Contour coding based rotating adaptive model for human detection and tracking in thermal catadioptric omnidirectional vision

Yazhe Tang and Youfu Li*

Department of Mechanical and Biomedical Engineering, City University of Hong Kong, Hong Kong, China

*Corresponding author: meyfli@cityu.edu.hk

Received 10 January 2012; revised 20 August 2012; accepted 28 August 2012;
posted 4 September 2012 (Doc. ID 161210); published 19 September 2012

In this paper, we introduce a novel surveillance system based on thermal catadioptric omnidirectional (TCO) vision. The conventional contour-based methods are difficult to be applied to the TCO sensor for detection or tracking purposes due to the distortion of TCO vision. To solve this problem, we propose a contour coding based rotating adaptive model (RAM) that can extract the contour feature from the TCO vision directly as it takes advantage of the relative angle based on the characteristics of TCO vision to change the sequence of sampling automatically. A series of experiments and quantitative analyses verify that the performance of the proposed RAM-based contour coding feature for human detection and tracking are satisfactory in TCO vision.

OCIS codes: 100.3008, 100.4999, 110.2970, 110.3080, 110.6820.

1. Introduction

In the past few decades, automatic surveillance systems have become more and more popular in a large variety of applications. However, most surveillance systems [1–4] rely on the conventional imaging system that requires proper illumination and has a limited field of view. In this paper, we propose a novel surveillance system for human detection and tracking with thermal catadioptric omnidirectional (TCO) vision, which consists of a thermal camera and a catadioptric omnidirectional sensor. A thermal imaging system is employed as it is a technology that enables detection of people and objects in darkness and in diverse weather conditions. Although its performance may be affected by some extreme weather conditions [5], such as dense fog, heavy rainfall and snow, thermal imaging cameras still allow us to see targets better than is possible with visible-light imaging systems. Therefore, a thermal camera is suitable for surveillance under diverse weather conditions. As the price of thermal cameras has reduced dramatically, they have become popular in civilian applications in recent years. To achieve an extended field of view, the catadioptric omnidirectional sensor is a good choice to capture the global information with a single image.

Compared with the conventional systems, detection and tracking in TCO vision are challenging as only limited information can be used in thermal imagery and the imaging distortion is severe in catadioptric omnidirectional vision (COV). Therefore, it is hard to apply the conventional concepts and algorithms to the TCO vision directly. To the best of our knowledge, there are very limited works that can be referenced by our system. Based on the characteristics of thermal vision, the contour can be considered as an informative and stable feature, while the other features, such as grayscale information and texture, are limited in thermal vision. Although detection and tracking in thermal imaging are difficult, more and more researchers pay increased attention to the thermal imagery vision driven by its merit. Recently, there has been extensive literature in human detection and tracking with thermal vision systems.

1559-128X/12/276641-12$15.00/0

In [6], the authors employ the support vector machine (SVM) for classification and Kalman filter to integrate with mean shift for tracking pedestrian in thermal imagery. Wang *et al.* extract the gray and edge cues, and then use the motion information to guide the fused cues for human tracking in infrared vision [7]. In [8], a two-stage template-based method combined with the Adaboosted classifier for pedestrian detection is presented. In [9], the SVM is integrated with histogram of oriented gradient (HOG) [10] to detect the pedestrian in thermal imagery. In [11], the authors adopt a generalized expectation-maximization (EM) algorithm to separate infrared images into background and foreground layers first, and incorporate with SVM for pedestrian classification. Then, they present a graph matching-based method for tracking purpose.

The omnidirectional vision has drawn lots of concerns in the computer vision community since last century because of its wide field of view, which is a unique advantage for the surveillance system. In [12], a fisheye omnidirectional tracking system is introduced, and the authors use optical flow to detect the target and employ color feature based kernel particle filter (KPF) to realize the single target tracking in omnidirectional vision. In [13], a catadioptric omnidirectional surveillance system that uses multi-background modeling and dynamic thresholding to spot the sniper in the battlefield is presented. The particle filter is proposed to incorporate with the color feature to realize tracking in COV [14,15]. Schulz *et al.* introduce a catadioptric omnidirectional pedestrian recognition system for vehicle automation, where a method of boosted cascade of wavelet-based classifiers is proposed to combine with a subsequent texture-based neural network [16]. In the human detection community, most state-of-the-art methods are based on the contour information, which is developed under the hypothesis of up-right in the conventional imaging system. Considering the inherent distortion of COV, the contour distribution of target varies as its azimuth angle changes in omnidirectional image. Therefore, the conventional contour-based methods are difficult to be applied to the omnidirectional vision for detection and tracking purposes. A common solution is to unwarp the distorted catadioptric omnidirectional image to a panoramic image or transform the coordinate of local area of omnidirectional image into a rectified image followed by using the conventional algorithms for detection and tracking purposes [17,18]. However, the computational load of this method is extensive as the interpolation is involved in unwarping and coordinate transformation (CT). Furthermore, unwarping has a risk to split the target located at the border of the panoramic image into two pieces, which likely leads to failure during the detection and tracking process. More importantly, unwarping and CT processes could cause raw information loss and generation of noises, which will degrade the performance of algorithm in TCO vision.

According to the characteristics of the proposed system, it seems that the contour information can be considered as stable in thermal vision but it is distorted in a catadioptric imaging system. A representative image is shown in Fig. 1. To solve this problem, we propose a contour coding based RAM that can take advantage of the characteristics of TCO vision to realize the rotating adaptivity of contour in TCO vision. The proposed RAM could work on the omnidirectional image directly but neither involves unwarping nor CT process. For tracking purposes, we propose to make use of the probability confidence of the classifier to calculate the observation likelihood of particle filter. Since no existing TCO database is available in public, we also collect a series of TCO datasets with diverse environment temperatures. For the merits of the proposed system, it should have a wide range of applications, such as surveillance in public, security in military bases, conservation of wild animals, and automatic drive assistance [19].

The paper is organized as follows. Section 2 presents a contour coding based RAM, which is able to integrate with the state-of-the-art contour features to form a series of contour coding based rotating adaptive features. Section 3 introduces the proposed human detection and tracking algorithm in a TCO system. Section 4 presents a TCO database first, then a series of experiments and quantitative analyses is given to verify the effectiveness of the proposed algorithm for human detection and tracking in TCO vision. Finally, we conclude the paper in Section 5.

2. Contour Coding Based Rotating Adaptive Model

Contour is an important cue that we can rely on in object detection and tracking. We can distinguish a human from other objects on the basis of its contour information. In conventional vision, the contour distribution of a human is generally following the hypothesis of up-right. With the effect of distortion in catadioptric vision, the distribution of contour is varied as the azimuthal angle of target is changed

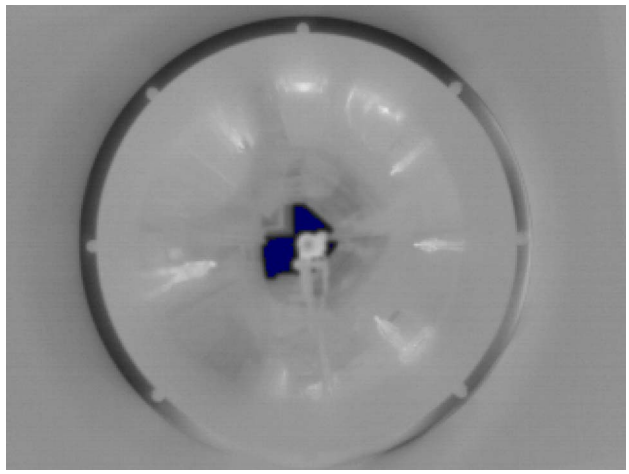


Fig. 1. (Color online) Representative image of TCO vision.

in TCO vision. Therefore, the conventional contour based methods cannot be applied to a TCO system directly for detection and tracking purposes. To solve this problem, we propose a contour coding based RAM which requires neither unwarping the distorted catadioptric omnidirectional image onto a conventional panoramic image nor making the CT for rectification, but is rotation invariant as the sampling sequence of the contour feature can change automatically based on its relative angle. According to the imaging characteristics of a TCO system, the effective imaging area is a circular image on the catadioptric mirror. Naturally, the polar coordinate is employed to fit the characteristics of TCO vision, and the center of the omnidirectional image is aligned with the origin of the polar coordinate. As shown in Fig. 2, the origin $O(0,0)$ of the image XY coordinates is located at the top left corner. Then, the center $O^l(0,0)$ of the polar coordinate X^lY^l can be obtained. As shown in Fig. 2, each point corresponds to an angle θ relative to the reference line in an omnidirectional image. With the relative angle θ, each sample model can adaptively change the initial sample point automatically. According to this principle, the proposed RAM can accommodate the rotation of a target in COV effectively. Then, the varied contour distribution can be recovered back to the reference plane through RAM. It should be noted that the proposed RAM can be integrated with the state-of-the-art contour coding features to form contour coding based rotating invariance features in TCO vision. Based on the complexity of coding mode, we combine the proposed model with the gradient information, Haar wavelet, and HOG to form three different levels of rotating adaptive contour features, respectively. With the different encoding levels, the performance of the proposed contour coding based rotating adaptive features is different.

The template of the proposed contour coding based rotating adaptive feature is shown in Fig. 3, which consists of multiple sector units to implement the contour coding. In order to extract the contour information uniformly, we employ the multiangle sectors

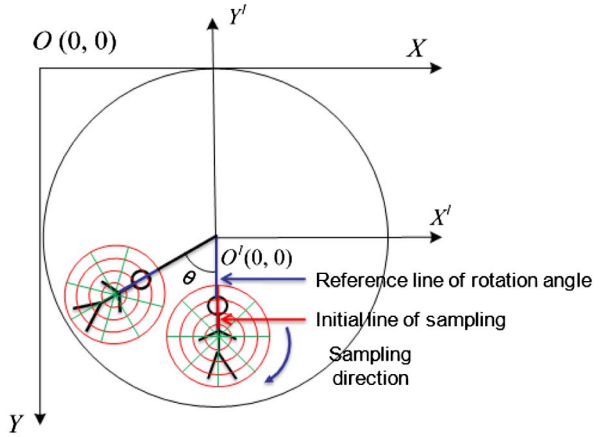


Fig. 2. (Color online) Principle of proposed contour coding based RAM.

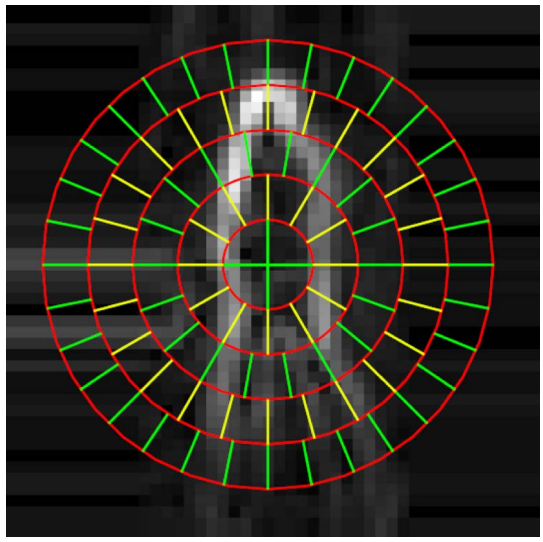


Fig. 3. (Color online) Template of contour coding based rotating adaptive feature.

as the coding unit since the single angle interval could cause oversampling at the inner ring and under sampling at the outer ring. Each ring corresponds to an angle interval φ that decreases with the radius r increases. In the following, we introduce three different types of contour coding based rotating adaptive features on the basis of the complexity of coding mode.

A. Gradient Coding Based Rotating Adaptive Feature

Gradient is an intuitive information cue we can utilize as the contour coding feature. With the sequential sampling, we could form a simple gradient coding based rotating adaptive feature. In the TCO vision, the contour distribution varies as the azimuthal angle of a target changes. To keep the contour information invariant, the sample sequence of a model should be varied accordingly. Based on the characteristics of TCO vision, the proposed RAM could take advantage of a relative angle to adaptively adjust the sampling sequence to form a rotating adaptive contour feature. The extracted gradient information $g(r,\theta)$ also contains geometric information (r,θ), combined into a single model through RAM. To effectively encode the contour information, we adopt two different configurations of radius intervals $ri_1 = 3$ with 50% overlap sampling and $ri_2 = 6$ with 25% overlap sampling to make the contour coding on the gradient map, respectively (the dimension of the template is 48×48). The small radius interval ri_1 could detect the minor change of the contour and ri_2 has a good response to larger variations. In order to reduce the effect of the area difference of multisectors in the model, we take the average of the coding unit for unit normalization. Then, the resulting feature vectors are divided by their average for normalization. Finally, the normalized feature vectors F_i $(i = 1, 2)$ are concatenated to form the final rotating adaptive contour feature F. The performance of the

proposed feature will be presented in the experiment section.

B. Haar Wavelet Based Rotating Adaptive Feature

Haar wavelet transform [20] is a very important contour coding based feature which has been well used for object detection and tracking in computer vision. It was developed based on the Haar wavelet theory, which is a sequence of rescaled "square-shaped" function [Fig. 4(a)].

The Haar wavelet's mother wavelet function $\psi(t)$ can be described as

$$\psi(t) = \begin{cases} 1 & 0 \le t < 1/2 \\ -1 & 1/2 \le t < 1 \\ 0 & \text{otherwise} \end{cases}. \tag{1}$$

Its scaling function $\phi(t)$ can be described as

$$\phi(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

Technically, the Haar wavelet is not continuous but jumping. Therefore, it is well used for analysis of the signal with sudden transition. Papageorgiou and Poggio [20] present the two-dimensional (2D) Haar wavelet transform for object detection in a conventional imaging system. In this paper, we propose to develop a Haar wavelet transform based rotating adaptive feature for omnidirectional vision by means of RAM. Conventionally, the rectangle Haar wavelet feature is adopted to meet the requirement of a traditional imaging system. To realize the rotating adaptivity of contour in TCO vision, we propose to employ the sector unit to extract the Haar wavelet transform based on the characteristics of omnidirectional vision. The proposed RAM-based Haar wavelets are shown in Fig. 4(b). It is obvious that the RAM-based Haar wavelet is different from the conventional as the area of the sector units is varied as the radial length is changed, such as radial and diagonal components [Fig. 4(b)]. To reduce the effect of difference of coding unit, we average the coding units to the same level for unit normalization. We adopt two sector configurations of radius interval $ri_1 = 4$ with 50% overlap sampling and $ri_2 = 8$ with 25% overlap sampling to encode the RAM-based Haar wavelet transform in TCO vision, respectively. Each Haar wavelet is divided into three components: angle, radial, and diagonal. For component normalization, the obtained component vectors are divided by their average. After normalization, the elements of vector much larger than 1 indicate strong intensity difference, the elements much less than 1 indicate consistent uniform region, and elements close to 1 are random pattern. Finally, the normalized component vectors are concatenated to form the final RAM-based Haar wavelet transform. Its performance will be presented in the experiment.

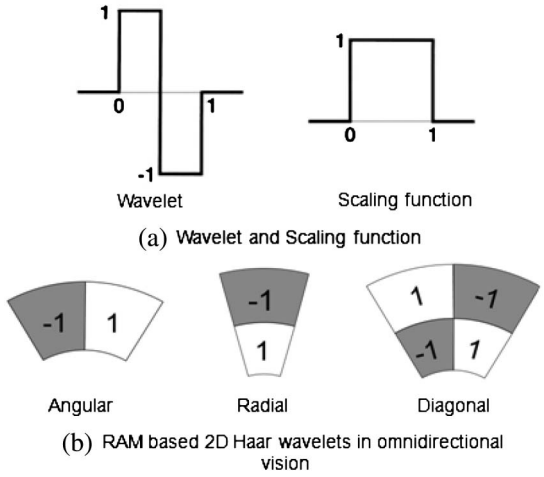


Fig. 4. Haar wavelet framework in omnidirectional vision. (a) Haar scaling function and wavelet and (b) three types of RAM-based 2D Haar wavelets in omnidirectional vision.

C. HOG-Based Rotating Adaptive Feature

Contour is a very useful feature that contains a lot of information, such as gradient and orientation. Previously, we introduced a simple RAM-based gradient coding feature that only encodes the gradient information but ignores the orientation of contour. Then, we presented the RAM-based Haar wavelet, which effectively encodes the grayscale intensity to reflect the local intensity variance of omnidirectional image. In this part, we present a RAM-based HOG [10] to adaptively represent the distorted contour information, which makes good use of the gradient intensity $g(x,y)$ and the orientation $\theta(x,y)$ [Eqs. (3) and (4)]. Due to the good performance of HOG, it has been widely used in conventional vision community. However, it is hard to simply integrate RAM with HOG to realize adaptive representation of contour in TCO vision since the distribution of contour and the orientation of gradient vary as the azimuthal angle of target changes. According to the characteristics of TCO vision, we develop an orientation transformation algorithm to effectively handle the varied orientation of gradient, which is able to effectively transform the varied orientation of gradient into the reference plane (Fig. 5). Then, we can make use of RAM to extract the recovered histogram of orientated gradient feature sequentially to obtain the HOG-based rotating adaptive feature:

$$g(x,y) = \sqrt{(I(x,y) - I(x-1,y))^2 + (I(x,y) - I(x,y-1))^2}, \tag{3}$$

$$\theta(x,y) = \tan^{-1}((I(x,y) - I(x,y-1))/(I(x,y) - I(x-1,y))). \tag{4}$$

The working principle of orientation transformation algorithm can refer to Fig. 5, and the right part

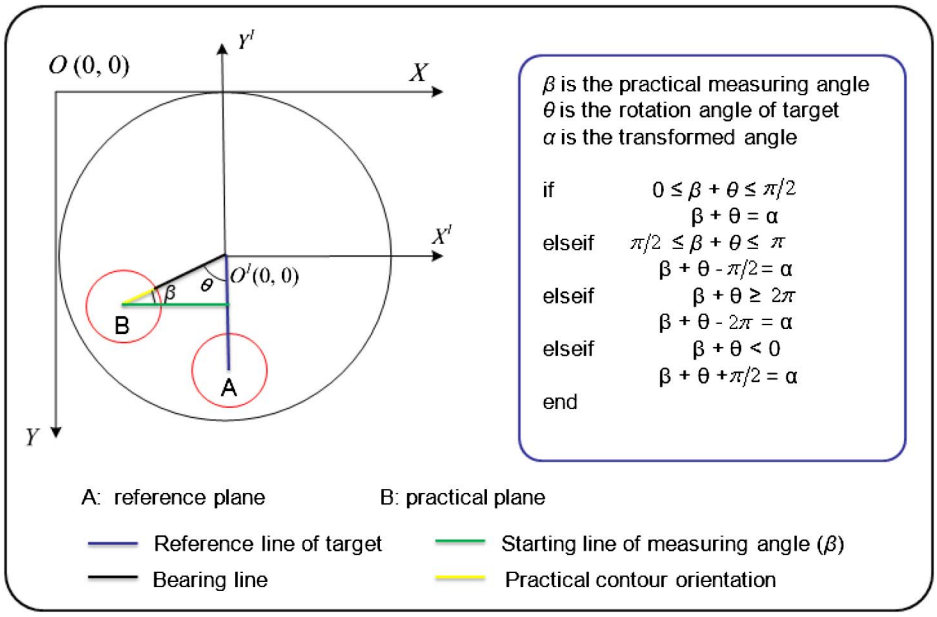


Fig. 5. (Color online) Working principle of transformation algorithm of gradient orientation in RAM.

of the figure details the transform items of the algorithm based on its azimuthal angle in omnidirectional image. As the effect of distortion in catadioptric vision, the gradient orientation of target varies with its azimuthal angle changes. In order to adapt the view habit of a human, we set A as the reference plane that lets the image of human stand vertically (Fig. 5). We set three key variables: the practical measuring angle β, the rotation angle of target θ, and the transformed angle α. The rotation angle is relative to the reference line and the measuring angle refers to the positive direction of the horizontal axis. Through the orientation transform algorithm, we can realize the transformation of gradient orientation from the practical plane to the reference plane. To better explain the principle of transform algorithm, we set the orientation $\pi/2$ as an example. When the target locates at the bottom left of the coordinate system, both rotation angle θ and practical angle β are less than $\pi/2$ but greater than 0. Therefore, it follows the item 1 of transform algorithm, and the transformed orientation α equals the sum of θ and β. Since the sum of the interior angle of triangle equals π, the transformed angle α is always equal to $\pi/2$. If the target falls on the upper left of the coordinate system, the rotation angle θ is greater than $\pi/2$ but less than π, the measuring angle β is negative but greater than $-\pi/2$. Therefore, the sum of rotation angle and measuring angle follows the item 1 of transform algorithm. In the same way, we can normalize the distorted orientation based on the transformation algorithm when the target locates at the top right and bottom right of the coordinate system. According to the proposed orientation transform algorithm, the varied orientation of gradient can be transformed into the reference plane to form the unified HOG descriptor in TCO vision. To normalize the coding unit, Dalal *et al.* [10] applied a Gaussian spatial window to weight the pixel based on its distance to the center of unit. Due to the varied area of coding units in RAM, we take the average of the orientation bins of the coding unit for unit normalization. Then, we adopt L2-norm to normalize the transformed feature vector for block normalization [10]. Finally, the transformed RAM-HOG is formed, which is robust to the rotation of contour. Experiment shows that the model configuration of $ri = 3.5$ with the 50% overlap sampling has the best performance in the TCO vision.

According to above, the RAM-based rotating adaptive features have been proposed, which should be able to well handle the distorted contour in TCO vision. Finally, their performance will be verified in the experiment section.

3. Detection and Tracking Algorithm

We introduced the contour coding based rotating adaptive features previously. For detection and tracking purposes, we adopt the well-know SVM as classifier and integrate the probability confidence of the classifier with the particle filter for tracking in TCO vision.

A. SVM Detector

SVM [21–24] is a very useful algorithm for pattern classification. It can discriminate the unlabeled samples correctly based on a trained model. Suppose the training sample set is provided as $\{x_i, y_i\}_{i=1}^{m}$, where y_i values in the set $\{-1, +1\}$, and m is the number of the training samples. The SVM works by maximizing the margin between two classes in feature space for minimization of the following objective function [22]:

$$\min J(\mathbf{w}, b, \xi_i) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i. \tag{5}$$

Subject to the constraints:

$$y_i[\mathbf{w}^T\varphi(\mathbf{x}_i)+b]\geq 1-\xi_i,\qquad \xi\geq 0,$$

where ξ_i is the slack variable, C is the penalty factor, $\varphi(\cdot)$ is the mapping from input space to feature space. By taking the Lagrangian of Eq. (5), the original problem can be derived as

$$\max\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1,j=1}^{l}\alpha_i\alpha_j y_i y_j k(\mathbf{x}_i,\mathbf{x}),\tag{6}$$

where $k(\mathbf{x}_i,\mathbf{x})=\varphi(\mathbf{x}_i)\cdot\varphi(\mathbf{x})$ is a kernel function. The function (6) follows the constraints of $\sum_{i=1}^{l}\alpha_i y_i=0$, where $\alpha_i\geq 0$ are the Lagrange coefficients which can be used to calculate the hyperplane of the SVM. Based on the formed hyperplane, the SVM could realize the sample classification for detection in computer vision.

B. Particle Filter

The particle filter [25–27] has been widely used for its advantage of nonlinear/non-Gaussian. It can approximately fit the continuous posterior distribution through a series of weighted particles $\{x_k^i,w_k^i\}_{i=1,2,\cdots,N}$. In Bayesian framework, the particle filter recursively obtains the state x_k at time k, given the available observations $z_{1:k}=z_1,z_2,\cdots,z_k$ up to time k. Suppose posterior $p(x_{k-1}|z_{1:k-1})$ at time $k-1$ is available, the posterior $p(x_k|z_{1:k})$ can be obtained recursively by prediction and update. The prediction stage makes use of the probabilistic system transition model $p(x_k|x_{k-1})$ to predict the posterior probability of state at time instant k. When the observation z_k is available, state posterior can be updated through the observation model $p(z_k|x_k)$.

The observation model characterizes the observation likelihood of particle filter. It is an important component to measure the probability confidence of the observed data for state updating. In this paper, we utilize the probability confidence q of the detector to calculate the observation likelihood of the particle filter for effective tracking in TCO vision. We define a parameter d to express the distance between candidate sample and standard positive sample [Eq. (7)]. Then, the observation likelihood $p(z_k|x_k)$ and the probability confidence of classifier are connected successfully as Eq. (8), where λ is variance:

$$d=1-q,\tag{7}$$

$$p(z_k|x_k^i)\propto\exp(-\lambda\cdot d^2),\tag{8}$$

$$w_k^i\propto w_{k-1}^i p(z_k|x_k^i).\tag{9}$$

The obtained observation likelihood $p(z_k|x_k)$ is used to affect the weight w_i of particles in Eq. (9). With the above relationship, it can be concluded that the distance d is inversely proportional to the weight of particles. Therefore, we can realize human tracking in TCO vision based on the integration of the probability confidence of the classifier and the particle filter.

4. Experiment

In this section, we present a series of experiments and quantitative analysis to verify the effectiveness of the proposed algorithm in TCO vision. For testing purposes, we establish a TCO database that covers a variety of environmental temperatures. In the following, we give a description of the TCO data collection and the experiments on the proposed algorithm.

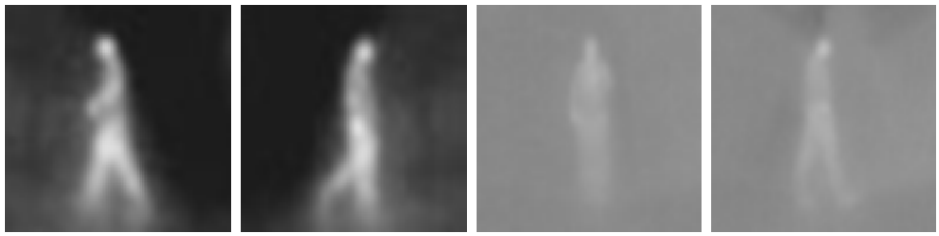
A. TCO Database

Due to the lack of publicly available TCO databases, we have to manually collect the positive (foreground) samples and TCO image sequences. The negative samples are obtained automatically from a set of TCO images not containing human. For data collection, we build a TCO sensor that consists of an FLIR Therma CAM PM 695 camera and a hyperboloid catadioptric omnidirectional mirror, as shown in Fig. 6. Some extracted representative samples are shown in Fig. 7. All TCO image sequences are of 320×240 grayscale images, and the extracted foreground/background samples are scaled to 48×48 in dimensions. The established TCO database contains several image sequences with different ambient conditions. Each set of image sequences contains hundreds of TCO frames that are sampled with 20 Hz by an FLIR thermal camera. In the following, we test the proposed rotating adaptive features on the established TCO database to verify their performance in TCO vision.

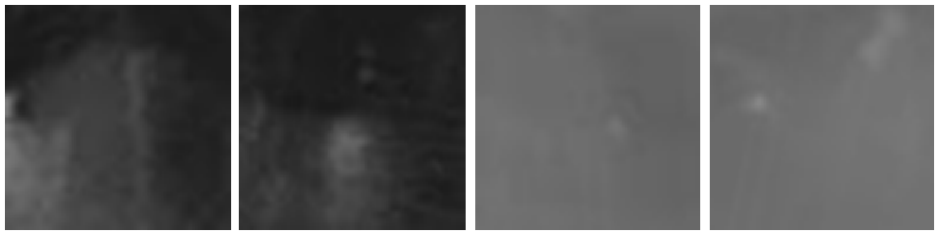
B. Detection

To verify the performance of RAM-based contour coding features for human detection in TCO vision, we present a comparison experiment with the CT-based contour coding features. First of all, it should be

Fig. 6. (Color online) Configuration of TCO surveillance system.



(a) positive samples



(b) negative samples

Fig. 7. Representatives of extracted TCO samples.

noted that the performance of features is highly dependent on the complexity of the coding mode. Since HOG makes good use of the gradient intensity and the gradient orientation, it should be able to robustly represent the contour of object. As shown in Fig. 8, HOG achieves the best performance among all the features. In addition, Haar wavelet outperforms the gradient due to its advanced coding mode on grayscale intensity level. Second, the sample extraction method (RAM or CT) affects the performance of features as well. For simplicity, the RAM-based and CT-based contour coding feature is abbreviated as RAM-feature and CT-feature, respectively. As Fig. 8 shows, RAM-features perform better than that of their corresponding CT-features because the transformed coordinate is hard to fall on the integer pixel precisely through the CT process. Therefore, it is necessary to calculate the intensity of the transformed point approximately through interpolation. However, this process is time consuming. Additionally, it is very easy to cause the information loss and

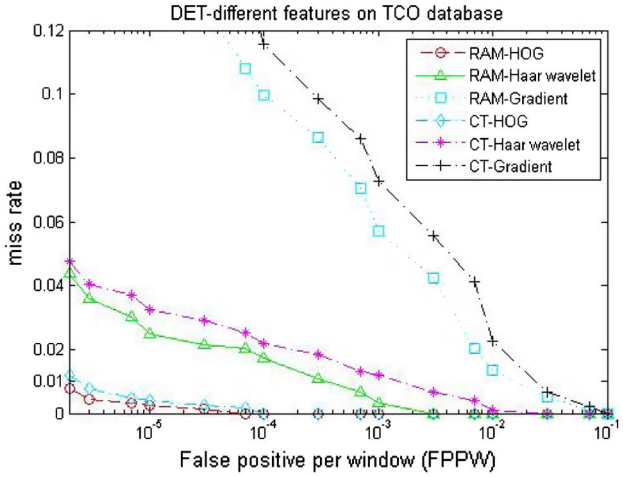


Fig. 8. (Color online) Performance of the RAM/CT-gradient, RAM/CT-Haar wavelet, and RAM/CT-HOG.

generation of noises during the CT process even though interpolation is involved. Hence, the CT process may lead to the performance degradation of the algorithm. On the contrary, RAM works directly on the original TCO image, which could maximally maintain the integrity of information. Therefore, the performance of RAM-features is superior to their corresponding CT-features in the TCO vision. It is worth mentioning that the CT-HOG adopted the conventional HOG descriptor, which applied a Gaussian spatial window to weight the pixel in the coding unit. Due to the varied area of coding units of RAM-HOG, we take the average of the orientation votes of the coding unit for unit normalization. As a matter of fact, the Gaussian weighting only improves the performance of HOG by 1% [10], so it is not enough to make up the information loss caused by CT. As shown in Fig. 8, it is proved that the proposed RAM-HOG has a better performance than that of CT-HOG in the TCO vision. Some representative detection examples of RAM-feature are given in Fig. 9. To test the robustness of the RAM-feature, we set some interferences in the experiment. In the dark circumstance, some small animals appear in the test scene. As expected, the RAM-features are still able to accurately distinguish the human from the interferences. Therefore, it can be concluded that the RAM-features perform robustly in TCO vision.

To further analyze the performance of RAM-feature, we make a comparison with the method proposed in [11], which presented a simple but effective algorithm for human detection in the conventional thermal vision. As shown in Table. 1, we select eight sets of representative image sequences from the TCO database for comparison. In this experiment, RAM-HOG and RAM-HW outperform the method [11] totally, and RAM-G performs better than the method [11] in average. The method in [11] has a good sensitivity in some particular image sets (set 2, 3, 4, and 8) that are collected in the night time with a high

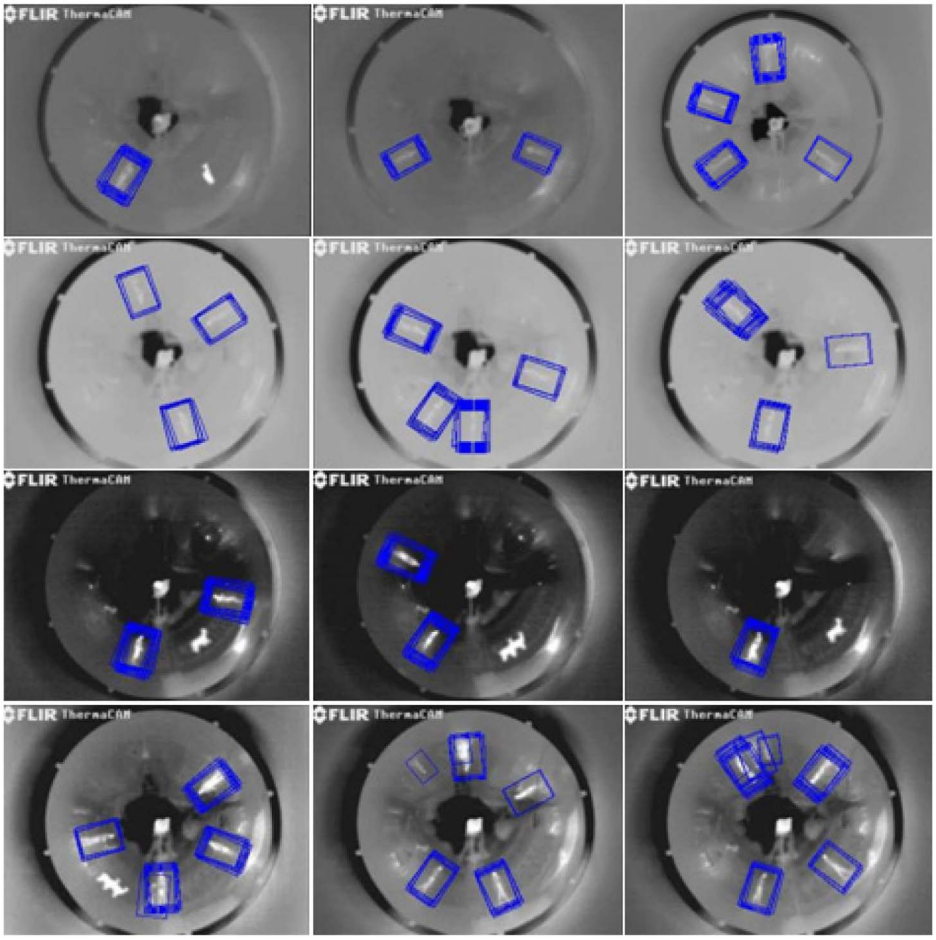


Fig. 9. (Color online) Representatives of the human detection on TCO vision database with RAM-feature (RAM-HOG).

contrast ratio. In contrast, the RAM-features are robust through all the experiments. In addition, the RAM-features have low false positives (FPs) as they contain a wealth of information to represent the characteristics of the target. It is notable that RAM-HOG achieves the best performance with the highest sensitivity 0.99 and the lowest FP among all the features. Through the above quantitative analysis and comparative experiments, the effectiveness of the proposed RAM-feature is verified for human detection in TCO vision.

C. Polarity Switch

Polarity switch is a particular phenomenon in thermal vision, when it occurs, hot and cold ranges of thermal sensors get reversed. Although this phenomenon is different from conventional thermal imaging obviously, the outline of the target is still visible in the thermal image. Therefore, the contour information is still stable in polarity switch. The experiment verifies that the proposed RAM-feature is robust to the phenomenon of polarity switch, as shown in Fig. 10.

Table 1. Detection Results of RAM-Gradient (RAM-G), Haar Wavelet (RAM-HW), and HOG (RMA-HOG) in TCO Vision Database (TP, True Positive; FP, False Positive; Sensitivity, Hit Rate)

			TP				FP				Sensitivity			
Dataset	Frames	People	RAM-G	RAM-HW	RAM-HOG	[11]	RAM-G	RAM-HW	RAM-HOG	[11]	RAM-G	RAM-HW	RAM-HOG	[11]
1	81	85	79	85	85	68	11	3	0	18	0.93	1	1	0.8
2	78	84	61	84	84	76	24	3	0	27	0.73	1	1	0.90
3	71	134	128	134	134	119	9	2	0	34	0.96	1	1	0.89
4	43	81	73	81	81	76	8	1	0	19	0.90	1	1	0.94
5	58	111	101	108	110	101	4	3	1	16	0.91	0.97	0.99	0.91
6	37	136	130	134	135	94	11	1	0	31	0.96	0.99	0.99	0.69
7	37	182	161	173	181	149	10	4	2	24	0.88	0.95	0.99	0.82
8	26	112	101	106	111	102	5	1	0	14	0.90	0.95	0.99	0.91
Total	430	925	834	905	921	785	82	18	3	183	0.90	0.98	0.99	0.86

D. Tracking

As discussed above, the performance of RAM-features for human detection in TCO sensors has been verified. In this section, we intend to further discuss the effectiveness of the proposed RAM-features on human tracking in TCO vision.

According to the characteristics of a thermal imaging system, a thermal image reflects the distribution of environment temperature field. With the difference of temperature field, the image of a human is salient against the background in thermal vision. Also, grayscale intensity of pixel corresponds to the value of temperature directly. Therefore, grayscale intensity is an intuitive feature that can be adopted in thermal vision. In this section, we compare the performance of the proposed RAM-feature with CT-feature and grayscale feature for tracking in TCO vision. Through tests on a number of datasets, the effectiveness of the RAM-feature for tracking in TCO vision is verified. For simplicity, we abbreviate the compared trackers as follows:

(1) "RAM-Gradient-PF, RAM-Haar-wavelet-PF, and RAM-HOG-PF" are RAM-Gradient, RAM-Haar wavelet, and RAM-Hog based particle filters, respectively.
(2) "CT-Gradient-PF, CT-Haar-wavelet-PF, and CT-HOG-PF" are CT-Gradient, CT-Haar wavelet, and CT-HOG based particle filters, respectively.
(3) "G-PF" is the grayscale feature based particle filter.

Both (1) and (2) are contour coding feature based particle filters, which utilize the probability confidence of the classifier to calculate the observation likelihood of particle filter. The observation likelihood of G-PF is determined by the Bhattacharyya distance. For a fair comparison, the above trackers are implemented with some identical parameter settings, such as particles number N and particle distribution variance λ. The dynamic models of trackers are random walking model as $x_k = x_{k-1} + v_k$, where v_k is a zero-mean Gaussian random variable. For the tracking experiments, we test the trackers on the established TCO database with diverse environment temperatures. Since the environment parameters vary in practice, the contrast ratio of thermal image becomes different. For example, thermal imaging system may achieve a higher contrast ratio in the night as the relatively low temperature during night. On the contrary, the contrast ratio may be relatively lower at the daytime as the environment temperature increases, which results in a blurry contour of human target. This is because the difference of temperature between background and foreground is reduced in the day. Therefore, the trackers may easily capture the target in the night but meet more challenges due to the vague outline in the day.

As described previously, it is difficult to apply the conventional contour based method on TCO vision unless the distorted omnidirectional image is rectified through the CT process. However, it is possibly hard to achieve a satisfactory tracking accuracy in TCO vision. The reason is that the original image is difficult to be approximated by the transformed image closely even though interpolation is employed. Therefore, working on the original omnidirectional image is able to achieve more important information and avoid noises generation during the CT process. To verify the tracking performance of the RAM-feature based trackers, we conduct a series of quantitative analyses in the following. The average root mean square errors (RMSE) with different particle numbers are presented to measure the stability of proposed RAM-features. As shown in Fig. 11(a), the RAM-Gradient-PF performs better than the CT-Gradient-PF in general, but their difference is small on account of the original encoding method of gradient feature. As Figs. 11(b) and 11(c) show, the RAM-Haar-wavelet-PF and RAM-HOG-PF achieve much higher tracking accuracy comparing to their corresponding CT-Haar-wavelet-PF, and CT-HOG-PF. The RMSE of RAM-Haar-wavelet-PF is around 3 and its accuracy is higher than that of the corresponding CT-Haar-wavelet-PF by 1.3 times. The RMSE of RAM-HOG-PF is around 2 and its tracking accuracy is almost 2 times higher than that of CT-HOG-PF. Furthermore, it is obvious that the RAM-features perform better than the grayscale feature, as shown in Fig. 11(d). Among them, the G-PF has the worst performance (RMSE is around 6.3). It can be concluded that the contour feature is more stable than grayscale feature in TCO vision. In addition,

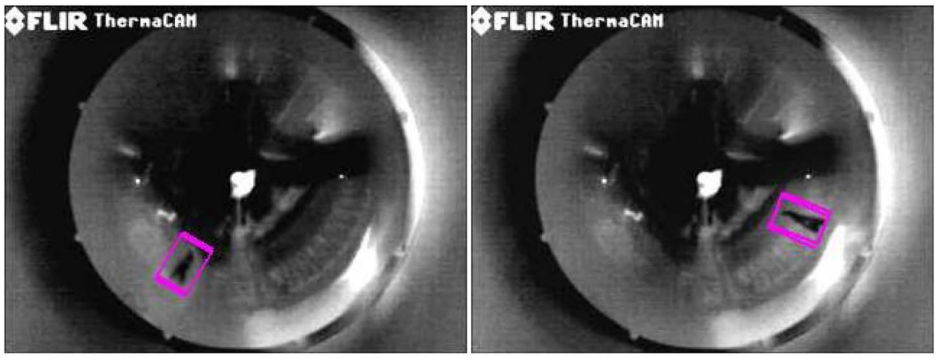


Fig. 10. (Color online) Experiment for human detection in TCO vision with polarity switch.

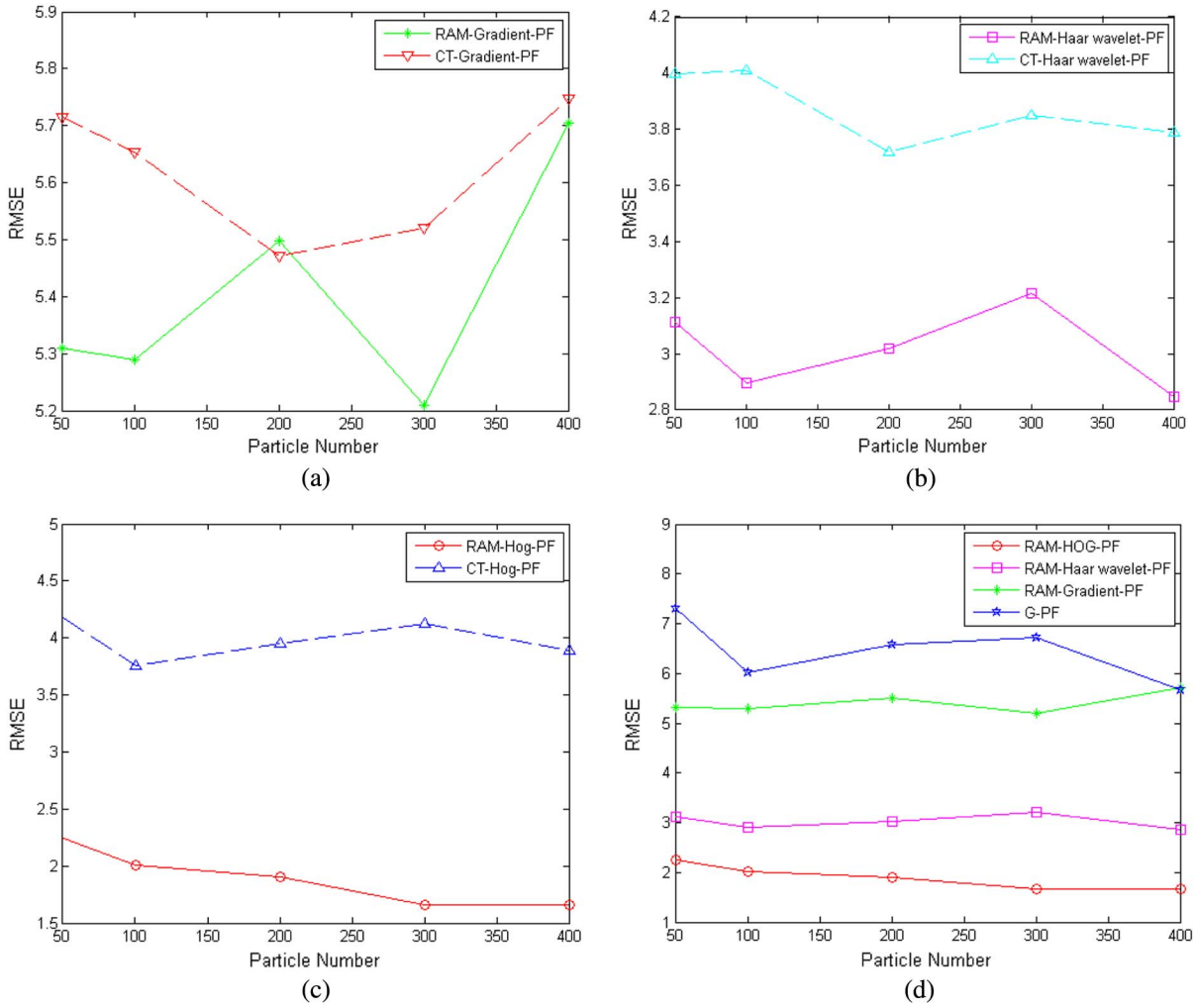


Fig. 11. (Color online) RMSE of RAM/CT-feature-PFs with the different number of particles. (a) RAM-Gradient-PF and CT-Gradient-PF. (b) RAM-Haar-wavelet-PF and CT-Haar-wavelet-PF. (c) RAM-HOG-PF and CT-HOG-PF. (d) RAM-Gradient-PF, RAM-Haar-wavelet-PF, RAM-HOG-PF, and G-PF.

it should be noted that the RAM-HOG-PF has the most stable performance among all of the trackers, and it achieves the best accuracy (RMSE = 1.66) using 300 particles. Therefore, it can be concluded that the RAM-feature based trackers have a satisfied stability in focusing on the center of human, while the grayscale based tracker is not stable to apply in TCO vision. To further discuss the effectiveness of RAM-feature based trackers, we present the following experiments and analyze the results in detail.

To verify the performance of the proposed RAM-feature, we test the algorithms on the designed experiments with different environment temperatures. In experiment I, two persons walk in the vicinity of TCO sensor at the forenoon. Due to the moderate environment temperature in the forenoon, we obtain a TCO image sequence with medium contrast ratio. At the beginning of the experiment, all the trackers track the human target successfully. Several frames later, G-PF tends to drift away from the target on account of it suffers from the interference of background, and its performance degrades seriously under the interference of a heat source point. Finally, the G-PF loses one target at the Frame 199 (Fig. 12, experiment I). In contrast, the RAM-feature based trackers track both targets successfully until the end of experiment even they also are affected by some interferences occasionally. The reason is that the RAM-feature based trackers adopt the contour feature, which is robust to the discrete interference of grayscale intensity compared to the grayscale feature, and it only responds to the correct contour distribution. Therefore, the contour coding based tracker has a great immunity from the numerous background interferences in the TCO vision. In experiment II, three persons appear in the scene at noon. As shown in Fig. 12 (experiment II), an image sequence with a low contrast ratio is obtained since the environment temperature reaches the peak at noon. It is obvious that the G-PF loses one of three targets very quickly due to the small intensity difference between foreground and background. Although there is a weak contour response of target in experiment II, the RAM-feature based trackers still can

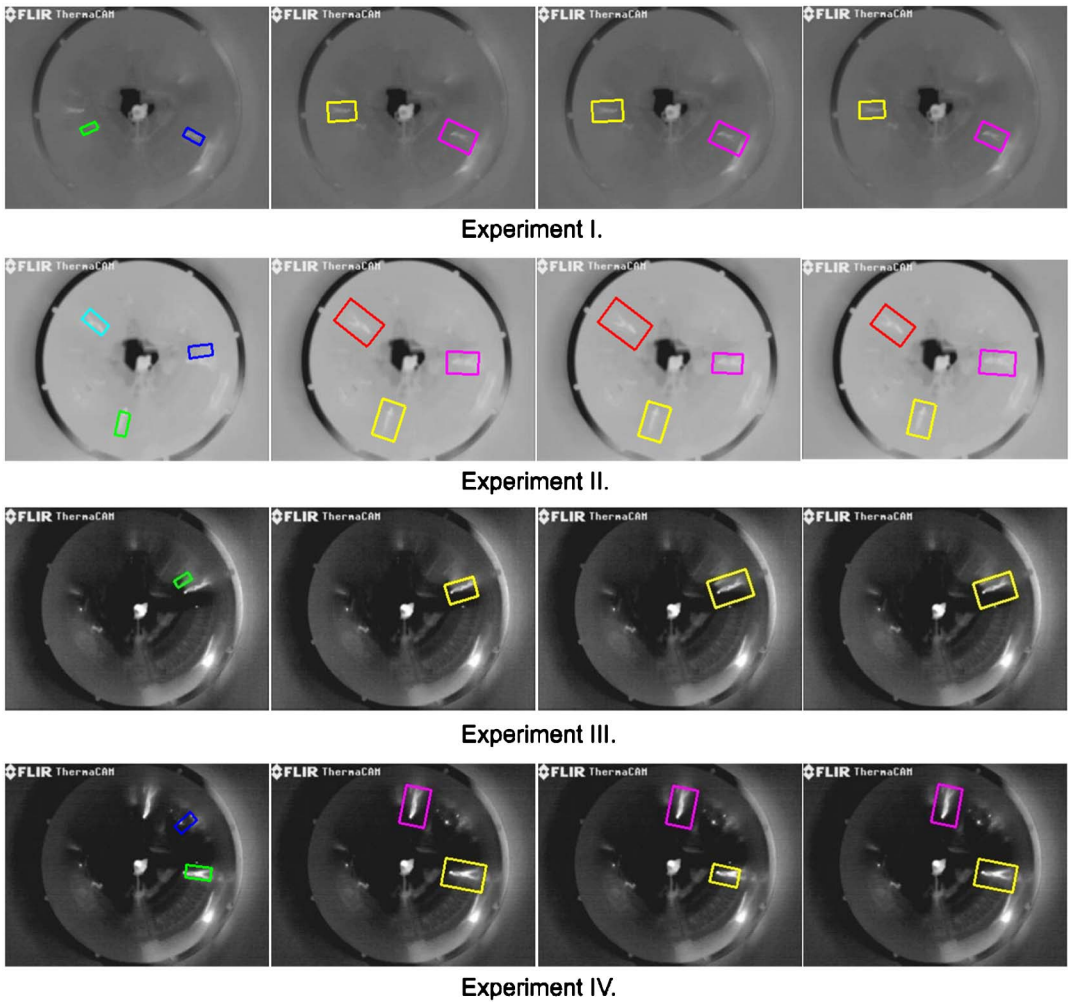


Fig. 12. (Color online) Experiments for human tracking in TCO vision. From left to right they are G-PF, RAM-G-PF, RAM-Haar-wavelet-PF, and RAM-HOG-PF. Experiment I: forenoon, experiment II: noon, experiment III, IV: night.

track the target stably and successfully survive until the end of the experiment. In experiments III and IV, the image sequences are collected at the different time period of night (experiment IV is collected later than experiment III). Since the environment temperature of night is lower than that in the day, both experiments III and IV have the optimal contrast ratio among all of the experiments. Therefore, the performance of the trackers should be improved obviously. However, the G-PF fails again as previously at the heat source points in experiments III and IV. From experiments III and IV we can conclude that there is very limited single channel grayscale information can be utilized in thermal vision, which directly results in a poor immunity from interference of G-PF even in the nighttime with the high contrast ratio. Therefore, grayscale information is unstable compared to the contour feature in TCO vision, especially under the settings with low number of particles. On the contrary, RAM-feature based trackers are able to track the target successfully through all the experiments. Furthermore, it should be noted that the RAM-HOG-PF achieves the best performance among all of the RAM-feature-PFs in the experiments. According to the above analysis, we can conclude that the RAM-feature-PFs have a stable performance compared to G-PF in the TCO vision.

Through the above experiments, we have analyzed the performance of RAM-features on human detection and tracking in the TCO system. Compared with the CT-feature, the proposed RAM-feature achieves much better performance in TCO vision. In addition, it is proved that contour feature is more stable than the grayscale feature for tracking in TCO vision. Through a variety of experiments, we can conclude that the effectiveness of the proposed RAM-feature has been verified for human detection and tracking in TCO vision.

As stated previously, CT-feature is time consuming as interpolation is involved. In contrast, the proposed RAM works on the original TCO image directly, which greatly improves the efficiency of the algorithm. With HOG descriptor as an example (because HOG achieves the best performance and its coding method is most complex), RAM-HOG achieves 70.36 s/frames for exhaustive detection (1650 times/frames) at a scale of 48×48 without optimization, and it is implemented in MATLAB on a PC of an Intel

Pentium 2.7 GHz with 2 G memory. In contrast, CT-HOG consumes 148.98 s/frames and 101.17 s/frames at the scale of 48 × 48 and 48 × 24, respectively. Therefore, RAM-feature is efficient and it should have a great potential for real-time application in surveillance if it is implemented in C/C++ and takes advantage of GPU processing.

5. Conclusion

In this paper, we introduced a novel TCO surveillance system that is able to work in darkness and has a wide field of view. Due to the inherent distortion of omnidirectional vision, most of the conventional contour features are difficult to be applied to the omnidirectional image directly unless the distorted omnidirectional image is rectified in advance. However, the rectified image may involve information loss and noise generation even when interpolation is implemented. To maximally maintain the integrity of information in TCO vision, a RAM is developed which could realize contour coding on the original TCO image directly. The effectiveness of the proposed RAM-feature is verified through a large number of experiments on human detection and tracking in TCO vision.

Since the proposed RAM-feature is developed based on the entire human model, its performance could be degraded when occlusion occurs. In our future work, we propose to develop a part based RAM-feature to handle occlusion. Then, the proposed algorithm should be robust to more challenging situations for human detection and tracking in TCO vision.

This work was supported by the Research Grants Council of Hong Kong (Project No. CityU 118311) and City University of Hong Kong (Project No. 7008176).

References

1. P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Proceedings of IEEE International Conference on Computer Vision* (IEEE, 2003), pp. 734–741.
2. J. Burchett, M. Shankar, A. Hamza, B. Guenther, N. Pitsianis, and D. Brady, "Lightweight biometric detection system for human classification using pyroelectric infrared detectors," Appl. Opt. **45**, 3031–3037 (2006).
3. R. Liu, E. Liu, J. Yang, Y. Zeng, F. Wang, and Y. Cao, "Automatically detect and track infrared small targets with kernel Fukunaga–Koontz transform and Kalman prediction," Appl. Opt. **46**, 7780–7791 (2007).
4. A. Abdi, M. Schmiedekamp, and S. Phoha, "Probabilistic color matching and tracking of human subjects," Appl. Opt. **49**, 4926–4935 (2010).
5. F. Xu, X. Liu, and K. Fujimura, "Pedestrian detection and tracking with night vision," IEEE Trans. Intell. Transport. Syst. **6**, 63–71 (2005).
6. M. Vollmer and K. P. Mollmann, "Selected applications in other fields," in *Infrared Thermal Imaging: Fundamentals, Research and Applications* (Wiley-Vch, 2010), pp. 566–579.
7. X. Wang, L. Liu, and Z. Tang, "Infrared human tracking with improved mean shift algorithm based on multicue fusion," Appl. Opt. **48**, 4201–4212 (2009).
8. J. Davis and M. Keck, "A two-stage template approach to person detection in thermal imagery," in *Proceedings of IEEE Workshop on Applications of Computer Vision* (IEEE, 2005), pp. 364–369.
9. F. Suard, A. Rakotomamonjy, and A. Bensrhair, "Pedestrian detection using infrared images and histograms of oriented gradients," in *Proceedings of IEEE Intelligent Vehicles Symposium* (IEEE, 2006), pp. 206–212.
10. N. Dalal and B. Triggs, "Histogram of oriented gradients for human detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2005), pp. 886–893.
11. C. Dai, Y. Zheng, and X. Li, "Layered representation for pedestrian detection and tracking in infrared imagery," Comput. Vis. Image Und. **106**, 288–299 (2007).
12. S. Yang, G. Min, and C. Zhang, "Tracking unknown moving targets on omnidirectional vision," Vis. Res. **49**, 362–367 (2009).
13. T. Boult, X. Gao, R. Micheals, and M. Eckmann, "Omnidirectional visual surveillance," Image Vis. Comput. **22**, 515–534 (2004).
14. J. Bazin, K. Yoon, I. Kweon, C. Demonceaux, and P. Vasseur, "Particle filter approach adapted to catadioptric images for target tracking application," in *Proceedings of British Machine Vision Conference* (Academic, 2009), pp. 1–15.
15. O. Jaime and B. Eduardo, "Omnidirectional vision tracking with particle filter," in *Proceedings of International Conference on Pattern Recognition* (IEEE, 2006) pp. 1115–1118.
16. W. Schulz, M. Enzwiler, and T. Ehlgen, "Pedestrian recognition from a moving catadioptric camera," in *Proceedings of the 29th DAGM conference on Pattern Recognition* (Academic, 2007), pp. 456–465.
17. A. Barczak, J. Okamoto, Jr., and V. Grassi, Jr., "Face tracking using a hyperbolic catadioptric omnidirectional system," Res. Lett. Inf. Math. Sci. **13**, 55–67 (2009).
18. Y. Tang, Y. Li, T. Bai, X. Zhou, and Z. Li, "Human tracking in thermal catadioptric omnidirectional vision," in *Proceedings of IEEE International Conference on Information and Automation* (IEEE, 2011), pp. 97–102.
19. W. Ye, H. Liu, F. Sun, and M. Gao, "Vehicle tracking based on co-learning particle filter," in *Proceedings of IEEE International Conference on Intelligent Robots and System* (IEEE, 2009), pp. 2979–2984.
20. C. Papageorgiou and T. Poggio, "A trainable system for object detection," Int. J. Comput. Vis. **38**, 15–33 (2000).
21. V. N. Vapnik, *The Nature of Statistical Learning Theory* (Academic, 1995).
22. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," Data Min. Knowl. Disc. **2**, 121–167 (1998).
23. V. N. Vapnik, *Statistical Learning Theory* (Academic, 1998).
24. J. Platt, "Probabilities for SV machines," in *Advances in Large Margin Classifiers*, A. J. Smola, P. Bartlett, B. Scholkopf, and D. Schuurmans, eds. (Academic, 2000), pp. 61–74.
25. M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," IEEE Trans. Signal Process. **50**, 174–188 (2002).
26. A. Doucet, J. de Freitas, and N. J. Gordon, "An introduction to sequential Monte Carlo methods," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, J. F. G. de Freitas, and N. J. Gordon, eds. (Academic, 2001), pp. 3–14.
27. M. Isard and A. Blake, "Condensation-conditional desity propagation for visual tracking," Int. J. Comput. Vis. **29**, 5–28 (1998).