VISUAL TRACKING VIA MULTI-TASK NON-NEGATIVE MATRIX FACTORIZATION

Yong Wang*, Xinbin Luo†, Shiqiang Hu†
*Hisilicon Technologies
†School of Aeronautics and Astronautics, Shanghai Jiao Tong University

ABSTRACT
We propose an online tracking algorithm in which the object tracking is achieved by using subspace learning and non-negative matrix factorization (NMF) under the particle filtering framework. The object appearance is modeled by a non-negative combination of non-negative components learned from examples observed in previous frames. In order to robust tracking an object, group sparsity constraints are included to the non-negativity one. In addition, the Alternating Direction Method of Multipliers (ADMM) algorithm is proposed for efficient model updating. Qualitative and quantitative experiments on a variety of challenging sequences show favorable performance of the proposed algorithm against 9 state-of-the-art methods.

Index Terms— non-negative matrix factorization, Alternating direction method of multipliers, subspace learning

1. INTRODUCTION
Object tracking plays a crucial role in numerous vision applications including human computer interaction, human activity analysis, traffic flow video processing, to name a few [1-10]. Tracking algorithms can be categorized as either generative [2-8] or discriminative [9-11] approaches. Generative tracking algorithms usually construct appearance models with image observations in offline or online settings. The tracking problem is formulated as searching for the region with the highest probability of being generated from the appearance model.

In this paper, a novel target representation is proposed based on the non-negative matrix factorization (NMF) [12]. Linear combinations of a set of non-negative basis are employed to model object appearance. The non-negative basis will efficiently capture the structure information of the target. In order to encode the characteristics in the tracking process, group sparsity is introduced. As shown in our experiments, our method is robust to illumination variation, pose change, background clutter and severe occlusion.

This work was jointly supported by the National Natural Science Foundation of China (No. 61374161) and China Aviation Science Foundation (No. 20142057006).

To apply our constrained NMF for visual tracking, we propose a tracking framework to capture the appearance property of the target. The workflow is shown in Fig.1. Our tracking algorithm is combined with the particle filtering and subspace learning framework. Specifically, given a new sample $y$ in a new frame, the algorithm iteratively and alternatively updates basis matrix $W$ and the approximation of $y$ with respect to $W$. The likelihood of each particle is derived from its reconstruction error using the learned basis $W$, and the maximum of the likelihood is chosen to the target.

Compared with existing approaches, the contributions of this work are three fold. First, we represent the tracked object using PCA bases, taking advantages of the strengths of subspace representation. Second, we propose to use group sparsity NMF for visual tracking. To the best of our knowledge, this is the first time group sparsity NMF has been used for object tracking. The model that takes the inliers and noise into consideration is robust under different tracking scenarios. Third, we develop a novel algorithm based on Alternating direction method of multipliers (ADMM) method. In the experiments, the proposed tracking algorithm demonstrated superior performances in comparison with 9 stat-of-the-art methods.

2. NMF GROUP SPARSITY-BASED TRACKING
In this section, we first briefly introduce particle filtering framework that our tracker is formulated within. And then
the subspace learning procedure is presented. Next we give a detailed description of both the principle and algorithm steps of our tracking model, followed by the implementation of the model using the ADMM algorithm.

2.1. Particle Filtering Tracking

In the particle filtering framework, there exist two fundamental steps: prediction and update. Let \( s_{s_1t} \) denote the state variable of the tracked object and \( y_t \) denote its corresponding observation in the \( t \)-th frame. Then the posterior probability can be recursively estimated by the following two rules:

\[
p(s_{s_1t}|y_{1:t-1}) = \int p(s_{s_1t}|s_{s_1t-1})p(s_{s_1t-1}|y_{1:t-1})ds_{s_1t-1} \tag{1}
\]

\[
p(s_{s_1t}|y_{1:t}) = \frac{p(y_t|s_{s_1t})p(s_{s_1t}|y_{1:t-1})}{p(y_t|y_{1:t-1})} \tag{2}
\]

where \( s_{s_1t} \) denotes a matrix of PCA basis vectors (\( d \) represents feature dimension and \( \eta \) denotes the number of PCA basis), \( s_{s_1t-1} \) stand for all available state vectors up to time \( t \) and \( y_{1:t} \) denote their corresponding observations. \( p(s_{s_1t}|s_{s_1t-1}) \) is a dynamic model that describes the state transition, and \( p(y_t|s_{s_1t}) \) is an observation model that estimates the likelihood of observing \( y_t \) at state \( s_{s_1t} \). The posterior \( p(s_{s_1t}|y_{1:t}) \) is approximated by \( K \) weighted particles.

2.2. Subspace Learning

At each instance \( t \) we model the tracking target and all candidates with \( k \) PCA basis vectors (\( W_t \)) and an error term (\( E_t \)) as:

\[
X_t = W_t H_t + E_t \tag{3}
\]

where \( X_t \in \mathbb{R}^{n \times N}, X_t = [x_1, x_2, \cdots, x_N] \) is the observation vector, \( N \) is the number of observations, \( W_t \in \mathbb{R}^{d \times k} \) denotes a matrix of PCA basis vectors (\( d \) represents feature dimension and \( k \) the number of PCA basis), \( H_t \in \mathbb{R}^{k \times N} \) denotes the corresponding coding vectors (target coefficients), and \( E_t \in \mathbb{R}^{n \times N} \) represents the error term. The most informative \( k \) orthogonal bases of the target subspace are composed to the PCA basis vectors to model the tracking target. We use the affine transformation to model the object motion between two consecutive frames.

2.3. NMF Representation

Recently the \( L_{1,\infty} \) norm has been proposed for joint regularization. Essentially, this type of regularization aims at learning a set of joint sparse models. The \( L_{1,\infty} \) norm is a matrix norm that penalizes the sum of maximum absolute values of each row. This regularizer encourages row sparsity: i.e., it encourages entire rows of the matrix to have zero elements. Therefore, we applied group sparsity penalty \( L_{1,\infty} \) to the row groups of the \( W_t \) in equation (3) to capture the shared features among all tasks over all particles. The \( L_1 \) loss penalty function is employed to penalize the differences between template and the noise. Thus, equation (3) can be re-written as the following form:

\[
\min_{W_t, H_t, E_t} ||X_t - W_t H_t - E_t||_F^2 + \lambda_1 ||W_t||_{1,\infty} + \lambda_2 ||H_t||_{1,1} + \lambda_3 ||E_t||_{1,1}, s.t. W_t \geq 0, E_t \geq 0. \tag{4}
\]

where \( \lambda_1, \lambda_2, \lambda_3 \) are tradeoff parameters controlling reliable construction of the observation, joint sparsity regularization and noise.

The novel of our tracking method is based on the NMF to implicitly combine holistic and part based methods. The target appearances are modeled as non-negative linear combinations of a set of non-negative basis that implicitly captures structure information. The row group sparsity which reflects the underlying assumption that target appearances across frames lies in the same subspace. Therefore, our representation model inherits the merits of nonnegativity constraint from NMF. In the following subsection we show that the solution to this problem can be obtained by performing a sequence of closed form optimization steps by the ADMM method. The detail of the ADMM algorithm can be found in [13] and [14].

2.4. Resolve Equation (4)

ADMM algorithm has been shown to be robust in machine learning applications and advantageous in resolving optimization problem of the sums of simple convex functions [7]. Resolving equation (4) in Algorithm 1 is essential to efficiently compute matrix \( B_t \) and \( W_t \) alternatively. We summarize our NMF group sparsity algorithm implemented by ADMM for resolving equation (4) in Fig.3.

3. EXPERIMENT

Performance of the proposed tracker has analyzed on 25 challenging video sequences and compared with seven state-of-the-art tracking works including the Incremental Visual Tracking.
Algorithm 2: group sparsity algorithm implemented by ADMM.

Input: X, W (it is omitted for clarity of the algorithm description in the following.)
Initialize H

While stopping criterion is not met do

\[ W = \arg \min_{W} \left( \frac{1}{2} \| X - WH \|_F^2 + \rho \| W - U \|_F^2 + \frac{\rho}{2} \| W - U \|_F^2 \right) \]

\[ U = \arg \min_{U} \left( \frac{1}{2} \| X - WH \|_F^2 + \rho \| W - U \|_F^2 + \frac{\rho}{2} \| W - U \|_F^2 \right) \]

\[ H = \arg \min_{H} \left( \frac{1}{2} \| X - WH \|_F^2 + \rho \| W - H \|_F^2 + \frac{\rho}{2} \| W - H \|_F^2 \right) \]

\[ V = \arg \min_{V} \left( \frac{1}{2} \| X - WH \|_F^2 + \rho \| W - V \|_F^2 + \frac{\rho}{2} \| W - V \|_F^2 \right) \]

\[ E = \arg \min_{E} \left( \frac{1}{2} \| X - WH \|_F^2 + \rho \| W - E \|_F^2 + \frac{\rho}{2} \| W - E \|_F^2 \right) \]

\[ F = \arg \min \left( \frac{1}{2} \| X - WH \|_F^2 + \rho \| W - F \|_F^2 + \frac{\rho}{2} \| W - F \|_F^2 \right) \]

end while

Output: \( W, H \).

3.1. Quantitative Comparison

The above-mentioned algorithms are evaluated using the center location error as well as the overlapping rate [18]. The center tracking errors are presented in Fig. 4 where the best and results are shown with bold red fonts, and the second best ones are shown with blue fonts. The average overlapping errors are presented in Fig. 5 where the best and results are shown with bold red fonts, and the second best ones are shown with blue fonts. The proposed tracking algorithm achieves the best or second best results in most sequences in terms of both success rate and center location error. Overall, the proposed tracker performs well against the other state-of-the-art algorithms.

3.2. Qualitative Comparison

Large pose variations with occlusions. In the basketball sequence, the player undergoes large pose variation and heavy occlusions. When the player is partially occluded by other similar players, the IVT, MIL, L1 and L1-APG methods do

Fig. 3. Implementation of the proposed NMF group sparsity learning algorithm using ADMM.

Fig. 4. The average tracking errors. The error is measured using the Euclidian distance of two center points from the ground truth. The last row is the average error for each tracker over all the test sequences.

Fig. 5. Average overlap rate. The best three results are shown in red, blue, and green fonts.
not perform well. Although, the WMIL track the object center location well, it cannot estimate the size of the objects. Our algorithm tracks the target objects reliably through the sequence.

Occlusions. The target in Caviar sequence undergoes heavy occlusions. In addition, the scale of the object in the caviar sequence changes significantly. The L2-RLS, MIL and MTT methods do not perform well when large scale change occurs. Due to significant scale changes in the caviar sequence, the CT shows limited tracking performance. When heavy occlusions occur, the WMIL and IVT methods start to drift away from the target object. On the other hand, our algorithm tracks the target objects well.

Illumination and pose variations: The objects in Singer and DavidIndoor sequences undergo large appearance changes due to illumination and pose variations. In the Singer sequence, the L1 methods do not perform well. The IVT, MIL, and L2-RLS approaches do not track the object reliably when illumination and pose variations occur together. In addition, the MTT-L01 and MTT-L02 methods do not perform well when scale and large illumination changes occur simultaneously. The WMIL does not deal with large scale changes well. Different from other tracking methods, our algorithm tracks the object favorably for various appearance changes.

Appearance changes: The target object in Shirt sequence undergoes various appearance changes including motion blurs, background clutter, and pose variations. When the target undergoes motion blurs, the L1 and MTT-L02 methods do not perform well. When background clutter occurs, the MTT-L01 and MIL methods drift away from the target objects. On the other hand, the IVT and WMIL methods fail to track the objects well when motion blurs occur. The CT and L1-APG methods do not perform well when large pose changes occur. In contrast, our algorithm performs well which can be attributed to use subspace learning and representation model to handle appearance changes.

Illumination and motion blur: The target objects undergo drastic illumination changes and motion blurs in Fish sequence. Most of trackers do not perform well. While our algorithm tracks the objects well due to the use of subspace model to update the template.

4. CONCLUSION

In summary, based on the subspace learning and NMF group sparsity constraint, we developed a robust tracking method which improved tracking accuracy. The accuracy improvement is achieved via a new object representation model for finding the sparse representation of the target. And it is solved by ADMM numerical solver. Numerous experimental results and evaluations demonstrate the proposed tracker performs favorably against existing state-of-the-art algorithms in the literature.
5. REFERENCES


[5] Zhang, T., Ghanem, B., Liu, S., Ahuja, N., "Robust visual tracking via multi-task sparse learning," In IEEE conference on computer vision and pattern recognition (pp. 1C8).


