Proceedings of the 2014 IEEE
International Conference on Robotics and Biomimetics
December 5-10, 2014, Bali, Indonesia

DCT Representations Based Appearance Model for Visual Tracking

Hongkai Chen, Weizheng Zhang, Xiaoguang Zhao and Min Tan

Abstract— **In this paper, we present a simple yet effective visual tracking algorithm with an appearance model based on 2D discrete cosine transform (2D-DCT) representations. The DCT has the properties of decorrelation and energy compaction, and is robust against geometry and illumination changes. Hence, it is suitable for appearance modeling and the features of our appearance model are extracted from an optimized low dimensional subspace. In order to adapt to the appearance change caused by environment change or ego motion, we also propose to update the observation appearance model through a nonlinear weighted method. Numerous experiments on some challenging video sequences demonstrated that our algorithm is effective and it considerably outperforms the other methods.**

I. Introduction

In recent years, visual tracking is one of the major research issue in computer vision and has many successful applications such as the video surveillance, robot vision and human robot interaction [1]. Although there are many works [2]–[12] proposed to solve the visual tracking problem during the last decades, it is still difficult to design a robust tracking system due to some factors, such as occlusion, illumination variation, deformation, scale and appearance change, etc. In general, a typical visual tracking problem covers three key components: appearance model, tracking strategy and appearance model update strategy.

The core problem of visual tracking is how to represent the target object. Target representation is implemented through the object appearance modeling. Hence, how to construct an effective appearance model plays a critical role in visual tracking. There are many ways to build an appearance model for the target. [13], [14] and [15] used interest points for tracking. In [16], Yilmaz *et al.* used contour feature to track object, but the high computational cost of the level set method limits its application in real-time scenarios. The histogram reflects the global statistic information of the target and the histogram representation is often used for appearance modeling due to its good anti-noise performance [17]. In view of this, Bradski proposed a tracking algorithm based on color histogram [18] and Nummiaro *et al.* proposed to use color histogram in particle filter framework for tracking [19]. Besides, histograms of oriented gradients (HOG) are also used for appearance modeling [20]. Wang *et al.* proposed a Gaussian Mixture Model fusing the spatial and the color information for the sake of using the statistical information of color feature as time changes [21]. Although it makes the target representation more accurate, it may fail when suffering from heavy posture change. Recently, the subspace method is proposed to use the history information of the target appearance to adapt to the object appearance change [9]. Furthermore, there are also some other appearance modeling methods such as tensor representation [22], MRF [23], saliency model [24], etc.

Besides, tracking strategy is also very important in visual tracking. In mean shift, non-parameter kernel density estimation method was adopted to handle the tracking problem [3]. It is considered as a local pattern matching optimization problem. However, it may fail when suffering from color clutter in background, illumination change and occlusion. Recently, the discriminative model based algorithms have been attracting much attention [1], [7], [10]. These methods take visual tracking as a binary classification task aiming at discriminate the interesting object from the background. Although they are suitable for situation with deformation, occlusion, scale and appearance changes, their high computational cost may limit their real-time applications. Furthermore, visual tracking can also be casted as a state estimation issue using the state space method. As a Bayesian sequential importance sampling technique, the particle filter can be used to estimate the posterior distribution of state variables [25]. It can handle non-Gaussian or nonlinear model and is able to run in real-time. This paper uses the particle filter as tracking framework. And the target state will be estimated sequentially via particle filter during the process of tracking.

In this paper, we present a simple yet effective visual tracking algorithm with an appearance model based on 2D discrete cosine transform (2D-DCT) representations. The DCT is robust against geometry and illumination, and has the properties of decorrelation and energy compaction. Thus the original image can be approximately described via using the low frequency components without introducing visual distortion in the reconstructed image. Hence, it is suitable for appearance modeling. In our appearance model, the features are extracted from an optimized low dimensional subspace. Besides, we also propose to update the observation appearance model through a nonlinear weighted method to adapt to the appearance change caused by environment change or ego motion. Numerous experiments on some challenging video sequences demonstrated that our algorithm is effective and it considerably outperforms the other methods.

The organizational structure of the remainder of this paper is as follows. In Section II, the detail of our target appearance

This work was supported by the National Natural Science Foundation of China (61271432, 61273337).

Hongkai Chen, Xiaoguang Zhao and Min Tan are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation,Chinese Academy of Sciences, Beijing 100190, China.

Weizheng Zhang is with State Grid Henan Electric Power Company, Zhengzhou 450052, China.

Email: {hongkai.chen, xiaoguang.zhao, min.tan}@ia.ac.cn

978-1-4799-7397-2/14/$31.00 © 2014 IEEE

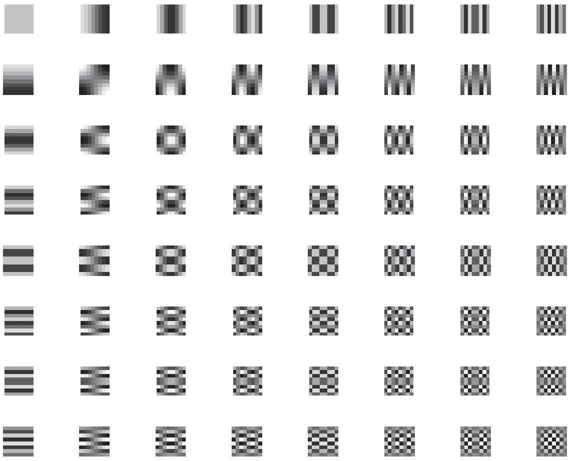

Fig. 1. This figure shows the 2-D DCT basis images for $N = 8$. Each patch represents one DCT basis image. Neutral gray represents zero, white represents positive amplitudes, and black represents negative amplitude.

model is given. Section III shows our tracking strategy with particle filter. The experimental results are shown in Section IV. Finally, we give some conclusions in Section V.

II. Target Appearance Model

In [26], Khayam claimed that the Discrete Cosine Transform (DCT) attempts to decorrelate the discrete signal via using a set of mutually uncorrelated cosine basis functions in a linear manner without losing compression efficiency. The DCT is often used for feature extraction due to its robustness against geometry and illumination changes. It has many applications in pattern recognition, computer vision, and multimedia [27], such as face recognition [28], image retrieval [29], etc. In this section, We will show how to construct an effective yet very simple appearance model based on 2D-DCT representations for visual tracking.

A. The One-Dimensional DCT

The 1D Discrete Cosine Transform (1D-DCT) and Inverse Discrete Cosine Transform (IDCT) are defined as follow

$$c(u) = \sum_{x=0}^{N-1} \alpha(u) f(x) cos[\frac{\pi(2x+1)u}{2N}] \tag{1}$$

$$f(x) = \sum_{u=0}^{N-1} \alpha(u) c(u) cos[\frac{\pi(2x+1)u}{2N}] \tag{2}$$

where $u \in \{1,2,\cdots,N-1\}$, $f(x)$ is the 1-D discrete sequence, N is the length of $f(x)$. In both equations (1) and (2), $\alpha(u)$ is defined as

$$\alpha(u) \quad = \quad \begin{cases} \sqrt{\frac{1}{N}}, & u=0 \\ \sqrt{\frac{2}{N}}, & others \end{cases} \tag{3}$$

We define $p(u,x)$ as $p(u,x) = \alpha(u)cos[\frac{\pi(2x+1)u}{2N}]$, then (2) can be rewritten as

$$f(x) = \sum_{u=0}^{N-1} c(u)p(u,x) \tag{4}$$

where $p(u,x)$ is the 1D-DCT cosine basis functions which are orthogonal.

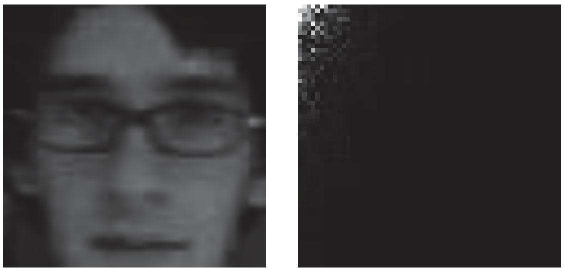

Fig. 2. The source image (80×80 pixels) is transformed from the pixel domain to the frequency domain through DCT transform. In general, images are highly redundant and relevant. Hence only a small part of the frequency components of the coefficient are not zero, most coefficient are zero (or close to zero) after transforming into the frequency domain. In this figure, from the top left to the bottom right corner, as frequency increasing, the values in the top left corner are relatively large, while the bottom right corner small. In other words, the energy of an image is almost concentrated in the low frequency components in the top left corner.

B. The Two-Dimensional DCT

The 2D Discrete Cosine Transform (2D-DCT) and Inverse Discrete Cosine Transform (IDCT) are defined as follow

$$c(u,v) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)q(u,v,x,y) \tag{5}$$

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} c(u,v)q(u,v,x,y) \tag{6}$$

where $q(x,y,u,v) = \alpha(u)\alpha(v)cos[\frac{\pi(2x+1)u}{2N}]cos[\frac{\pi(2y+1)v}{2N}]$, and $q(u,v,x,y)$ is the 2D-DCT cosine basis functions which are orthogonal. $u \in \{1,2,\cdots,N-1\}$, $v \in \{1,2,\cdots,N-1\}$. $f(x,y)$ is the 2D discrete signal (e.g. digital image) with size of $N \times N$. The definition of $\alpha(u)$ and $\alpha(v)$ are the same as (3). And Fig. 1 shows the cosine basis functions for N=8 [29].

In Fig. 1, each patch represents one DCT basis image $q_i(x,y)$ corresponding to the DCT coefficient $c(u,v)$. According to (6), an original image can be expressed as a linear combination of DCT basis image blocks after projected onto the DCT coefficient space, and the corresponding weighted coefficient is $c(u,v)$. Then we can use a certain order (such as zig-zag sequence [29]) to scan the 2D coefficients and retain a one dimensional vector $\vec{C} = (c_0,c_1,\cdots,c_{N\times N-1})^T$. Thus, the (6) can be rewritten as

$$f(x,y) = \sum_{i=0}^{N\times N-1} c_i q_i(x,y) \tag{7}$$

C. The Properties of DCT

The Discrete Cosine Transform (DCT) has some properties which are of particular value to digital image processing applications.

1) Decorrelation: the DCT is an orthogonal transformation and can transform the covariance matrix into diagonalization. It means the components are uncorrelated (i.e. decorrelation characteristic). Besides, the DCT is an effective approximation of the KL transform, and is often used in the area of image processing for the purpose of feature selection.

2) Energy compaction: after DCT, we can obtain a coefficient matrix with the same size of the input image. In the coefficient matrix, the position and amplitude of every element reflects the spatial frequency and energy of an image. Low frequency components of the coefficient matrix will be located in the top left corner, while high frequency components in the bottom right corner.

As shown in Fig. 2, the energy is concentrated in the low frequency components and the DCT coefficients in these regions have much larger values. As claimed in [26], the original image can be approximately described via using the low frequency components without introducing visual distortion in the reconstructed image. That is to say, low frequency components describe most of the original image information, while the high frequency components describe the detail. Or rather, the high frequency components provide the details and the low frequency components offer a framework.

D. DCT Representations

As shown in (4) and (7), the discrete signal (one-dimensional discrete sequence signal or digital image) can be expressed as a linear combination of mutually uncorrelated DCT cosine basis functions. Based on these functions, the original signal is converted to the orthogonal DCT coefficient space. Besides, the output of the DCT is often very sparse, especially suitable for signal compression and dimension reduction for data.

At the given time t, the target region in the image frame f_t is denoted by R_t. Do DCT operation on the image region R_t then produce a coefficient matrix $\mathbf{C}$ with the same size of R_t. In this paper, we use the zig-zag sequence to scan the top left part of the 2D coefficients and retain a one dimensional vector. Then reserve the first K (here K = 64) elements in the top left corner, thus the other coefficient components with small coefficient will be discarded. The one dimensional vector is $\vec{C}_t = (c_t^0, c_t^1, \cdots, c_t^{K-1})^T$. Thus we can use the first K DCT coefficient to describe the target image region. And the target image region R_t can be approximately expressed by

$$R_t = \sum_{i=0}^{K-1} c_t^i q_i(x,y) \tag{8}$$

This operation is equivalent to the image projected onto an optimized low dimensional subspace, and the corresponding feature vector represents the target appearance model and the position in the subspace.

E. Update of Observation Appearance Model

In visual tracking, the appearance of the target may change over time due to some factors, such as illumination variation, pose change, etc. If we use a fixed observation appearance model, the tracking algorithm can not adapt to the environment change and ego motion, and it will cause tracking failure. If we directly update the observation appearance model using the newest image sample to adapt to the appearance change of the target, the appearance model will degenerate when encountering occlusion. Finally, the tracker will result in drift.

In this paper, we propose a method to update the observation appearance model via a nonlinear weighted method. The appearance of the target may change over time. However, the change is very small between two consecutive video frames, i.e. only a small details change. As we know, the high frequency components provide the details and the low frequency components offer a framework. So the small details change occur in high frequency components. Hence, we assign large weight to high frequency components, and the low frequency components is assigned small weight. The appearance model can be updated as follow:

$$\vec{C}^* = (\vec{\mathbf{1}} - \vec{\alpha}) \cdot \vec{C}^* + \vec{\alpha} \cdot \vec{C}_t^* \tag{9}$$

where $\vec{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_K)^T$. α_i is a weight coefficient and its corresponding DCT coefficient value is c_t^i. Before transformed to a one dimensional vector, the position coordinate of c_t^i in 2-D DCT coefficient matrix $c(u,v)$ is (u,v). So α_i can be computed by $\alpha_i = exp\{-\frac{(u-1)^2+(v-1)^2}{\lambda^2}\}$ where λ is a constant coefficient. And $\vec{\mathbf{1}} = (1,1,\cdots,1)^T$ is a column vector with the same size of $\vec{\alpha}$. $\vec{C}^*$ is the target appearance model. $\vec{C}_t^*$ is the observation appearance of the tracked target at the t-th video frame.

III. Visual Tracking with Particle Filter

Visual tracking can be viewed as a Bayesian inference task in Markov model with hidden state variables [9]. The task of visual tracking is to estimate the state of target in image frame. We use a rectangular bounding box to describe the target in the video frame, namely the state of the target. Each target image region is resized to a standard 32×32 pixels sub-image patch. Over the given time t, the state of the target is defined as $X_t = \{x_t, y_t, sx_t, sy_t\}$, where (x_t, y_t) represents the pixel coordinate of the target rectangular bounding box center in the image frame f_t. sx_t and sy_t represent the ratio of the target's width and height to the sub-image's width and height, respectively. Define Z_t as the observed appearance at the t-th frame. Given a set of observed appearance sequence $Z_{1:t}$, the posteriori probability density of the target state X_t can be estimated by

$$p(X_t|Z_{1:t}) \propto p(Z_t|X_t)p(X_t|Z_{t-1}) \tag{10}$$

$$p(X_t|Z_{t-1}) = \int p(X_t|X_{t-1})p(X_{t-1}|Z_{1:t-1})dX_{t-1} \tag{11}$$

where $p(X_t|Z_{1:t-1})$ is the prior probability predicted through using the historical observed appearance, and $p(Z_t|X_t)$ denotes observation appearance model using DCT representations, aiming at estimating the likelihood of observation Z_t at the state X_t. $p(X_t|X_{t-1})$ is the state transition model between two consecutive states and can be formulated by a Gaussian distribution around its previous state X_{t-1}, i.e. $P(X_t|X_{t-1}) = N(X_t;X_{t-1},\Sigma)$, where Σ is a diagonal covariance matrix.

Then the optimal state of the tracked target $\hat{X}_t$ can be obtained by the Maximum a Posteriori (MAP) estimation over the N samples at each time t by:

$$\hat{X}_t = \arg\max_{X_t^i} p(X_t^i|Z_{1:t}), i = 1,2,\cdots,N \tag{12}$$

Particle filter is an approximate algorithm for solving Bayesian estimation by non-parametric Monte Carlo method, and is also known as Sequential Monte Carlo method. It is suitable for nonlinear, non-Gaussian and multimode state estimation problem of dynamic system, and is a very active research field in visual tracking. In our particle filter framework, we employ the standard Condensation algorithm [25], which maintains N weighted particles $X_t^i(i = 1,2,\cdots,N)$ at the t-th frame f_t, to simulate a posteriori probability distribution. These particles are drawn from an importance distribution function $q(X_t|X_{0:t-1},Z_{1:t})$, then the posteriori probability density of the target state X_t can be represented as

$$p(X_t^i|Z_{1:t}) \propto \sum_{i=1}^{N} \omega_t^i \delta(X_t - X_t^i) \tag{13}$$

where $\delta(\cdot)$ is a Dirac function and the weights $\omega_t^i(i = 1,2,\cdots,N)$ of particles are updated recursively as

$$\omega_t = \omega_{t-1} \frac{p(Z_t|X_t)p(X_t|X_{t-1})}{q(X_t|X_{0:t-1},Z_{1:t})} \tag{14}$$

If we make $q(X_t|X_{0:t-1},Z_{1:t})$ equal to $p(X_t|X_{t-1})$, we can obtain $\omega_t = \omega_{t-1}p(Z_t|X_t)$. That is to say the weights of particles over time t are given by the observation likelihood.

In order to avoid degeneracy for each particle, the particles are resampled according to their weights. Those particles with higher weights have much great possibility to be resampled, and the particles with lower weights have lower possibility to be resampled.

IV. Experiments

Our algorithm is tested on 10 challenging video sequences using VS2008 on a PC with Inter i5-2400 CPU (3.1GHz) with 4GB memory, and runs at 20 frames per second (FPS). The adopted video sequences shown in Table I are from previous works [8]–[10], [30], [31] which are all publicly available. The challenges of these sequences include illumination variation, occlusion, fast direction change, pose change, viewpoint change, scale and appearance change. Besides, we also test some state-of-the-art tracking algorithms using the source code provided by the authors for comparison. These tracking algorithms include the color-based particle filter (CPF) [2], IVT [9], MILTracker [10] and CT [12] algorithms.

To quantitatively evaluate the performance of the proposed algorithm with other trackers, we use two metric in this paper. The first metric is the center location error which is defined as the average Euclidean distance between the center location of the tracked target and the manually labeled ground truth data. The other is the success rate using bounding box overlap. The overlap score is defined as

TABLE I
The tracking video sequences used in our experiments

Sequence	Main Challenges
david_indoor	moving camera,illumination change,scale change
Dudek	illumination change,scale change,pose change
Faceocc1	moving camera,occlusion
Faceocc2	occlusion,heavy pose change
fish	illumination change,fast direction change
juice	moving camera,fast direction change,scale change
person	moving camera,occlusion
rubikscube	moving camera,viewpoint change
Singer1	illumintation variation,scale change
Sylvester	illumination change,pose change

$$S = \frac{|r_t \cap r_a|}{|r_t \cup r_a|} \tag{15}$$

where r_t is the tracked bounding box of the target, and r_a is the manually labeled ground truth bounding box. $\cap$ and $\cup$ denote the intersection and union of two regions respectively. $|\cdot|$ measures the number of pixels in the region. Then we count the number of successful image frames whose overlap score S is large than the given threshold S_{thr} (i.e. $S > S_{thr}$, here we set $S_{thr} = 0.5$) [32]. The success rate is defined as the ratio between the successful frames and the total frames.

Both Table II and Table III are the quantitative results. Table II shows the average center location errors of tracking algorithm and Table III shows the success rate.

A. Scale, Pose and Illumination Variation

The target in *david_ind* sequence undergoes illumination, scale and pose change. The IVT, CT and our methods perform well on this sequence. For the *Singer1* sequence, the scale changes gradually. However it undergoes heavy illumination variation. When the stage light changes drastically, most of the other algorithms fail to track the target. As the light and scale changing, all the other trackers fail. However, the proposed tracker is robust to scale and illumination changes due to the target appearance can be modeled well by DCT representations and the update method with nonlinear weight. Besides, Our tracker performs well on the *Dudek* and *Sylvester* sequences where the targets undergo scale, pose and illumination variation.

B. Occlusion and Pose Change

The target in *Faceocc2* sequence undergoes heavy occlusion and large pose change. The CPF, IVT and CT methods perform well on this sequence. Our method performs best not only in *Faceocc2*, but also in the *Faceocc1* sequence and *person* sequence.

C. View Point Variation and Fast Direction Change

The object in *juice* sequence undergoes fast direction change because of fast moving camera. Besides, its scale also changes gradually. Only the IVT method and our

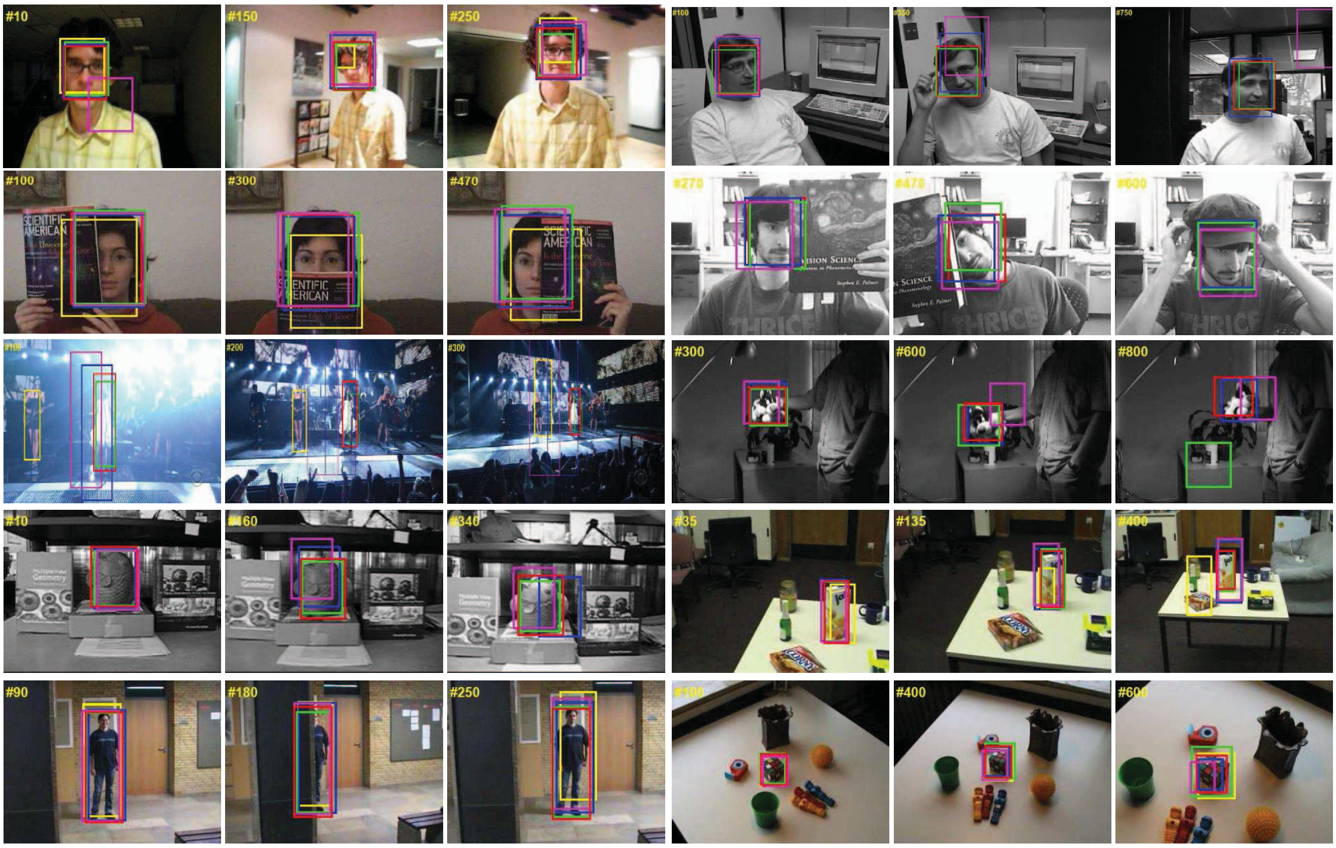

Fig. 3. Screenshots of sampled tracking results. The results of proposed tracker, CPF, IVT, MILTracker and CT algorithms are indicated by red, yellow, green, pink and blue boxes, respectively.

TABLE II

AVERAGE CENTER LOCATION ERRORS OF TRACKING ALGORITHMS.

Sequence	Frames	CPF	IVT	MIL	CT	Our
david_indoor	471	19	**6**	25	12	9
Dudek	1145	-	18	151	20	**14**
Faceocc1	892	21	19	28	18	**14**
Faceocc2	812	-	14	18	**10**	11
fish	476	-	**5**	13	26	8
juice	405	31	**5**	7	7	6
person	306	12	**3**	9	7	4
rubikscube	717	6	**3**	9	4	**3**
Singer1	352	114	13	62	24	**7**
Sylvester	1346	-	99	45	13	**10**
Average	-	34	18	37	14	**9**

tracker perform well on this sequence. For the *fish* sequence, the illumination changes gradually and this sequence also undergoes fast direction change. Only the IVT, MILTracker and our tracker perform well on this sequence. In addition, in the *rubikscube* sequence, our tracker achieves the best result both in average location error and success rate metric.

V. CONSLUSIONS

In this paper, we proposed a simple yet effective visual tracking algorithm with an appearance model based on 2D discrete cosine transform (2D-DCT) representations. The DCT is robust against geometry and illumination, and has the properties of decorrelation and energy compaction. The original image can be approximately described via using the low frequency components without introducing visual distortion in the reconstructed image. Hence, it is suitable for appearance modeling. The features of our appearance model are extracted from an optimized low dimensional subspace. Besides, we propose to update the observation appearance model through a nonlinear weighted method to adapt to the appearance change caused by environment change or ego motion. Besides, it has the potential to work in real-time applications. Numerous experiments on some challenging video sequences demonstrated that our algorithm is effective and it considerably outperforms the other methods.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: A survey,” *Acm Computing Surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.

[2] P. Prez, C. Hue, J. Vermaak, and M. Gangnet, *Color-based probabilistic tracking*. Springer, 2002, pp. 661–675.

[3] D. Comaniciu, V. Ramesh, and P. Meer, “Kernel-based object tracking,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 5, pp. 564–577, 2003.

[4] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi, “Robust online appearance models for visual tracking,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 10, pp. 1296–1311, 2003.

[5] S. Avidan, “Support vector tracking,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 8, pp. 1064–1072, 2004.

TABLE III

SUCCESS RATE (SR)(%): PERCENTAGE OF CORRECT TRACKED FRAMES BASED ON THE OVERLAP SCORE CRITERION ($> 50\%$).

Sequence	Frames	CPF	IVT	MIL	CT	Our
david_indoor	471	23	**88**	68	87	87
Dudek	1145	-	88	18	90	**91**
Faceocc1	892	92	92	69	96	**100**
Faceocc2	812	-	91	82	97	**100**
fish	476	-	**100**	87	23	98
juice	405	36	81	42	44	**82**
person	306	88	**97**	90	90	94
rubikscube	717	98	99	72	89	**100**
Singer1	352	13	34	20	25	**100**
Sylvester	1346	-	45	24	**83**	79
Average	-	58	82	57	72	**93**

[6] R. T. Collins, L. Yanxi, and M. Leordeanu, "Online selection of discriminative tracking features," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1631–1643, 2005.

[7] S. Avidan, "Ensemble tracking," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, Conference Proceedings, pp. 494–501 vol. 2.

[8] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, Conference Proceedings, pp. 798–805.

[9] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125–141, 2008.

[10] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1619–1632, 2011.

[11] B. Chenglong, W. Yi, L. Haibin, and J. Hui, "Real time robust l1 tracker using accelerated proximal gradient approach," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, Conference Proceedings, pp. 1830–1837.

[12] K. Zhang, L. Zhang, and M.-H. Yang, *Real-time compressive tracking*. Springer, 2012, pp. 864–877.

[13] S. Battiato, G. Gallo, G. Puglisi, and S. Scellato, "Sift features tracking for video stabilization," in *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*. IEEE, Conference Proceedings, pp. 825–830.

[14] T. Mathes and J. H. Piater, *Robust non-rigid object tracking using point distribution manifolds*. Springer, 2006, pp. 515–524.

[15] W. He, T. Yamashita, H. Lu, and S. Lao, "Surf tracking," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, Conference Proceedings, pp. 1586–1592.

[16] A. Yilmaz, X. Li, and M. Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 11, pp. 1531–1536, 2004.

[17] J.-K. Kamarainen, V. Kyrki, J. Ilonen, and H. K-lviinen, "Improving similarity measures of histograms using smoothing projections," *Pattern Recognition Letters*, vol. 24, no. 12, pp. 2009–2019, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865503000394

[18] G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," 1998.

[19] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," *Image and Vision Computing*, vol. 21, no. 1, pp. 99–110, 2003.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, Conference Proceedings, pp. 886–893 vol. 1.

[21] W. Hanzi, D. Suter, K. Schindler, and S. Chunhua, "Adaptive object tracking based on an effective appearance filter," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 9, pp. 1661–1667, 2007.

[22] L. Xi, H. Weiming, Z. Zhongfei, Z. Xiaoqin, and L. Guan, "Robust visual tracking based on incremental tensor subspace learning," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, Conference Proceedings, pp. 1–8.

[23] L. Wen-Chieh and L. Yanxi, "A lattice-based mrf model for dynamic near-regular texture tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 5, pp. 777–792, 2007.

[24] A. Borji, S. Frintrop, D. N. Sihite, and L. Itti, "Adaptive object tracking by learning background context," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, Conference Proceedings, pp. 23–30.

[25] M. Isard and A. Blake, "Condensationconditional density propagation for visual tracking," *International journal of computer vision*, vol. 29, no. 1, pp. 5–28, 1998.

[26] S. A. Khayam, "The discrete cosine transform (dct): theory and application," *Michigan State University*, 2003.

[27] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *Computers, IEEE Transactions on*, vol. C-23, no. 1, pp. 90–93, 1974.

[28] J. Xiao-Yuan and D. Zhang, "A face and palmprint recognition approach based on discriminant dct feature extraction," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, no. 6, pp. 2405–2415, 2004.

[29] G. K. Wallace, "The jpeg still picture compression standard," *Consumer Electronics, IEEE Transactions on*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[30] D. A. Klein, D. Schulz, S. Frintrop, and A. B. Cremers, "Adaptive real-time video-tracking for arbitrary objects," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, Conference Proceedings, pp. 772–777.

[31] K. Junseok and L. Kyoung-Mu, "Visual tracking decomposition," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, Conference Proceedings, pp. 1269–1276.

[32] 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.