Multi-Channel Sparsity Histogram based Particle Filter for Hand Tracking

Xiaowei An
Shandong University of Science and Technology
Qingdao, Shandong, China 266590
anxiaowei2017@aliyun.com

Quanquan Liang✉
Shandong University of Science and Technology
Qingdao, Shandong, China 266590
quanquan_sdust@163.com

Jie Tian
Shandong Normal University
Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center (National Supercomputer Center in Jinan)
Jinan, China 250014
tianjiesdu@gmail.com

ABSTRACT

Sparse representation is a crucial tool for modeling object appearance in the tracking process. Various sparsity models are coded by single feature with fixed local patches, which neglects the potential information inside the appearance model. This way also gives a low robust effect on appearance variations such as partial occlusion, illumination changes and scale variation. In order to resolve above problems, a novel sparsity representation based on multi-channel color features is presented in this paper. Pure color pixel vectors in the local patch are decomposed into several separate dictionaries, which can keep the structural attributes of original model appearance. After constructing sparsity histogram in the local patches, a cosine histogram measurement is proposed for the comparison between template and candidate under the particle filter framework. Finally, a scheme for template update by incremented 2DPCA learning is employed for appearance variation. In order to show the proposed method robust performance, this work employs hand as tracking target that traditional skin color based way which is easily effected by lots of factors and difficult to track the hand target. Qualitative experimental result shows that the proposed tracking algorithm gives a better performance in dynamic scenes.

KEYWORDS

Multi-Channel, Sparsity Histogram, Particle Filter, Hand Tracking

ACM Reference format:
Xiaowei An, Quanquan Liang✉, and Jie Tian. 2017. Multi-Channel Sparsity Histogram based Particle Filter for Hand Tracking. In *Proceedings of 10th EAI International Conference on Mobile Multimedia Communications, Chongqing, China, July, 2017 (MOBIMEDIA 2017),* 5 pages.
DOI: 10.475/123_4

MOBIMEDIA 2017, Chongqing, China

DOI: 10.475/123_4

1 INTRODUCTION

Object tracking is an important research problem in the computer vision, and lots of algorithms have been employed for numerous vision applications for the past years [14]. Generally speaking, estimation of target position is decided by appearance model and localization strategy that own the same importance in the whole process. The main challenges that confused the tracking process is partial occlusion, fore/background clutter, illumination changes, pose and scale variation. Therefore, it is a crucial task to describe the model appearance and optimize the iterative localization steps[9]. With the development of sparsity theory, numerous model schemes have given different explanations in the object tracking field. During the past few years, statistical method based sparse model is more popular under the tracking framework. In Liu et al.[6] provide the sparsity-based mean shift model which selects K discriminative features to model the local patch sparsity histogram. In Zhong et al.[13] propose a sparsity-based model which consists of discriminative and generative features that treat the observation model as sparse representation. However, these sparsity performances are limited by the sparse decomposition which is known that fixed single feature is easy to lose potential structural information in the original color space. Otherwise, [1][2] take a linear combination to model the appearance by a training dictionary, but the template update scheme is only related to holistic template for consideration of the sparsity representation. Since the sparsity appearance model owns the potential structural information about original color pixel vectors which are from the local patches, this paper proposes a robust tracker incorporating color multi-channel histograms for dictionary sparsity construction. For updating template, 2 dimensional principle component analysis(2DPCA)[4] utilizes iterative sparse error matrix to describe the template basis effectively and efficiently. Therefore, the proposed appearance model and

MOBIMEDIA 2017, July 13-14, Chongqing, People's Republic of China

DOI 10.4108/eai.13-7-2017.2270649

observation model based particle filter framework result in low-sensitive and high-correlation among local appearances.

The proposed method offers several advantages in the tracking process:

1)It offers a novel representative format that consists of local model sparsity representation. Obviously, the observation model in the tracking framework is also replaced by multi-channel sparsity histogram evaluation.

2)The multi-channel sparse representation largely exploits the relationship of local variant appearances in appearance model so that it is able to detect the target accurately and is able to run in real time.

3)To address the problem of observation noise, this paper proposes cosine distance as the effective way to find the similarity between multi-channel sparsity histograms .

2 MULTI-CHANNEL SPARSITY HISTOGRAM BASED APPEARANCE MODEL

2.1 MultiChannel Sparsity Representation

With the sparse assumption, this paper proposes that the given signal is decomposed into several linear combinations of a few basis vectors in the sparsity feature space which can be described by original color multi-channel features. Commonly, the appearance of object exists motion smoothness that owns underlying correlation on the local patch regions between consecutive frames. To dig out the potential structure information, this paper splits the region of interest(ROI) into several smaller image patches called image blocks that come from $R/G/B$ channels respectively. This paper employs the first reference image as the root of $Kmeans$ clustering that is plausible to achieve the abundant compactness of data. Considering above factors, movement of micro-distance around the target ROI is the other good choice for enriching composition of clustering . Suppose that the object image patches are collected from 1-th frame to the t-th frame, each of them is partitioned into M overlapping image blocks, and each image block $y_{i,x\in(r,g,b)} \in R^{G\times 1}$ is then stacked into column vectors, where superscript G is the size of the patch.

The sparse coefficient vector $\beta_{i,x\in(r,g,b)} \in R^{G\times 1}$ of each patch can be calculated under the $Kmeans$ sparsity dictionary D by Equ. 1

$$\min_{\beta_{i,x\in(r,g,b)}} ||y_{i,x\in(r,g,b)} - D\beta_{i,x\in(r,g,b)}||_2^2 + \lambda||\beta_{i,x\in(r,g,b)}|| \quad (1)$$

where the sparse coefficient vector $\beta_{i,x\in(r,g,b)}$ corresponding to image block $y_{i,x\in(r,g,b)}$ and λ is a non-negative constant that controls the sparsity of $\beta_{i,x\in(r,g,b)}$.

2.2 Sparsity Histogram Composition

After calculating multi-channel sparsity coefficients in all image local patches, fusion histogram of different t channel can be constructed as Equ. 2:

$$H(h_j^t)_{t\in(r,g,b)} = \sum_i \left|a_{i,j}^t\right| \quad (2)$$

where $H(h_j^t) \in \mathbf{R}^{c\times 1}$ is the sparsity coding histogram for one candidate under the dictionary in which has c atoms. $a_{i,j}^t$ is the j-th coefficient of i-th image patch. In order to keep the robustness of multi-channel attribute, this paper takes the sparsity reconstruction error into consideration. Similar to SCM tracker[13] and the paper[12], fusion histogram should reduce the effect which is from large reconstruction error. For each channel histogram $H(h_j^t)$, this paper adopts a descriptor Q^t to show the degree of reconstruction as following Equ. 3:

$$Q^t = \begin{cases} 1 & {e^t}_i < e_0 \\ 0 & otherwise \end{cases} \quad (3)$$

where ${e^t}_i$ is the reconstruction error of the i-th patch under the t-th channel. so each channel histogram is denoted by Equ. 4

$$H(h_j^t) = H(h_j^t) \otimes Q^t \quad (4)$$

where $\otimes$ stands for point multiplication based element-wise.

3 BAYESIAN INFERENCE TRACKING FRAMEWORK OVER SPARSE CODING HISTOGRAM

3.1 Generic Particle Filter Method

Object tracking is usually reckoned as a Bayesian inference in a Markov model with hidden state variables. The particle filter is popularly used to solve Hidden Markov Chain(HMC) and nonlinear filtering problems arising in signal processing and Bayesian statistical inference. The filtering problem consists of the estimated internal states in dynamical systems when partial observations are made, and random perturbations are present in the sensors as well as in the dynamical system[14][7].

Let X_t denote the state vector which describes the target motion variables in this work, and affine transformation is employed to described six state variables $X_t = \{x_t, y_t, \beta_t, \epsilon_t, \tau_t, s_t\}$, which denote x, y *transition, rotation angle, scale, aspect ratio, and skew direction* at time t.[11]

let $\{Z_t\}$ denotes the corresponding observation vectors. The Bayesian inference can be described as following Equ. 5:

$$\begin{aligned} p(X_t|Z_{1:t-1}) &= \int p(X_t|X_{t-1})p(X_{t-1}|Z_{1:t-1})dX_{t-1} \\ p(X_t|Z_{1:t}) &= \frac{p(Z_t|X_t)p(X_t|Z_{1:t-1})}{p(Z_t|Z_{1:t-1})} \end{aligned} \quad (5)$$

Where $X_{1:t} = \{X_i\}_{i=1:t}$ represents target motion state vectors up to t-th time and $Z_{1:t} = \{Z_i\}_{i=1:t}$ stands for the corresponding observations. $p(X_t|X_{t-1})$ is the state transition model between the reference frame and candidate frame. $p(Z_t|X_t)$ is the observation model that evaluates the similarity of a target choice.

Usually speaking, particle filter treats the posterior $p(X_t|Z_{1:t})$ as N weighted sampling particles $\{X_t^i, W_t^i\}_{i=1,\ldots,N}$. According to the different global weights $\{W_t^i\}_{i=1,\ldots,N}$ in the distribution, the state $\widehat{X}_t = \sum_{i=1}^N W_t^i X_t^i$ can be predicted.

3.2 Likelihood Measurement Based On Multi-Channel Sparsity Histogram

Once a sparsity collection of target ROI has been determined between consecutive frames, the most crucial task is to evaluate which candidate ROI is the best match. As we known that Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them[10]. Suppose that there are two multi-channel sparsity histograms denoting a pair of image patches, resorting to "cosine similarity", the likelihood can be shown over them:

$$\rho(H_r^t, H_c^t) = \frac{< H_r^t, H_c^t >}{||H_r^t||||H_c^t||} \tag{6}$$

Here H_r^t and H_c^t are multi-channel sparsity of histograms from reference and candidate frame, respectively. Observation model that reflects the variations of target appearance caused by pose changes, illumination variations and partial occlusion is shown as Equ. 6 in the inference framework.

3.3 Dynamic Update Scheme with Template

Dynamical update scheme with the template is very necessary for appearance variation. Common Bayesian tracking depends on the matching score between consecutive frames with involving some imprecise samples that lead to tracking drift problem. For handling the appearance change efficiently, Lim et al.[5] exploits one dimensional principle component analysis approach for updating the tracking template. However, $2D$ images directly with the linear algebra method preserve much more sufficient spatial information that facilitates the dimension compression of original features. In this paper, initial N optimal filtering states are able to initialize templates.

Considering accumulated N normalized optimal filtering results $F = \{F_k\}(k = 1, 2, ..., N)$, whose size is $m \times m$. A column of orthogonal vectors $U_{m\times m}$ project the original F into new features G as shown in Equ. 7 and Equ. 8.

$$\begin{aligned} F &= [F_1\ F_2\ F_3\ F_4\ F_5\ F_6\ \cdots F_N] \\ G &= FU_{m\times m} \end{aligned} \tag{7}$$

2DPCA[4] can construct the whole co-variance matrix C_G,

$$\begin{aligned} C_G &= E[(G - E(G))^T(G - E(G))] \\ &= E[(FU - E(FU))^T(FU - E(FU))] \\ &= U^T E[(F - E(F))^T(F - E(F))]U \\ &= U^T C_F U \end{aligned} \tag{8}$$

In fact, orthogonal vector $U_{m\times m}$ is a matrix rather than a scalar that results in sparse basis vector for modeling reconstruction. If new normalized reference ROI are obtained, that is immediately projected into discriminative feature space τ^t by left-projection $U_{m\times m}^T$ and right-projection $U_{m\times m}$. Meanwhile, the updated dictionary can be effectively composition by $Kmeans$ clustering[3] under the τ^t subspace, whereas the new multi-channel histogram is also robustly inferred.

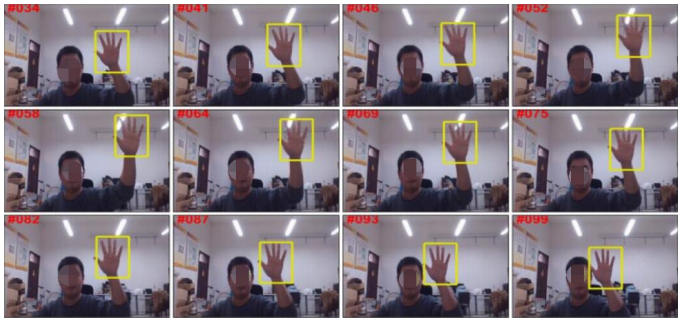

Figure 1: Hand motion with stable pose

4 EXPERIMENTS AND ANALYSIS

To demonstrate the performance of the proposed method, experiments on hand tracking are presented in the paper. Here are several scenarios tested on the sequences. For each sequence, the location of the target object is labeled manually in the first frame. Sliding windows are overlapping type under the same normalization, which produces a cluster of 6×6 image patches in size. Affine transform parameters in the particle tracking are all normalized into the same size 32×32 pixels. Number of Sampled particle is 600. The proposed algorithm is implemented using $Matlab$ on a computer environment of $2.66 - GHz$ Intel Pentium(R) CPU and 8.0 GB memory.

4.1 Qualitative Performance

For the scenarios of the hand tracking, several principal factors that have influences on the appearance of the hand are considered. According to the characteristics of the cases, qualitative discussion is detailed in the following.

4.1.1 Stable Pose Motion. As shown in Fig. 1, hand is moving with stable pose. Meanwhile, the 52-th frame shows that the proposed algorithm also supports a certain illumination-invariant stability. For the reason that multi-channel histogram takes a fusion of multi-channel sparse representation.

4.1.2 Small Pose-variation. As shown in Fig.2, hand is moving with small pose-variation. In this sequence, hand moves under small scale motion and with fingers closing. The proposed algorithm is robust to locate the hand position and label the hand completely.

4.1.3 Large Pose-variation. As shown in Fig. 3, hand is moving with large pose-variation which is the most challenging task for the proposed algorithm.

4.1.4 Similar Color Region. As shown in Fig. 4, hand is moving away from the face where exists similar color to the target hand. But the proposed algorithm keep the tracker without drifting away.

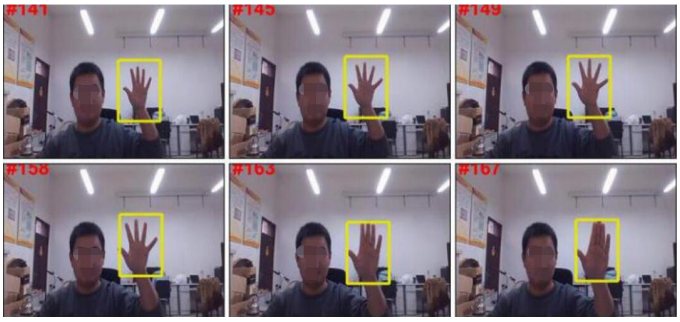

Figure 2: Hand motion with small pose-variation

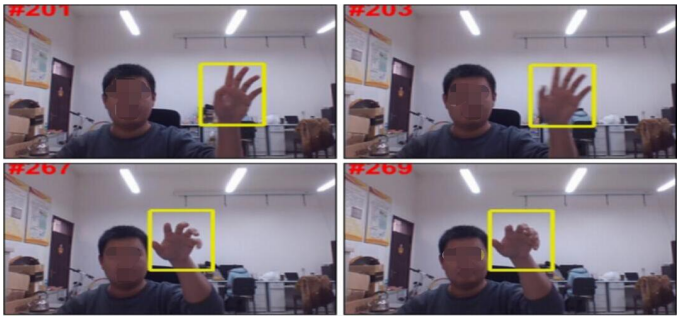

Figure 3: Hand motion with large pose-variation

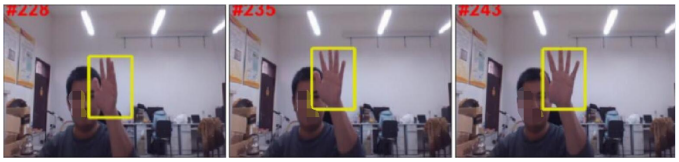

Figure 4: Hand motion near the region of similar color

4.2 Discussion on Stability of Performance

The proposed method is able to achieve robust performance for above scenarios with many challenging factors. Processing time is most cost by composition of dictionary and construction of multi-channel histogram. Target is represented by sparsity approximation using Least absolute shrinkage and selection operator (Lasso)[8] in this method. Therefore, it is possible to improve the proposed method by Orthogonal Matching Pursuit (OMP)[8] that computational cost is much less than that of the other minimization with standard convex programming for sparsity coding problems. In addition, the size of the patch is a trade-off between computational efficiency and effectiveness of modeling appearance. The proposed algorithm also encounter some failure cases that the other object occluded the hand target fully.

5 CONCLUSIONS

This paper presents a Multi-Channel Sparsity Histogram based Particle Filter for Hand Tracking algorithm. Considering the potential characteristics of overlapping patches, sparsity structural information among patches is exploited to model more accurate model of the target appearance. To dynamically depict the model appearance, this paper takes the sparsity reconstruction error into the update scheme. In order to construct more robust template, online learning by 2DPCA solution is employed to make the update scheme suitable for scenarios where exist appearance variation and occlusion as well. Based on the cosine evaluation between histograms, the particle filter achieves a better performance in the hand tracking process. Currently, we are working on a new algorithm that merges more discriminitive features which can dig out more and more sparsity information for the robust tracking.

ACKNOWLEDGMENTS

This research was financially supported by the "Open Research Fund from Shandong provincial Key Laboratory of Computer Network under Grant No.: SDKLCN-2016-02" and "Science and Technology Research Program for Colleges and Universities in Shandong Province, China under Grant No.: J17KA075"

REFERENCES

[1] Tianxiang Bai and Y. F Li. 2012. *Robust visual tracking with structured sparse representation appearance model.* Elsevier Science Inc. 2390–2404 pages.

[2] Hui Ji. 2012. Real time robust L1 tracker using accelerated proximal gradient approach. In *IEEE Conference on Computer Vision and Pattern Recognition*. 1830–1837.

[3] Pilsung Kang and Sungzoon Cho. 2009. K -Means Clustering Seeds Initialization Based on Centrality, Sparsity, and Isotropy. In *IDEAL (Lecture Notes in Computer Science)*, Emilio Corchado and Hujun Yin (Eds.), Vol. 5788. Springer, 109–117. http://dx.doi.org/10.1007/978-3-642-04394-9

[4] X. Li, Y. Pang, and Y. Yuan. 2010. L1-norm-based 2DPCA. *IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society* 40, 4 (2010), 1170–1175.

[5] Jongwoo Lim, David A Ross, Ruei-Sung Lin, and Ming-Hsuan Yang. 2004. Incremental Learning for Visual Tracking.. In *Nips*, Vol. 17. 793–800.

[6] Baiyang Liu, Junzhou Huang, Lin Yang, and C Kulikowsk. 2011. Robust tracking using local sparse appearance model and K-selection. In *Computer Vision and Pattern Recognition*. 1313–1320.

[7] Katja Nummiaro, Esther Koller-Meier, and Luc Van Gool. 2003. An adaptive color-based particle filter. *Image and vision computing* 21, 1 (2003), 99–110.

[8] Bruno A Olshausen and David J Field. 1997. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision research* 37, 23 (1997), 3311–3325.

[9] Fatih Porikli and Alper Yilmaz. 2012. *Object Detection and Tracking.* Springer Berlin Heidelberg. 12 pages.

[10] Lokesh Sahu and Biju R. Mohan. 2014. An improved K-means algorithm using modified cosine distance measure for document clustering using Mahout with Hadoop. *2014 9th International Conference on Industrial and Information Systems (ICIIS)* (2014), 1–5.

[11] Rudolph van der Merwe, Arnaud Doucet, Nando de Freitas, and Eric A. Wan. 2000. The Unscented Particle Filter. In *NIPS*.

[12] Zhongpei Wang, Hao Wang, Jieqing Tan, Peng Chen, and Chengjun Xie. 2016. Robust object tracking via multi-scale patch

based sparse coding histogram. *Multimedia Tools & Applications* (2016), 1–23.

[13] Ming Hsuan Yang, Huchuan Lu, and Wei Zhong. 2012. Robust object tracking via sparsity-based collaborative model. In *Computer Vision and Pattern Recognition*. 1838–1845.

[14] Alper Yilmaz, Omar Javed, and Mubarak Shah. 2006. Object tracking: A survey. *Acm Computing Surveys (CSUR)* 38, 4 (2006), 13.