propose combining the strength of subspace learning [3] and sparse representation for modeling object appearance. In their work the raw pixel templates used in in [1, 2] are replaced with the orthogonal basis vectors of the subspace, which will interfere with the accuracy of object representation.

We in this paper address this problem by proposing a tracking method based on an \(L_0\) regularized object representation. The \(L_0\) regularized object representation is able to reduce the redundant features while keeping the most important part. Furthermore, the estimation of the \(L_0\) regularized parameters can be efficiently conducted by the proposed APG algorithm.

\**L_0\** Regularized Object Representation:** We assume that the target region \(y \in \mathbb{R}^{d \times 1}\) can be represented by an image subspace with corruption,

\[
y = D\alpha + e,
\]

where the columns of \(D \in \mathbb{R}^{d \times n}\) are orthogonal basis vectors of the subspace, \(\alpha\) is the sparse coefficient vector, and \(e\) represents additive errors modeled by a Laplacian noise.

We propose an \(L_0\) regularized prior to select useful features, which is defined as

\[
\min_{\alpha, e} \frac{1}{2} \|y - Da - e\|^2_2 + \lambda \|e\|_1 + \gamma \|\alpha\|_0,
\]

where \(D^\top D = I\), \(\|\cdot\|_0\) denotes the \(L_0\) norm which counts the number of non-zero elements, \(\|\cdot\|_2\) and \(\|\cdot\|_1\) denote \(L_2\) and \(L_1\) norms, respectively, \(\gamma\) and \(\lambda\) are regularization parameters, and \(I\) is an identity matrix. The term \(\|e\|_1\) is used to reject outliers (e.g., occlusions), while \(\|\alpha\|_0\) is used to select the useful features. We note that if we set \(\gamma = 0\), (2) is reduced to (4).

\**Analysis on the Effectiveness of \(L_0\) Representation:** The benefit of the \(L_0\) norm regularized prior is that it is able to reduce the redundant features while keeping the most important part, thereby facilitating the tracking result.

When there are no errors (e.g., occlusion) in the observation \(y\), i.e., \(e \approx 0\), we can think of \(L_p\) regularized error metric in general,

\[
\min_{\alpha} \frac{1}{2} \|y - Da\|^2_2 + \gamma \|\alpha\|^p_p, \quad \text{where} \quad D^\top D = I,
\]

and the solutions for different \(p\) are given in the following theorem.

**Theorem 1** Assume that \(D \in \mathbb{R}^{d \times d}\) and \(D^\top D = I\). The solution of (3) when \(p = 0\) is given by

\[
\alpha = H_{\gamma}(D^\top y),
\]

where \(p = 1\), the solution is

\[
\alpha = S_\gamma(D^\top y),
\]

and when \(p = 2\), the solution becomes

\[
\alpha = \frac{D^\top y}{1 + 2\gamma}.
\]

Here \(S_\gamma(x) = \text{sign}(x) \max(|x| - \gamma, 0)\), and \(H_\gamma(x)\) is a hard thresholding operator, which is defined as \(H_\gamma(x) = x\) if \(x > \gamma\) and 0 otherwise.

Based on Theorem 1, we have the following corollary.

**Corollary 1** We assume \(D\) is redundant and contains all possible basis. Let \(u^*\) denote the non-zero elements of \(D^\top y\). If we set \(\gamma = \frac{1}{2}\min\{|u_i^*|^2\}\), the solution of \(L_0\) regularized error metric (i.e., (3) when \(p = 0\)) can exactly recover the data \(y\).

Figure 1 shows the tracking results by using LS method [4] (i.e., \(\gamma = 0\) in (2)), \(L_0\) and \(L_2\) norm under the same dictionary \(D\), respectively. We note that using \(L_0\) regularized method is able to find the good candidate when there exists occlusion, then facilitating the tracking results.

![Figure 1: Coefficients and reconstruction results by using LS method, \(L_0\) and \(L_2\) norm under the same dictionary \(D\). The coefficients by using \(L_0\) norm are more sparse than those by \(L_2\) norm and LS method, and the reconstruction result and the best candidate are also better. The rectangles in the last image represent MAP states for particle filers.](image-url)