Hindawi Publishing Corporation
Advances in Multimedia
Volume 2014, Article ID 343860, 11 pages
http://dx.doi.org/10.1155/2014/343860

*Research Article*

# Object Tracking with Adaptive Multicue Incremental Visual Tracker

**Jiang-tao Wang,[1] De-bao Chen,[1] Jing-ai Zhang,[1] Su-wen Li,[1] and Xing-jun Wang[2]**

[1] *School of Physical and Electronic Information, Huaibei Normal University, Huaibei 235000, China*
[2] *Shandong Huisheng Group, Weifang 261201, China*

Correspondence should be addressed to Jing-ai Zhang; ellazhangja@126.com

Received 27 February 2014; Revised 25 August 2014; Accepted 28 August 2014; Published 23 September 2014

Academic Editor: Constantine Kotropoulos

Generally, subspace learning based methods such as the Incremental Visual Tracker (IVT) have been shown to be quite effective for visual tracking problem. However, it may fail to follow the target when it undergoes drastic pose or illumination changes. In this work, we present a novel tracker to enhance the IVT algorithm by employing a multicue based adaptive appearance model. First, we carry out the integration of cues both in feature space and in geometric space. Second, the integration directly depends on the dynamically-changing reliabilities of visual cues. These two aspects of our method allow the tracker to easily adapt itself to the changes in the context and accordingly improve the tracking accuracy by resolving the ambiguities. Experimental results demonstrate that subspace-based tracking is strongly improved by exploiting the multiple cues through the proposed algorithm.

## 1. Introduction

Due to the wide applications in video surveillance, intelligent user interface, human motion understanding, content-based video retrieval, and object-based video compression [1–3], visual tracking has become one of the essential and fundamental tasks in computer vision. During the past decades, numerous and various approaches have been endeavored to improve its performance, and there is a fruitful literature in tracking algorithms development that reports promising results under various scenarios. However, visual tracking still remains a challenging problem in tracking the nonstationary appearance of objects undergoing significant pose, illumination variations, and occlusions as well as shape deformation for nonrigid objects.

When we design an object tracking system, usually two essential issues should be considered: which searching algorithm should be applied to locate the target and what type of cue should be used to represent the object. For the first issue, there are two well-known searching algorithms which had been widely studied in the last decade. These are often referred to as particle filtering and mean shift. The particle filter performs a random search guided by a stochastic motion model to obtain an estimate of the posterior distribution describing the object's configuration [4]. On the other hand, mean shift, a typical and popular variational algorithm, is a robust nonparametric method for climbing density gradients to find the peak of probability distributions [5, 6]. The searching paradigms differ in those two methods as one is stochastic and model-driven while the other is deterministic and data-driven.

Modeling target appearance in videos is a problem of feature extracting and is known to be a more critical factor than the search strategy. Developing a robust appearance system which can model the target appearance changes adaptively has been the matter of primary interest in the recent visual tracking research. The Incremental Visual Tracker (IVT) [7] has been proved to be a successful tracking method by incorporating an adaptive appearance model. In particular, the IVT modeled the target appearance as a low-dimensional subspace based on the probabilistic principal component analysis (PPCA), where the subspace is updated adaptively based on the image patches tracked in the previous frames. In this model, the intensity differences between target reference and candidates are computed to measure the observation weight. The IVT alleviates the burden of constructing a target

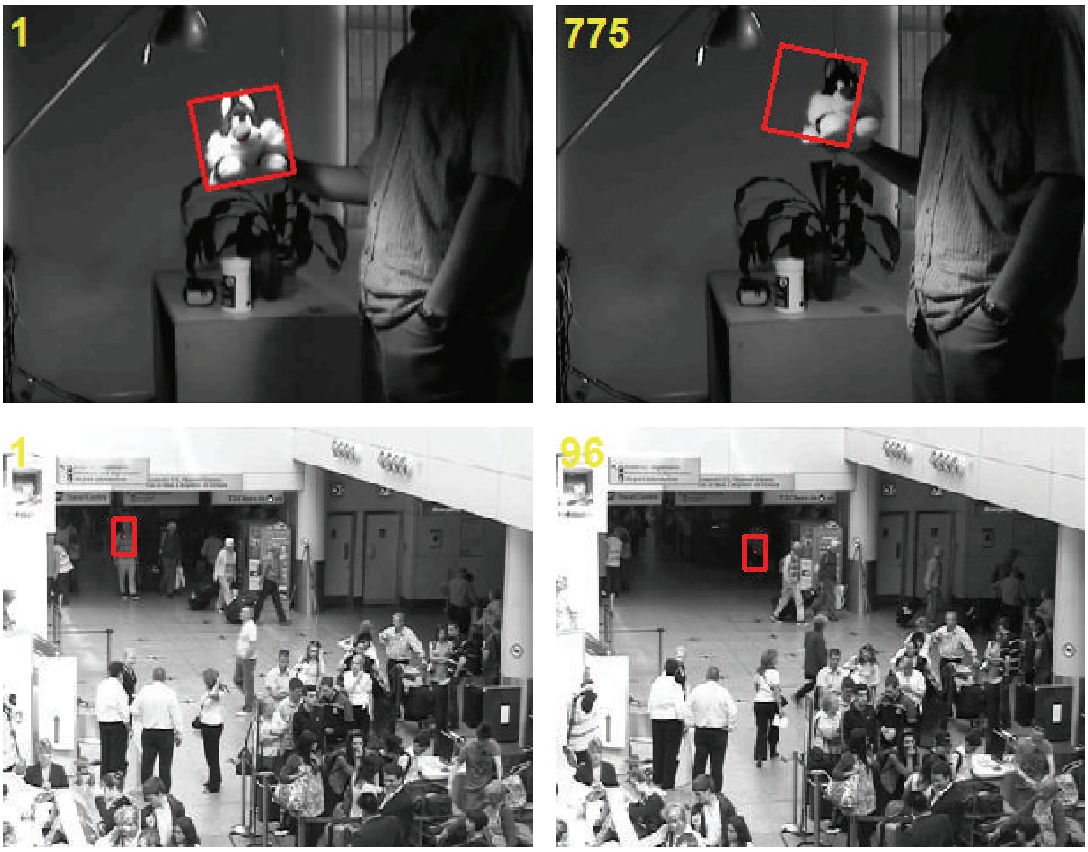


Figure 1: Two cases for the IVT tracker failure.

model prior to tracking with a large number of expensive offline data and tends to yield higher tracking accuracies. However, since only the intensity feature is employed to select the optimal candidate as the target, it may fall into trouble when the target is moving into shadow or undergoing large pose changes (as shown in Figure 1).

In this work, a multicue based incremental visual tracker (MIVT) is proposed to confront the aforementioned difficulties. In a sense, our work can be seen as an extension of [7]. Compared to the classical IVT method, the main contributions of our algorithm are as follows. First, with color (or gray) and edge properties, our representation model describes the target with more information. Second, an adaptive multicue integration framework is designed considering both the target and the background changes. When one cue becomes not discriminative enough due to target or background changes, the other will compensate. Third, the proposed multicue framework can be effectively incorporated in the particle filter tracking system, so as to make the tracking process more robustly.

The rest of the paper is organized as follows. Section 2 reviews the related multicue fusion works. Section 3 gives an overview of the IVT tracking algorithm. In Section 4, we first propose our multicue appearance modeling scheme, and then we implement the presented MIVT tracking framework. In Section 5, a number of comparative experiments are performed. Section 6 concludes the whole paper.

## 2. Related Work

There is a rich literature in visual tracking and a thorough discussion on this topic is beyond the scope of this paper. In this section, we review only the most relevant visual tracking works, focusing on algorithms that operate on multiple cues. Up to now, a number of literatures have been published about the fusion of multiple cues. In general, there are two key issues that should be solved in multicue based tracking algorithm: (1) what cues are used to represent the target's feature, (2) how the cues are integrated. Here, we focus on the second key problem.

The simplest case is that different cues are assumed to be independent, so as to use all cues in parallel and treat them as equivalent channels; this approach has been reported in [8, 9]. Based on this idea, in [10], two features, intensity gradients, and the color histogram were fused directly with fixed equal weights. A limitation of this method lies in that it does not take account of the single cue's discriminative ability.

To avoid the limitation of above methods, in [11], Du and Piater proposed Hidden Markov Models (HMM) based multicue fusing approach. In this approach the target was tracked in each cue by a particle filter, and the particle filters in different cues interacted via a message passing scheme based on the HMM, four visual cues including color, edges, motion, and contours were selectively integrated in this work. Jia et al. [12] presented a dynamic multicue tracking scheme by integrating color and local features. In this work, the weights were supervised and updated based on a Histograms of Oriented Gradients (HOG) detection response. Yang et al. [13] introduced a new adaptive way to integrate multicue in tracking multiple human driven by human detections, these defined dissimilarity function for each cue according to its discriminative power and applied regression process to adapt the integration of multiple cues. In [14], a consistent histogram-based framework was developed for the analysis of color, edge, and texture features.

In [15], Erdem et al. carried out the integration of the multiple cues in both the prediction step and the measurement step, and they defined the overall likelihood function so that the measurements from each cue contributed the overall likelihood according to its reliability. Yin et al. [16] designed an algorithm that combined CamShift with particle filter using multiple cues and an adaptive integration method was adopted to combine color information with motion information. Democratic integration was an architecture that allows the tracking of objects through the fusion of multiple adaptive cues in a self-organized fashion. This method was given by Triesch and von der Malsburg in [17]. It was explored more deeply in [18, 19]. In this framework, each cue created a 2-dimensional cue report, or saliency map; the cues fusion was carried out by resulting fused saliency map which was computed as a weighted sum of all the cue reports. Pérez et al. [20] utilized a particle filter based visual tracker that fused three cues: color, motion, and sound. In their work, color cues were served as the main visual cue, and according to the scenario under consideration, color cues were fused with either sound localization cues or motion activity cues. A partitioned sampling technique was applied to combine different cues; the particle resampling was not implemented on the whole feature space but in each single feature space separately. This technique increased the efficiency of the particle filter. However, in their case, only two cues could be used simultaneously, this restricted the flexible selection of cues and the extension of the method. Wu and Huang [21] formulated the problem of integrating multiple cues for robust tracking as the probabilistic inference problem of a factorized graphical model. To analyze this complex graphical model, a variational method was taken to approximate the Bayesian inference. Interestingly, the analysis revealed a coinference phenomenon of multiple modalities, which illustrated the interactions among different cues; that is, one cue could be inferred iteratively by the other cues. An efficient sequential Monte Carlo tracking algorithm was employed to integrate multiple visual cues, in which the coinference of different modalities was approximated.

Despite that subspace representation models have been successfully applied in handling the small appearance variations and illumination changes, they still usually fail in handling rapid appearance, shape, and scale changes. To overcome this problem for the classical IVT tracker, in this paper, we design a novel multicue based dynamic appearance model for the IVT tracking system, and this model can adapt to both the target and the background changes. We implement this model by fusing multiple cues in an adaptive observation model. In each frame, the tracking reliability is utilized to measure the weight of each cue, and an observation model is constructed with the subspace models and their corresponding weights. The appearance changes of the target are taken into account when we update the appearance models with tracking results. Therefore, online appearance modeling and weight update of each cue can adapt our tracking approach to both the target and background changes, thereby generating good performances.

## 3. Review of the IVT

The IVT models the target appearance as a low-dimensional subspace based on the probabilistic principal component analysis (PPCA) and uses particle-filter dynamics to track the target.

Let the state of the target object at time $t$ is represented as

$$X_t = \{x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t\}, \tag{1}$$

where $x_t$, $y_t$, $\theta_t$, $s_t$, $\alpha_t$, and $\varphi_t$ denote $x$, $y$ translation, rotation angle, scale, aspect ratio, and skew direction at time $t$. And the state dynamic model between time $t$ and time $t-1$ can be treated as a Gaussian distribution around the state at $t-1$; then, we have

$$p(X_t \mid X_{t-1}) = N(X_t; X_{t-1}, \Psi), \tag{2}$$

where $\Psi$ is a diagonal covariance matrix whose elements are the corresponding variances of state parameters.

Based on (1) and (2), the particle filter can be carried out to locate the target. In this stage, first the particles are drawn from the particle filter, according to the dynamical model. Then, for each particle, extract the corresponding window from the current frame and calculate its reconstruction error on the selected eigenbasis and weight $w$ through (3) and (4), which is its likelihood under the observation model:

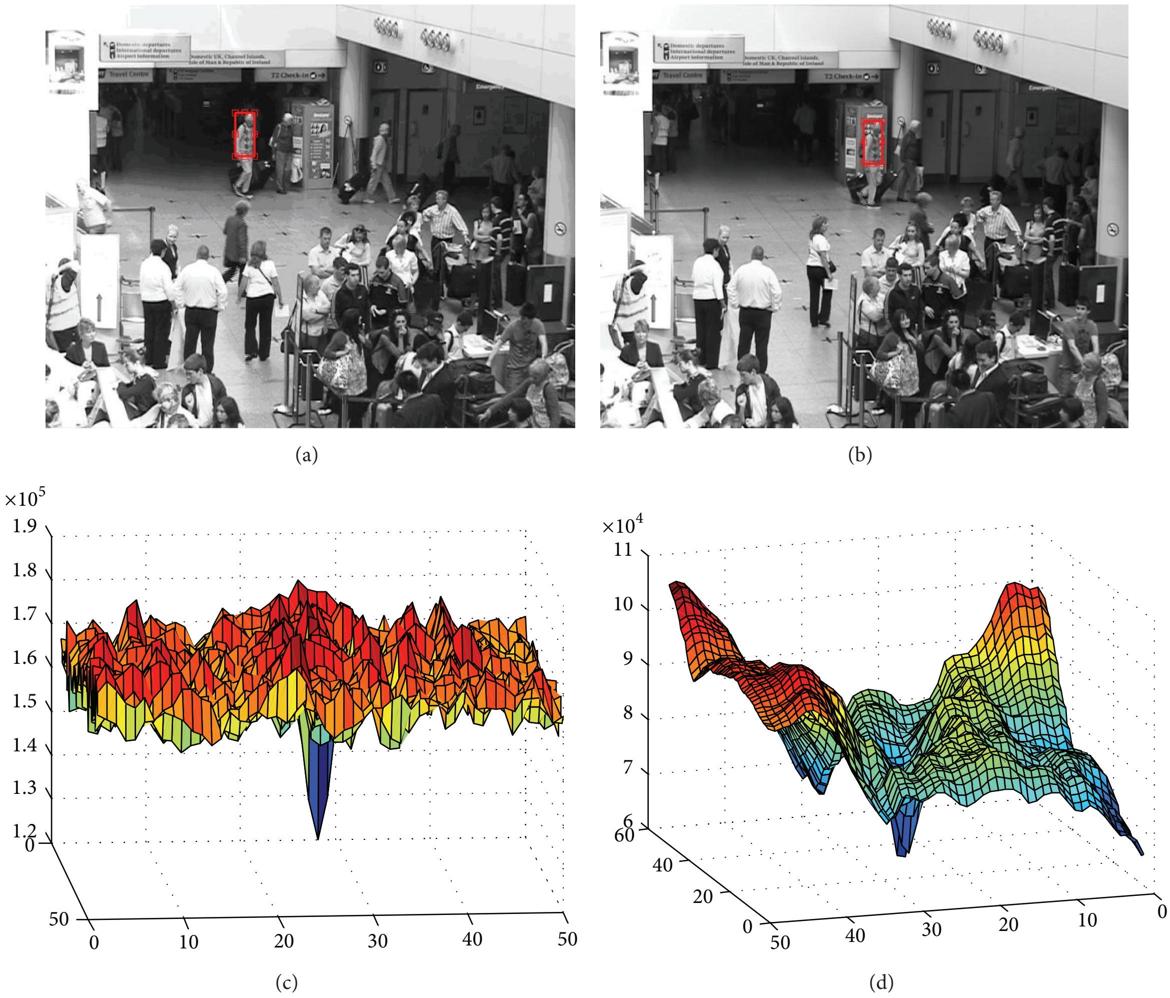
$$R_e = (I_t - \mu) - UU^T (I_t - \mu), \tag{3}$$

$$w_i^t \propto \exp\left(\frac{-\|R_e^t\|^2}{2\sigma_w^2}\right), \quad i = 1, \ldots, M, \tag{4}$$

where $I_t$ is an image patch predicated by $X_t$, and it was generated from a subspace spanned by $U$ and centered at $\mu$. In (4), $\sigma_w^2$ is the variance of the reconstruction error and $\|\cdot\|$ denotes the *L2-norm*.

Finally, the image window corresponding to the most likely particle is stored as the real target window. When the desired number of new images has been accumulated, perform an incremental update (with a forgetting factor) of the eigenbasis, mean, and effective number of observations.

The key contribution of the IVT lies in that an efficient incremental method was proposed to learn the eigenbases online as new observations arrive. This method extends the Sequential Karhunen-Loeve (SKL) algorithm to present a new incremental PCA algorithm that correctly updates the eigenbasis as well as the mean, given one or more additional training data. A detailed description of this method can be found in [7].


(a)

(b)

(c)

(d)

FIGURE 2: Different features show different discriminative abilities. (a) The target window within a red bounding box in reference image. (b) Current target window after object moving. (c) Image error between reference image and candidate image around current target window with edge cue. (d) Image error between reference image and candidate image around current target window with gray.

## 4. Multicue Fusion

For a robust tracking system with multicue, different cue's significance should be consistent with its tracking reliability. And this significance also should be self-adapting under dynamic environment when the object undergoing significant pose, illumination variations, and occlusions as well as shape deformation, so as to ensure that the most reliable cue in current time is always adapted to track the target. In this work, we aim to develop a multicue integrating framework which is flexible enough to exploit any valid image feature, such as gray, texture, edge direction, and motion information; and this framework does not restrict the target's feature type. However, it is impossible to apply all the types of features simultaneously, for simple, only two types of features, gray and edge, are used in the rest of the paper.

Figure 2 shows that difference cues may have different discriminative abilities. In Figure 2(a), the image within red rectangle is set as the target reference. After some time, the target moves to the current position of Figure 2(b). To evaluate the discriminative ability of various features, we generate target candidate uniformly around the current target position with same scale as the reference image. Then, the sum pixel error between the candidate and reference is calculated under two feature spaces: edge and gray. As shown in Figures 2(c) and 2(d), it indicates that different cues may yield different discriminative abilities. From the figure we also can find that here two distances exist: (1) the Euclidean distance between the position of candidate and the position of the real target in the image plane, (2) the distance between the reference model and candidate model in feature space (reconstruction error). When single cue is used, small feature distance means that the candidate approximates the real object more closely. However, this does not work when multiple visual cues are adopted, since different cues may have different sensitive characters to the change of object appearance and environment.

In this section, we introduce a method to evaluate the reliabilities of cues based on the above analysis, and this method can be effectively embedded in the particle filter

```
Initialization
Locate the target manually in the first frame, and use a single particle to indicate this location. Set the initial relative sharpness factors as $R_0 = 1/N$ for $N$ cues. Initialize the eigenbasis to be empty, and the mean to be the appearance of the target in the first frame.
for $t = 1$ to $T$
    (1) Spread the target states at time $t - 1$ to time $t$ using the state dynamic model.
    (2) For each new state $x_t^i$ corresponding to particle $i$ at time $t$, find its corresponding weight $w_t^{(i,f)}$ in feature space $f$ based on its likelihood under the observation models.
    (3) Based on each cue's relative sharpness factor $R_{t-1}^f$, $f = 1, \ldots, N$. Combine multiple cues by calculating the new weight for each particle as $\widetilde{w}_t^i = \sum_{f=1}^N w_t^{(i,f)} R_{t-1}^f$.
    (4) Store the image window corresponding to the most likely particle. When the desired number of new images have been accumulated, perform an incremental update (with a forgetting factor) of the eigenbasis, mean, and effective number of observations.
    (5) Update the relative sharpness factor for each cue at time $t$ as $R_t$ based on the estimated target state and the particle distribution.
end for
```

Algorithm 1: Multicue based IVT algorithm.

tracking framework. In our approach, we treat each particle as a target candidate, and the image reconstruction error of this candidate serves as particle's weight. Thereby, the particle distribution and its weights can be referred as a 3D map (Figures 2(c) and 2(d)). For a point $p(x, y, z)$ in this map, the $x$ and $y$ coordinates give the point projecting position on the image plane, and $z$ coordinate describes the weight for the particle on this point. Particles with same position on the image plane may have different weighs under various feature spaces, for that different cues may lead to different reconstruction errors. In other words, particles with same distribution may show difference maps under difference cues. For cues with enough discriminability between the target and background, obvious height differences exist among point on target position and point on other positions in the 3D map. These 3D maps are then analyzed to obtain the sharpness factor of the terrain. The sharpness factor is used to evaluate the significance of the cue.

We denote the distance created by reconstruction error as $d_m^n$, $m = 1, \ldots, M$ and $n = 1, \ldots, N$. Here, $m$ is the number of particle; $n$ is the cue index which the distance belongs to. This distance can be gotten by (3) and (4), where $d_m^n \propto w_m^n$. Then we calculate the Euclidean distance $t_m$, $m = 1, \ldots, M$ for every particle as:

$$t_m = \sqrt{\left(x_m - x_0\right)^2 + \left(y_m - y_0\right)^2}, \tag{5}$$

where $(x_m, y_m)$ and $(x_0, y_0)$ are the coordinates of particle and target in the image plane. The sharpness factor for particle $m$ under feature space $n$ can be defined as:

$$r_m^n = \frac{d_m^n}{t_m^n}. \tag{6}$$

So the mean sharpness factor for the entire particle under feature space $n$ is

$$\widehat{r}^n = \frac{1}{M} \sum_{m=1}^{M} r_m^n. \tag{7}$$

Here, $\widehat{r}^n$ gives the tracking ability for the $n$th feature space, because more large value of $\widehat{r}^n$ indicates that the current reconstruction error map is more steep, therefore the target can be distinguished from other candidates more clearly. Otherwise, the current reconstruction error map is more flat, thus the target and other candidates may be confused. To compare the discriminative ability among various feature spaces, the relative sharpness factor among different features is defined as:

$$R_n = \frac{\widehat{r}^n}{\sum_{n=1}^{N} \widehat{r}^n}. \tag{8}$$

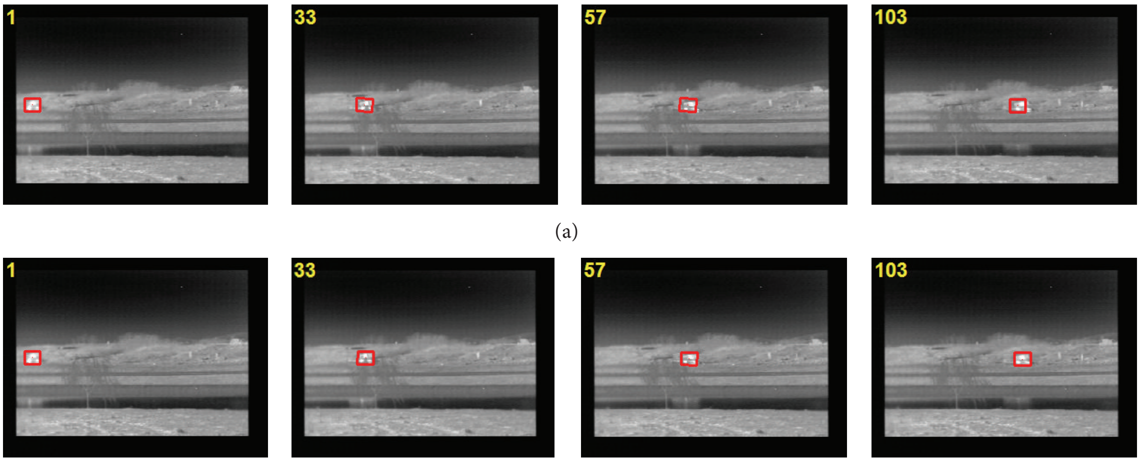
This RSF (relative sharpness factor) gives the significance for the $n$th cue.

The general algorithm is given as  in Algorithm 1.

## 5. Experimental Results and Analysis

We implemented the proposed approach in MATLAB based on the code of classical IVT from http://www.cs.toronto.edu/~dross/ivt/. The proposed method is tested on several video sequences, which include difficult tracking conditions such as complex backgrounds, occlusions, and non-rigid object's appearance changes. In order to test the effectiveness of the proposed adaptive appearance model, we compare the tracking results of our presented method with other approaches. For multicue method, the multicue appearance models with intensity and edge cues are used, as for the single cue tracker, the single feature model with intensity cue is applied. The number of particles used for our method is same to the other trackers, 300 particles are adopted for all experiments except for the two long sequences where it is 500. In all cases, the initial position of the target is selected manually.

The first test sequence is an infrared (IR) image sequence; it shows a tank moving on the ground from left to right. Some samples of the tracking results are shown in Figure 3.



(a)

(b)

Figure 3: Tracking results for seq.1. The first row: results for IVT. The second row: results for our method.

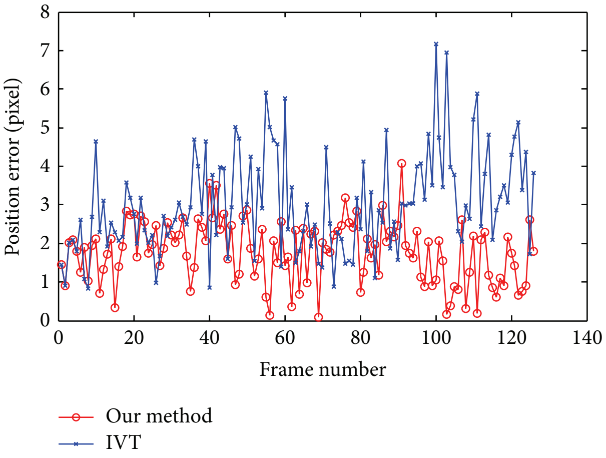


Figure 4: Position error for seq.1.

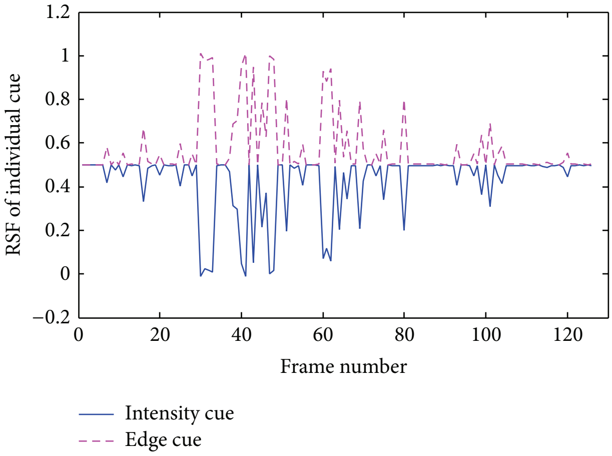


Figure 5: RSF for each cue throughout seq.1.

Here, the first row gives the results of classical IVT and the second row shows the results of our proposed method. The frame indices are 1, 33, 57, and 103 from left to right. The target-to-background contrast is very low and the noise level is high for these IR frames. In Figure 4, the tracking errors for both the two methods are given, we can see that our tracker is capable of tracking the object all the time with small error. The RSF for the two cues are demonstrated in Figure 5, it shows that the edge weight is higher than the intensity weight in general, and this also gives low target-to-background contrast.

The second test sequence shows a moving person and it presents challenging appearance changes caused by shadows of the trees. Figure 6 shows the tracking results using both methods, where the first row gives the results of classical IVT and the second row shows the results of our proposed method. The person is small in the image and undergoes sharp appearance changes when he walks into the shadow. From Figure 7, we see that large position error aroused for the classical IVT, on comparisons, our method keeps low error still the person walk out the shadow. In Figure 8, the RSF of edge and intensity cues are given.

The third test sequences is from http://www.cs.toronto.edu/~dross/ivt/, it shows a moving animal toy undergoing drastic view changes, for that the toy frequently changes its view as well as its scale. The tracking results are illustrated in Figure 9, where rows 1 and 2 correspond to IVT and rows 3 and 4 correspond to our tracker. In which eight representative frames (1, 100, 240, 427, 538, 605, 693, and 775) are shown. We can see that our proposed tracker performs well throughout the sequence. In contrast to our method, IVT fails when the target changes its pose drastically.

The fourth test sequence is an infrared (IR) image sequence from the VIVID benchmark dataset [22]. In this sequence, cars run through large shadows caused by the trees on the roadside, and the target-to-background contrast is low, but the noise level is high. Some samples of the final tracking results are demonstrated in Figure 10. Four representative

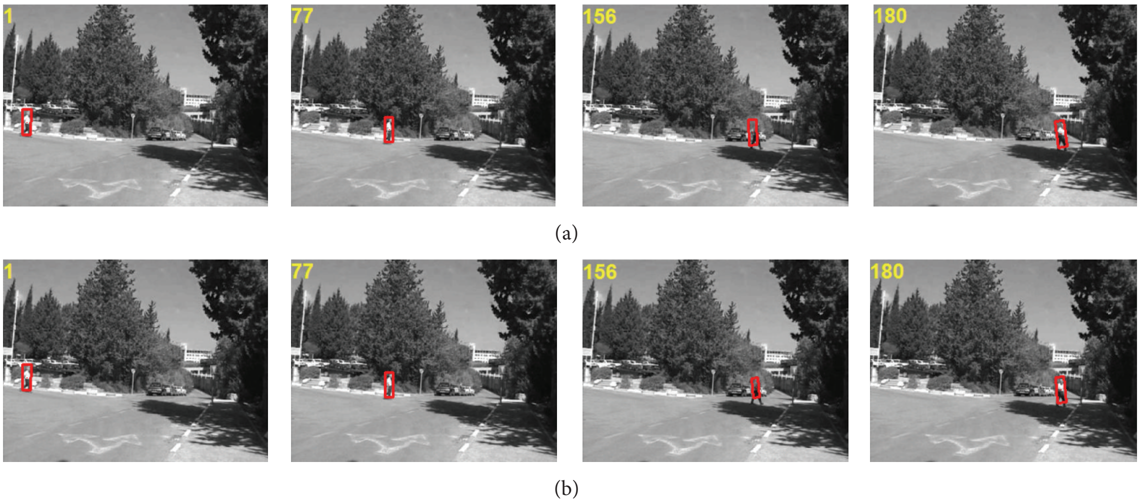


Figure 6: Tracking results for seq.2. The first row: results for IVT. The second row: results for our method.

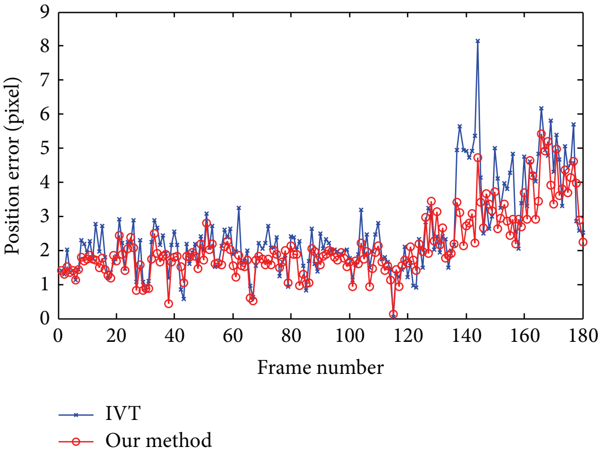


Figure 7: Position error for seq.2.

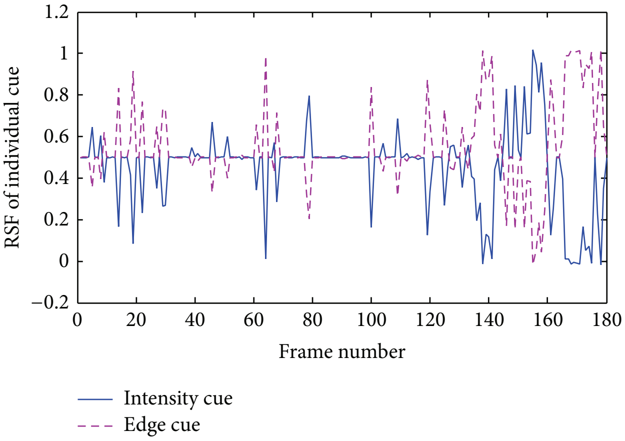


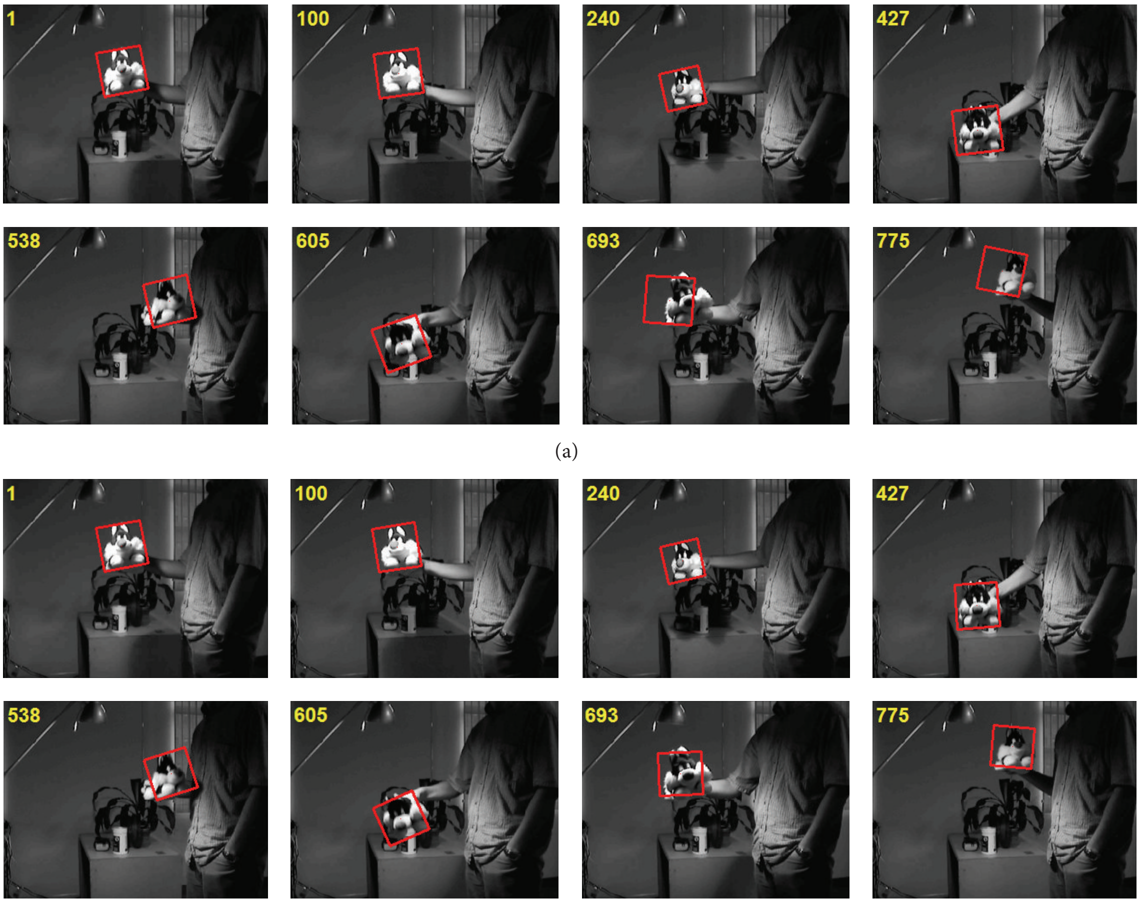
Figure 8: RSF for each cue throughout seq.2.

frames of the video sequence are shown, with indices 1, 51, 72, and 102, it corresponds to the Figure 781, 831, 852, and 882 in the dataset. From Figure 9, we see that our tracker is capable of tracking the object all the time even the car runs out the shadows. In comparison, IVT tracker fails when the car runs out the shadows and is unable to recover it.

The fifth test sequence is obtained from PETS 2007 benchmark dataset http://www.cvg.reading.ac.uk/datasets/index.html. It shows a walking passenger in subway station who undergoes large appearance changes. Some samples of the tracking results are shown in Figure 11. The frame indices are 1, 58, 96, and 113 from left to right; the indices of them in the dataset are 884, 941, 979, and 996 respectively. From the results, we can see that the tracking process of IVT cannot distinguish the actual person of interest from the background for the large appearance changes. On the other hand, our framework provide good tracking results, it can overcome the effect of appearance changes and tracking the target successfully.

Figure 12 gives some representative tracking results for three sequences which have been tested in [7]. The first row shows results for sequence “trellis70.” The indices of them in the dataset are 40, 155, 225, and 303 from left to right. The second row provides representative results for the sequence “car4,” and the frame indices are 0, 210, 419, and 638. Tracking results for the sequence “davidin300” are depicted in the last row, frame 5, 150, 300, and 460 of the dataset are given. As can be seen in the figure, our method performs well under challenging conditions such as variations of views, scale, and illumination changes. To straightforwardly make comparisons among other tracker, we quantitatively evaluate our tracking algorithm on the sequence “duedk” and sequence “car11” which can be found in [7]. In Table 1, center location RMS errors of three tracker: the proposed method, IVT tracker, and a multicue tracker which described in [15] (here, we call it CSR) are provided.

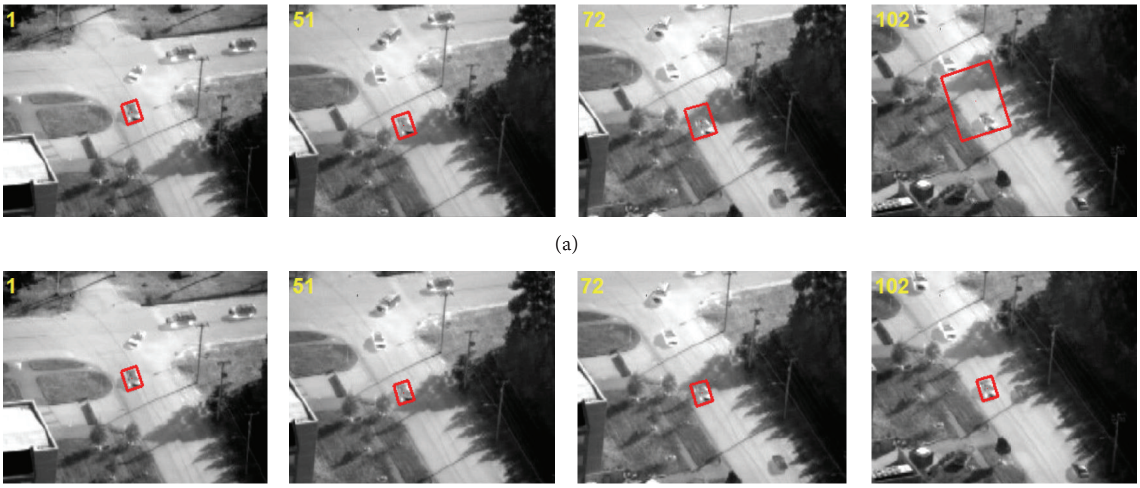
Finally, we investigate the runtime of the IVT algorithm and the proposed method. As can be seen from Table 1, the IVT tracker can track target with perfect real-time processing



(a)

(b)

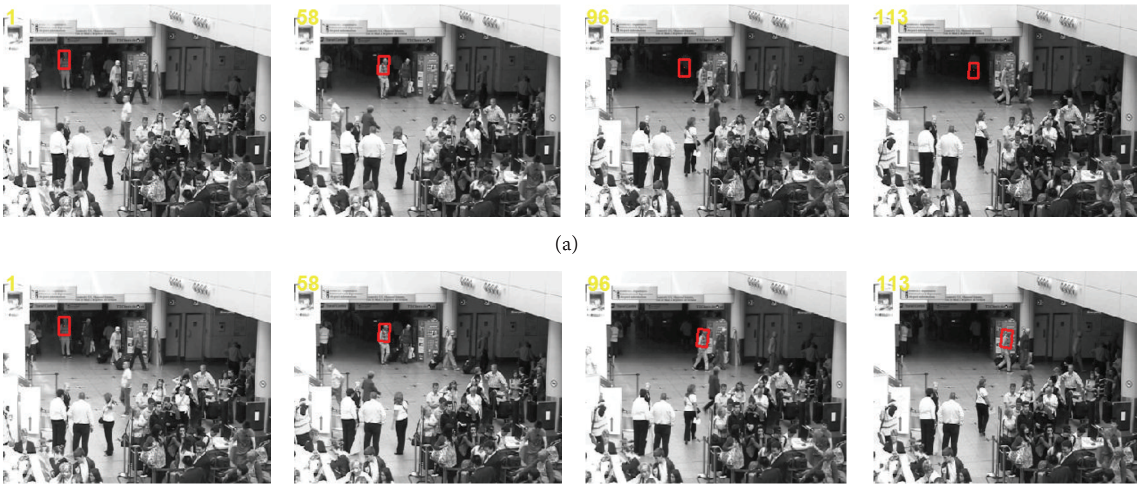
Figure 9: Tracking results for seq.3. The first two rows: results for IVT. The last two rows: results for our method.



(a)

(b)

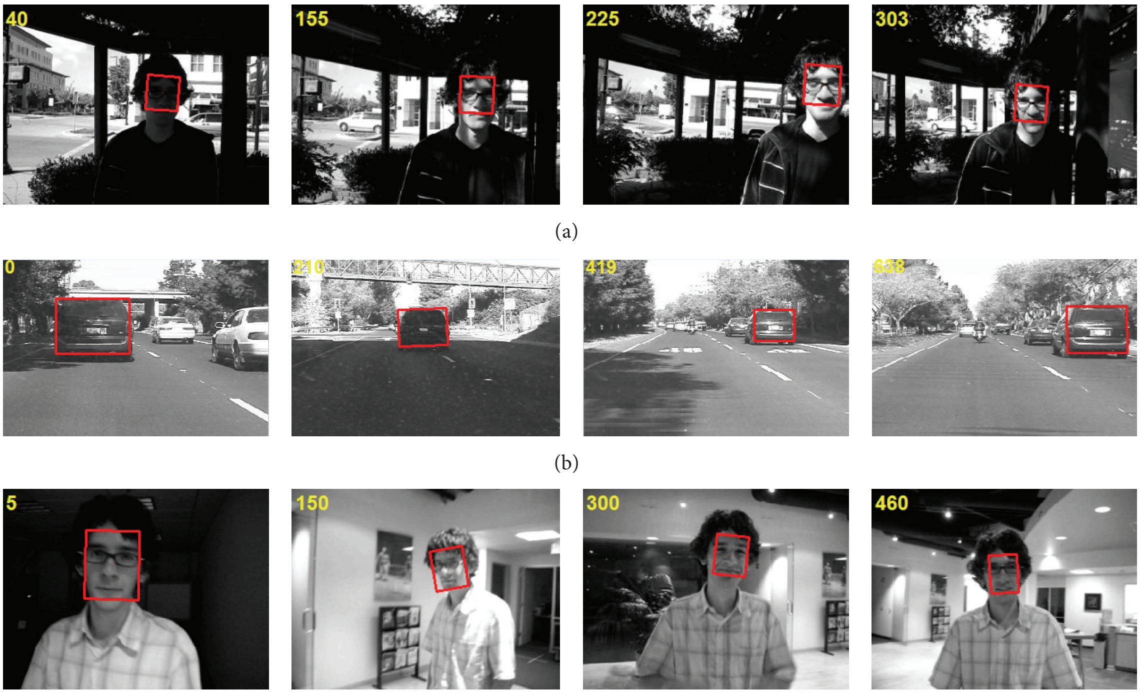
Figure 10: Tracking results for seq.4. The first row: results for IVT. The second row: results for our method.



(a)

(b)

Figure 11: Tracking results for seq.5. The first row: results for IVT. The second row: results for our method.



(a)

(b)

(c)

Figure 12: Tracking results for some sequences which had been tested in [7]. The first row: results for the sequence "trellis70." The second row: results for the sequence "car4." The last row: results for the sequence "davidin300."

Table 1: Center location RMS errors (in pixel) and running speed (in frame per second) for three trackers.

| Video sequence | IVT | | CSR | | Our method | |
|---|---|---|---|---|---|---|
| | RMS error | Running Speed | RMS error | Running Speed | RMS error | Running Speed |
| dudek | 18.19 | **21.32** | 19.21 | 3.2 | **15.32** | 13.5 |
| Car11 | 2.84 | **28.04** | 2.55 | 5.6 | **2.51** | 18.2 |

speed. In contrast, our method and CSR are slower than the IVT. With the same number of particles, the IVT algorithm using single intensity cue run at an average speed of 25.4 fps, in comparison, our method using both intensity and edge cues runs at an average speed of 16.7 fps. This means a loss in the runtime performance as increasing the number of cues.

As illustrated in above experiment results, we can see that the presented approach outperforms the IVT and the CSR algorithm in terms of tracking accuracy. This mainly stems from our adaptive cue integration scheme, for that, at each frame, the target is determined by using particles under all the cues, but additionally considering their discriminative reliabilities, rather than by just using particles under single cue which itself may provide poor or inaccurate measurements. The advantage of our formulation is its adaptive nature which lets us easily combine different target views, but generally with a loss of computational efficiency. It would be interesting to focus on developing more efficient solutions to this problem in future work.

## 6. Conclusion

In this work, we presented a novel tracker to enhance the IVT algorithm by employing a multicue based adaptive appearance model. First, we carry out the integration of cues both in feature space and image geometric space. Second, considering both the target and background changes, the integration of cues directly depend on the dynamically-changing reliabilities of visual cues. These two aspects of our method allow the tracker to easily adapt itself to the changes in the context and accordingly improve the tracking accuracy by resolving the ambiguities. In this way, our adaptive appearance model can ensure when one cue becomes not discriminative enough due to target or background changes, the other will compensate. In the last, the proposed multicue framework effective utilizes the merits of particle filter, so as to make robust tracking on less computing cost. Experimental results demonstrate that subspace tracking is strongly improved by exploiting the multiple cues through the proposed algorithm.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work is jointly supported by the National Natural Science Foundation of China (nos. 61203272, 41275027, and 61304082) and the Natural Science Foundation of Anhui Province (1308085MF82).

## References

[1] R. Wang and J. Popovic, "Real-time hand-tracking with a color glove," *ACM Transactions on Graphics*, vol. 28, no. 3, article 63, 2009.

[2] V. Thomas and A. K. Ray, "Fuzzy particle filter for video surveillance," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 937–945, 2011.

[3] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," *Image and Vision Computing*, vol. 29, no. 1, pp. 1–14, 2011.

[4] M. Isard and A. Blake, "CONDENSATION: conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[5] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[6] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–575, 2003.

[7] 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.

[8] H. Wang and D. Suter, "Efficient visual tracking by probabilistic fusion of multiple cues," in *Proceedings of the 18th International Conference on Pattern Recognition*, pp. 892–895, August 2006.

[9] P. Li and C. Francois, "Image cues fusion for object tracking based on particle filter," in *Articulated Motion and Deformable Objects*, vol. 3179 of *Lecture Notes in Computer Science*, pp. 99–110, 2004.

[10] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 232–237, June 1998.

[11] W. Du and J. Piater, "A probabilistic approach to integrating multiple cues in visual tracking," in *Proceedings of the 10th European Conference on Computer Vision*, vol. 2, pp. 225–238, 2008.

[12] G. Jia, Y. Tian, Y. Wang, T. Huang, and M. Wang, "Dynamic multi-cue tracking with detection responses association," in *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*, pp. 1171–1174, October 2010.

[13] M. Yang, F. Lv, W. Xu, and Y. Gong, "Detection driven adaptive multi-cue integration for multiple human tracking," in *Proceedings of the International Conference Computer Vision*, pp. 1550–5499, 2009.

[14] P. Brasnett, L. Mihaylova, D. Bull, and N. Canagarajah, "Sequential Monte Carlo tracking by fusing multiple cues in video sequences," *Image and Vision Computing*, vol. 25, no. 8, pp. 1217–1227, 2007.

[15] E. Erdem, S. Dubuisson, and I. Bloch, "Visual tracking by fusing multiple cues with context-sensitive reliabilities," *Pattern Recognition*, vol. 45, no. 5, pp. 1948–1959, 2012.

[16] M. Yin, J. Zhang, H. Sun, and W. Gu, "Multi-cue-based CamShift guided particle filter tracking," *Expert Systems with Applications*, vol. 38, no. 5, pp. 6313–6318, 2011.

[17] J. Triesch and C. von der Malsburg, "Democratic integration: self-organized integration of adaptive cues," *Neural Computation*, vol. 13, no. 9, pp. 2049–2074, 2001.

[18] M. Spengler and B. Schiele, "Towards robust multi-cue integration for visual tracking," *Machine Vision and Applications*, vol. 14, no. 1, pp. 50–58, 2003.

[19] C. Shen, A. van den Hengel, and A. Dick, "Probabilistic multiple cue integration for particle filter based tracking," in *Proceedings of the 7th Digital Image Computing: Techniques and Applications*, pp. 399–408, 2003.

[20] P. Pérez, J. Vermaak, and A. Blake, "Data fusion for visual tracking with particles," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 495–513, 2004.

[21] Y. Wu and T. S. Huang, "Robust visual tracking by integrating multiple cues based on co-inference learning," *International Journal of Computer Vision*, vol. 58, no. 1, pp. 55–71, 2004.

[22] VIVID database, http://vision.cse.psu.edu/data/vividEval/datasets/datasets.html.

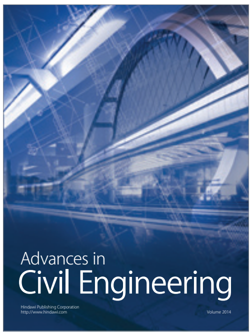
Journal of
Engineering

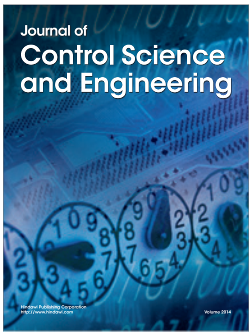
The Scientific World Journal

International Journal of
Rotating Machinery

Journal of
Sensors

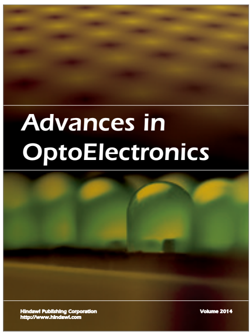
International Journal of
Distributed Sensor Networks


Advances in
Civil Engineering


Journal of
Control Science and Engineering

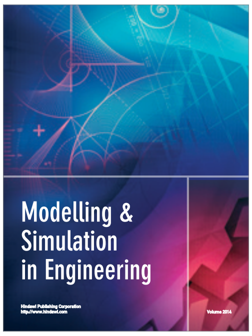
Journal of
Robotics

Journal of
Electrical and Computer Engineering

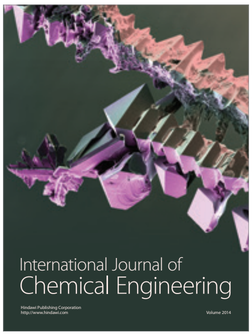

Advances in
OptoElectronics

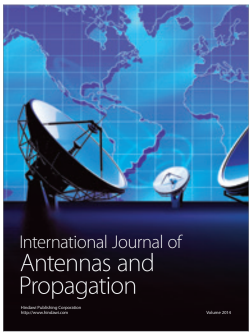
VLSI Design

International Journal of
Navigation and Observation

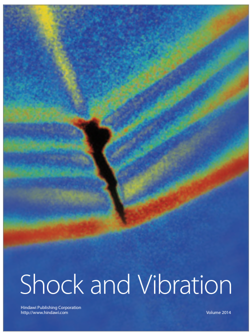

Modelling &
Simulation
in Engineering

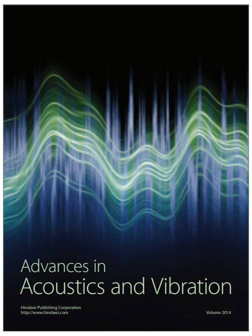
International Journal of
Aerospace Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration