On the Performance of Handoff and Tracking in a Camera Network

Yiming Li, Bir Bhanu and Vincent Nguyen
Center for Research in Intelligent Systems, University of California, Riverside, California, USA
*yimli@ee.ucr.edu, bhanu@*cris.ucr.edu*, vnguy023@student.ucr.edu*

Abstract

Camera handoff is an important problem when using multiple cameras to follow a number of objects in a video network. However, almost all the handoff techniques rely on a robust tracker. State-of-the-art techniques used to evaluate the performance of camera handoff use either annotated videos or simulated data, and the handoff performance is evaluated in conjunction with a tracker. This does not allow a deeper understanding into the performance of a tracker and a handoff technique separately in the real-world settings. In this paper, we evaluate three camera handoff techniques, two different color-based trackers in seven real-life cases, with varying numbers of cameras, number of objects and the changing environmental conditions. We also perform experiments on annotated videos to provide the ground-truth for all the scenarios. This evaluation of performance isolates the effect of tracking and handoff techniques and clarifies their role in a video network.

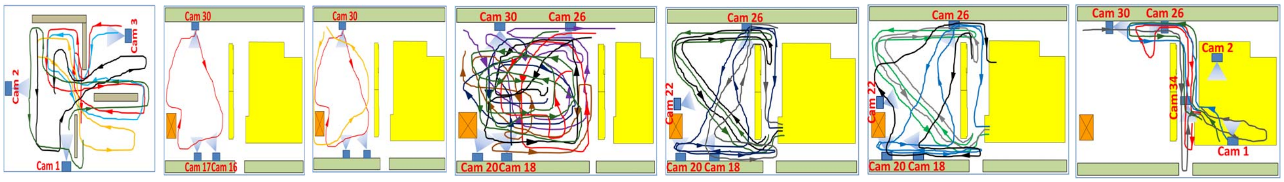
1. Introduction

Due to the increasing demand for video surveillance, there has been a large body of work in the area of tracking multiple persons with multiple cameras [1-5]. As the number of cameras involved increases, camera handoff, the process of transferring the tracking task from one camera to another, becomes very important for video networks. With camera handoff, a camera network can take the advantage of multiple cameras and get rid of the limitations caused by the limited field-of-view (FOV) of a single camera. A few camera handoff techniques are developed based on different principles [4-7]. To analyze the performance of these camera handoff techniques, there are some approaches that are evaluated on annotated videos, where the trackers are manually labeled by humans, to get rid of the effect of the tracking performance. Others evaluate the camera handoff performance based on simulated data. However, in real applications, we can neither use annotated videos nor simulated data, which makes it important to evaluate the camera handoff performance in conjunction with the tracking performance. Some approaches have done evaluation in this way [5, 6]. However, it makes it hard to deduce the real performance of a handoff technique. To our knowledge, up to now, there has been no work that (a) evaluates the camera handoff performance on different trackers and (b) separates the tracking performance from the handoff performance. In this paper, we perform a wide variety of experiments so as to have a better evaluation of the camera handoff techniques. We compare three camera handoff approaches based on two different trackers. We provide results in all the cases and compare with the ground-truth to make the effect of a tracker easier to observe.

2. Camera handoff techniques

There are many papers on the *camera assignment problem* [4-7], i.e. how to assign different cameras to perform different tasks, such as using different cameras to follow different objects in a video network. From all these works, there are some that specifically focus on the camera handoff problem [5-7], i.e. when to transfer the right of following an object from one camera to another. In this paper, we select three of these camera handoff techniques to compare their performances on different trackers. They are, namely, the utility-based approach [5], the weakly acyclic game (WAG) approach [6] and the co-occurrence occurrence ratio (COR) approach [7]. We choose these three approaches to compare because the principles behind them cover different aspects of the camera handoff problem. Also, all of them have no requirement for camera calibration, making it easy to compare them with one another. The utility-based approach and the WAG approach apply ideas from game theory in economics. They are very flexible to different scenarios because they can use the user-supplied criteria and focus on the best available camera selection, while the COR approach considers handoff when the objects are on the boundary of FOVs. The utility-based approach and the WAG approach are different mainly in the sense of computational

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.889



C1: ($N_C = 3, N_P = 5$) C2: ($N_C = 3, N_P = 1$) C3: ($N_C = 3, N_P = 2$) C4: ($N_C = 4, N_P = 6$) C5: ($N_C = 4, N_P = 4$) C6: ($N_C = 4, N_P = 4$) C7: ($N_C = 4, N_P = 4$)

Figure 1: Illustration for trajectories of persons in all cases. Cx: ($N_C = y, N_P = z$) means that in case x there are y cameras and z persons. Each color stands for the trajectory of the person in clothes of that color.

complexity, so that we can compare them as the number of cameras and objects increases. In the following, we only describe the principle of these three approaches. For detailed technical description, the interested reader is referred to [5-7].

2.1. The utility-based approach

This approach [5] views the camera selection and handoff problem in a game theoretic manner. The authors map the camera handoff problem to a classical vehicle-target problem in game theory by viewing the cameras that can "see" an object as the multiple players in a game. In game theory, a game refers to the interactions among multiple agents, i.e. players, and the welfare that a player can get is called the utility.

In this formulation, camera utility $U_{C_j}(a)$, person utility $U_{P_i}(a)$, and the global utility $U_g(a)$ are designed to make it a *potential game*:

$$U_g(a) = \sum_{C_j \in C} U_{C_j}(a) \qquad (1)$$

$$U_{P_i}(a) = U_g(a_i, a_{-i}) - U_g(C_0, a_{-i}) \qquad (2)$$

$$U_{C_j}(a) = \sum_{i=1}^{n_P} \sum_{l=1}^{N_{Crt}} Crt_{il} \qquad (3)$$

where $a = (a_i, a_{-i})$. a is the camera assignment result. a_i stands for the camera used to track person P_i, while a_{-i} stands for the camera assignment for all the other persons other than P_i. n_P is the number of person that camera C_j can see. Crt_{sl} are the user-supplied criteria. The final assignment result is given in the form of a mixed strategy $p_i^l(k) = \frac{e^{\frac{1}{\tau}\bar{U}^l_{P_i}(k)}}{e^{\frac{1}{\tau}\bar{U}^1_{P_i}(k)} + \dots + e^{\frac{1}{\tau}\bar{U}^{n_C}_{P_i}(k)}}$ where

$$\bar{U}^l_{P_i}(k+1) = \begin{cases} \bar{U}^l_{P_i}(k) + \dfrac{1}{p_i^l(k)}(U_{P_i}(a(k)) - \bar{U}^l_{P_i}(k), a_i(k) = A_i^l \\ \bar{U}^l_{P_i}(k) & ,otherwise \end{cases}$$

is the predicted person utility in the $(k+1)^{th}$ iteration step. This approach applies a potential game model. It requires $U_{P_i}(a)$ to be aligned with $U_g(a)$.

2.2. The WAG approach

In this approach [6], the authors model the camera handoff problem as a weakly acyclic game. Unlike the above utility-based approach, this model does not require the alignment of local and global utility. So, it is more flexible to use this approach. For the convenience of comparison, we will use the same criteria for these two approaches in the experimental part. The main idea of this approach is to find the better reply path iteratively. A better reply path means a set of updating of the camera actions which can make the global utility increasing. The algorithm is briefly given below as Algorithm 1.

Algorithm 1: WAG for camera assignment and handoff
1. For the camera which can "see" the person P_i, initialize the action of camera j, $a_j(0)$, randomly, and calculate the utility $U_{C_j}(a)$. $a_j(0)$ can take value from $A_j = \{Use, Standby\}$. Use this action as the camera's baseline action $a_j^b(0)$. Initialize the global baseline utility as $U_g^b(0) = U_g(0)$.
3. At each iteration k, each camera updates its action $a_j(k)$ with probability ε, or stay at its baseline action $a_j^b(k-1)$ with probability $(1-\varepsilon)$.
4. If $U_g(k) > U_g(k-1) + \delta$ $a_j^b(k+1) = a_j(k); U_g^b(k+1) = U_g(k)$ If $U_g(k) \le U_g(k-1) + \delta$ $a_j^b(k+1) = a_j^b(k); U_g^b(k+1) = U_g^b(k)$ where U_{C_j} and U_g are computed as (3) and (1) respectively and δ is the improvement step.
5. Repeat Steps 4 to 5 until there's no better reply path.
6. Perform the camera assignments and handoffs according to the set of camera action assignment $a = \{a_1, a_2, \dots, a_{N_C}\}$.   Repeat Steps 1 to 6 for every time instant.

2.3 The COR approach

This approach [7] decides camera handoff by calculating the COR. If the COR is higher than some predefined threshold, then the two points are considered to be in correspondence. When one point is getting close to the edge of the FOV of one camera, the system will hand-off to another camera that has its corresponding point. The COR is defined as $R(x, x') = \frac{p(x,,x')}{p(x)}$ where $p(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_t} K_2(x - x_t^i)$ is the mean probability that a moving object appears at *x*, i.e. the occurrence at *x*. K_2 is claimed to be circular Gaussian kernel. $p(x, x') = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_t} K_2(x - x_t^i)\sum_{i=1}^{N'_t} K_2(x' - x'^i_t)$ is the co-occurrence at *x* in one camera and *x'* in another. If two points *x* and *x'* are in correspondence, then the calculated COR will be 1. On the contrary, if the *x* and *x'* are completely independent of each other, then $p(x, x') = p(x)p(x')$, which leads the COR $R(x, x')$ to be $p(x')$. If we chose some threshold θ_r such that $p(x') < \theta_r < 1$, then by comparing with θ_r, the correspondence of two points in two camera views can be determined. Another threshold θ_0 is needed to be compared with *p(x)* to decide whether a point is detected. Thus, camera handoff is done of by calculating the correspondence of pairs of points in the views of different cameras.

3. Experiments

3.1. Data

We perform all the experiments using the commercially available AXIS cameras. The experiments include seven scenarios which are summarized in Table 1. The trajectories in each case are shown in Figure 1. Note that these figures are for illustration purposes only. They only give a rough idea of the trajectories for the convenience of readers. For some example views in these cases, the readers are referred to Figure 2 and Figure 3.

Table 1: *Experiment scenarios.*
($\boldsymbol{N_C}$: number of cameras; $\boldsymbol{N_P}$: number of persons; **Id**: indoor; **Od**: outdoor; **Of**: overlapped FOVs; **Nf**: Non-overlapped FOVs; **D**: 9am-5pm; **N**: 5pm-7pm; **Dc**: distinct colors; **Rc**: random colors; **L**: number of frames)

Cases	N_C	N_P	Id	Od	Of	Nf	D	N	Dc	Rc	L
C1	3	5	√		√		√		√		173
C2	3	1		√	√		√		√		103
C3	3	2		√	√		√		√		128
C4	4	6		√	√		√		√		697
C5	4	4		√	√	√	√			√	964
C6	4	4		√	√	√		√		√	1194
C7	4	4	√	√	√	√	√			√	1600

3.2. Trackers and face detector

We compare the above 3 camera handoff approaches on 2 color-based trackers, the Camshift tracker [8] and the particle filter tracker [8], which are referred to as T1 and T2 in the rest of this paper. There are many trackers that have been developed, the reason we choose these two because they belong to two different categories, i.e. kernel tracking and point tracking [8]. We use color as the feature for tracking because this is one of the most popular features and is easy to extract. Some trackers use shape or other templates as the feature, but they require training and are hard to implement for real-life experiments in real time. For data association, we first initialize the tracker manually and provide the objects' IDs to identify different objects. Afterwards, the objects are identified by the colors. The colors are extracted by selecting a small region on the person's upper body (on the person's coat) manually and then calculating the mean RGB value of that region.

For the utility-based and WAG approaches, we use the same criteria as those in [5, 6], namely, the size, position and view of the person. Face detection is done by implementing the OpenCV face detector [9] in an area above the upper body (defined by the bounding box returned by the tracker) whose size is 1/2 (1/2 of the length and same width) of the upper body's bounding box.

3.3. Evaluation metrics

1) ***Tracking error***: The overlap of the bounding box returned by the tracker and that of the ground-truth is less than 30% or the former one is over 1.5 times of the latter one. 2) ***Face detection error***: The overlap of the bounding box returned by the face detector and that of the ground truth is less than 30% or the former one is over 1.5 times of the latter one. 3) ***Camera handoff error***: A camera handoff error occurs when there is a handoff in the ground-truth but no hand-off is detected in the experimented approach or there is no hand off in the ground truth but there is one in the experimented approach. 4) ***Ground-truth for tracking:*** We manually label the bounding boxes for different persons. The bounding boxes are for upper body only, since we only use the color of a person's upper body as the feature for tracking. We also manually label bounding boxes for visible faces (over 200 pixels) as the ground truth for face detection. 5) ***Ground truth for camera handoff***: We enumerate all the camera assignment possibilities, based on the annotated video, and exhaustively choose the best one according to the corresponding criteria.

All the results for a case are averaged over all the cameras and all the persons for all the frames that are involved, which is defined as the error rate.

To clearly observe the effect of different trackers, we first tabulate the tracking error rates of the two tested trackers in the above 7 cases in Table 2. The face detection error rates are listed in Table 3. We observe that T2 works better than T1. Color-based trackers are less robust when the background is dark. That is why in C6 the error rate is relatively high.

Table 2: *Tracking error rates (%) in all experimental cases.*

	C1	C2	C3	C4	C5	C6	C7
T1	16.7	2.0	12.2	20.8	25.3	35.2	19.6
T2	8.6	1.2	9.3	14.9	22.8	30.7	14.3

Table 3: *Face detection error rates (%) on different trackers in all experimental cases.*

	C1	C2	C3	C4	C5	C6	C7
T1	19.5	12.3	15.5	25.5	30.1	30.0	25.7
T2	15.9	11.1	12.4	21.3	32.3	29.1	19.8

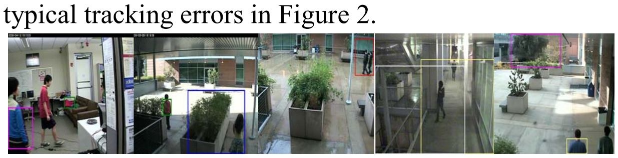
3.4. Evaluating camera handoff performance

We summarize the results in Table 4. We can observe that when the tracking error rate is high, there is not much difference among the camera handoff approaches. When the tracking error is less than 15%,

Table 4: Camera handoff performance in all experimented cases
(For C1-C7: (x, y, z), x: the number of correct handoffs, y: number of false alarms, z: number of false dismissal; For the overall performance: (x, y), x: the overall number of correct handoffs, y: the overall number of error handoffs)

Approach / tracker	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**Overall**
Utility-based / T1	25, 15, 7	8, 0, 0	12, 1, 2	90, 19, 65	36, 12, 60	39, 21, 50	79, 22, 34	289, 308
Utility-based / T2	28, 10, 4	8, 0, 0	13, 1, 2	116, 12, 39	50, 10, 46	32, 18, 57	99, 19, 14	349, 232
WAG / T1	24, 14, 8	8, 0, 0	13, 1, 1	92, 18, 63	45,16, 51	30, 18, 59	78, 18, 35	290, 302
WAG / T2	30, 12, 2	8, 0, 0	12, 0, 2	122, 17, 33	67, 16, 29	34, 17, 55	102, 13, 11	377, 207
COR / T1	18, 12, 14	5, 1, 3	10, 8, 4	69, 20, 86	32, 12, 64	29, 20, 60	50, 10, 63	213, 377
COR / T2	19, 12, 13	5, 1, 3	12, 13, 2	77, 19, 78	32, 9, 64	32, 21, 57	69, 11, 44	246, 347
Ground truth	32	8	14	155	96	89	113	507, 0

the utility-based approach and the WAG approach are less sensitive to tracking errors than the COR approach. As the number of cameras and persons increases, the WAG approach is more robust than the utility-based approach when a small number of iteration is allowed (15 in our experiments). T1 is easier to be distracted by other objects than T2. As long as it is distracted to some background objects, such as the trees or the ground, the tracker return the same result afterwards, making the number of camera handoff s to not change any more. We show some typical tracking errors in Figure 2.



A: T1 C1 B: T1 C3 C: T2 C5 D: T1/T2 C6 E: T1/T2 C7

Figure 2: Typical errors caused by different trackers in several cases. In A B and E, the trackers are distracted to other clusters with similar color. In C, the tracker cannot differentiate two persons having similar colors. In D, the environment is too dark to identify the two persons.

There are also some cases when there are tracking errors, but the system gives correct camera handoffs. This is because the criteria calculated from the inaccurate tracking result still give the correct direction for utility changes. So, when the system hands off to another camera, there is still a chance to start tracking correctly. However, this is more likely to happen in the two game theoretic approaches because of the multi-criteria that are used. With only position considered, the COR approach is less robust in this sense. Some examples are shown in Figure 3.

Figure 3: A correct camera handoff for the person in blue with tracking error by the WAG approach in case 7. The colored boxes are for the camera used to track the person whose bounding box is the same color. In this case, although the tracking results from camera 4 are incorrect, the system detect the person's face in Frame b and based on the criteria supplied, it hands off camera 4 to camera 5 for the person in blue, which has the same result as in the ground truth.

4. Conclusions

We performed experiments in 7 cases with various conditions (number of cameras and persons, lengths of videos, changing environmental conditions) to evaluate the effect of different trackers and handoff techniques. All the experiments are done in real-time (15fps) with real-life data. It is shown that the game theoretic approaches are less sensitive to tracking errors. Future work will be devoted to evaluate performance of algorithms over days and weeks.

5. Acknowledgement

This work was partially supported by NSF grants 0551741 and 0622176, 0905671 and ONR grants (DoD Instrumentation and the Aware Building). The contents of the information do not reflect the position or policy of the US Government.

References

[1] J. Krumm *et al*. Multi-camera multi-person tracking for Easy Living. *IEEE International Workshop on Visual Surveillance*, 2000.
[2] K. Heath and L. Guibas. Multi-person tracking from sparse 3D trajectories in a camera sensor network. *ICDSC* 2008.
[3] L. Martijn and G. M. Dariu. Multi-person tracking with overlapping cameras in complex, dynamic environments. *British Machine Vision Conference* 2009.
[4] L. Tessens *et al*. Principal view determination for camera selection in distributed smart camera networks. *ICDSC* 2008.
[5] Y. Li and B. Bhanu. Utility-based dynamic camera assignment and hand-off in a camera network. *ICDSC* 2008.
[6] Y. Li and B. Bhanu. Task-oriented camera assignment in a video network. *ICIP* 2009.
[7] Y. Jo and J. Han, A new approach to camera hand-off without camera calibration for the general scene with non-planar ground, *VSSN* 2006.
[8] A. Yilmaz *et al*. Object tracking: A survey. *ACM Computing Survey*, 2006.
[9] http://opencv.willowgarage.com/wiki/FaceDete