Coopetitive Multimedia Surveillance

Vivek K. Singh[1], Pradeep K. Atrey[1], and Mohan S. Kankanhalli[1]

School of Computing, National University of Singapore,
Singapore
{vivekkum, pradeepk, mohan}@comp.nus.edu.sg

Abstract. 'Coopetitive' interaction strategy has been shown to give better results than similar strategies like 'only cooperation', 'only competition' etc [7]. However, this has been studied only in the context of visual sensors and for handling non-simultaneous events. In this paper, we study this 'coopetitive' strategy from a multimedia surveillance system perspective, wherein the system needs to utilize multiple heterogeneous sensors and also handle multiple simultaneous events. Applying such an interaction strategy to multimedia surveillance systems is challenging because heterogeneous sensors have different capabilities for performing different sub-tasks as well as dissimilar response times. We adopt a merit-cum-availability based approach to allocate various sub-tasks to the competing sensors which eventually cooperate to achieve the specified system goal. Also, a 'coopetition' based strategy is adopted for effectively utilizing the information coming asynchronously from different data sources. Multiple simultaneous events (e.g. multiple intrusions) are handled by adopting a predictive strategy which estimates the exit time for each intruder and then uses this information for enhanced scheduling. The results obtained for two sets of surveillance experiments conducted with two active cameras and a motion sensor grid are promising.

1 Introduction

Recently a fair amount of interest has been generated on devising effective interaction and feedback mechanisms for multisensor surveillance[7, 3, 1]. For example, Singh and Atrey [7] propose the use of 'coopetitive' interaction approach for sensor interaction. Coopetition is a process in which the sensors compete as well as cooperate with each other to perform the designated task in the best possible manner. Intuitively, the process is similar to that of two partners in the card game of 'bridge' trying to outbid each other (compete), even though they are doing so for the benefit of the team (cooperate) in a bigger context. Such an interaction mechanism helps in improving the system performance by employing the best sensor for each (sub) task and also makes available the remaining sensors for undertaking other tasks if required.

While the above mentioned coopetitive strategy promises to be an effective interaction approach it has so far been studied for interaction between cameras alone. Similarly, other sensor interaction works like [3, 1] have also focused on interaction mechanisms between cameras only.

On the other hand, effective surveillance often requires multi-modal information which is obtained from more than one type of sensors. Hence, while adopting the salient features of coopetitive interaction as described in [7], in this paper we tackle the problems from a heterogeneous sensor perspective. Our key contributions are :

1. Applying the 'coopetitive' interaction strategy to multimedia surveillance systems wherein multiple heterogenous sensors are employed which may have different functional capabilities as well as dissimilar response times.
2. Enhancing the system capability to handle multiple simultaneous events e.g. handling multiple simultaneous intruders in monitored space.

In heterogeneous sensor environments each sensor has a different capability for handling the different sub-tasks i.e. divisible components of the system goal. While some sensors may be able to accomplish multiple types of sub-tasks, others may be able to do only one such type of sub-task. We adopt a suitability-cum-availability strategy for appropriately allocating the various sub-tasks to each of these sensors. Firstly, for each sub-task, a list of suitable sensors is made. Sensors are then allocated from this suitability list based on their availability.

Also, dissimilar sensors may have different response times. For example in a typical surveillance setup, a camera could be working at 4 frames/sec but a motion sensor may be providing information 6 times per second. Thus, to effectively utilize the information arriving asynchronously from such disparate sources, we adopt a 'coopetition' based strategy. The sensors 'compete' to provide information about newer events of interest. However, only the genuinely 'new' information is accepted by the system and the rest is termed as recurrent. Thus only the 'winning' sensor is allowed to trigger responses required for handling newer events of interest.

Handling multiple events which lead to concurrent sub-tasks is also a non-trivial issue as the system must decide an order for handling them such that the global output is maximized. To resolve this issue we employ a predictive strategy which allows us to estimate the deadlines for completing each sub-task beyond which they can not be completed satisfactorily. For example, in a multi-intruder scenario, the system evaluates the estimated time of exit for each intruder and then checks if the currently focused intruder shall still remain in monitored area even if it focuses on a newer intruder first. This additional information opens the doors to various predictive scheduling strategies which can in turn help to maximize the system performance.

To measure the effectiveness of our proposed approach we conduct experiments using two cameras and a grid of motion-sensors. However, it may be worth noting that we currently connect all the sensors to a central computer which acts as central coordinator/controller. Thus, we presently circumvent many complexities of distributed decision making like network delay, coordination cost etc. which are beyond the scope of the current work.

We define the system goal for the experiments as obtaining atleast three high resolution frontal facial images of any intruder entering the monitored premises. Hence the two major sub-tasks are localizing the intruder and focusing on him

(for obtaining his facial images). The motion sensor grid provides the localization information about the intruder while the cameras can provide both localization information as well as the facial data. The requirements for effective sub-task allocation among dissimilar sensors as well as the need to handle multiple simultaneous intrusion events makes these surveillance scenarios quite challenging.

Garcia et al in [3] described coordination and conflict resolution between cameras. Collins group [1] has done some pioneering work in multi-camera cooperation strategies. However, both of these works do not employ the use of competition which is we advocate as an integral part of coordination together with cooperation. Also, they deal only with cameras while we we want to be able to handle heterogenous sensors. We have earlier described 'coopetitive' interaction strategy with a homogeneous sensor (cameras) perspective in [7]. However, in that work we do not handle the issues posed by dissimilarity between sensors and also do not handle multiple simultaneous events e.g. multiple intrusions which we do in this work.

Doran et al [2] have described different types of cooperation from an artificial agent perspective. Murphy et al [6] illustrate cooperation between heterogeneous robots based on 'emotions'. Hu et al [4], discuss the relationships between local behaviors of agents and global performance of multi-agent systems. While these works provide insights into different types of cooperation from an agent perspective they do not discuss the practical issues which are faced in implementation of such interactive strategies across heterogeneous sensors. Lam et al [5] have described a predictive method for scheduling tasks on panning cameras. Their intention is to schedule resource intensive tasks at a time when the system load is least. While their idea of scheduling is interesting, this work does not deal with the either the interaction strategy or the practical issues of of heterogeneous sensing. Hence, we realize that there are no existing works which tackle the problem of effective interaction between multiple sensors while also considering the issues posed by their heterogeneousness.

The outline of the remainder of this paper is as follows. Section 2 describes how we tackle the issues of appropriate sub-task allocation, asynchronous data utilization and multiple simultaneous events handling. Section 3 describes the experimental results obtained for the surveillance experiments. The conclusions and possible future works have been discussed in Section 4.

2 Proposed Work

2.1 Coopetitive Framework for Heterogeneous Sensors

The generic algorithm for 'coopetitive' interaction strategy is shown in figure 1. Upon arrival of each new surveillance object S_Obj (i.e the object or the person under observation), the system divides the overall goal into sub-tasks. These sub-tasks are allocated to the available sensors based on a suitability-cum-availability strategy. Initially, the sensors 'compete' to be allocated the various sub-tasks and later 'cooperate' with each-other to better perform the respective allocated sub-

tasks. The roles i.e. sub-task allocation can be swapped if environment changes or it is realized that the current sub-task allocation is no longer appropriate.

A key point to note is that 'coopetition' does not specify any particular type of sensor, measure of merit or even type of scenario. For example, this work describes 'coopetition' between two cameras and motion sensor grid but nothing restricts us from using audio sensors, pressure sensors etc. which may be suitable for some other scenario. Similarly, the measure of merit can also be chosen based on the system task. For example in tracking-like applications, resolution of facial images obtained, body blob size, physical proximity to the sensor, and camera resolution/quality all form reasonable measures of merit for handling competition between sensors for task allocation. Lastly, 'coopetition' does not restrict itself to any particular type of application. It is equally relevant to multiple search teams operating in rescue effort or multiple satellites providing data about approaching tsunami/storms etc.

More details about 'coopetitive' strategy in general can be found in [7]. Here we focus on issues which are unique to heterogeneous multimedia sensor systems.

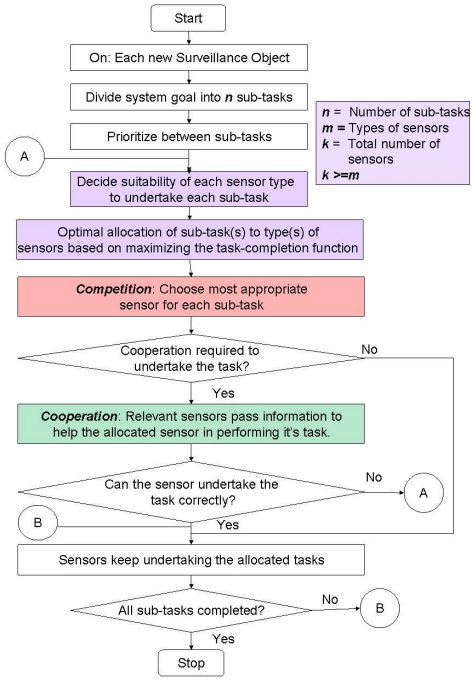

Fig. 1. Proposed 'coopetitive' strategy for heterogeneous sensor interaction

Firstly, the system needs to allocate sub-tasks to sensors based on the suitability of each type of sensor for the various sub-tasks. Let there be n different

sub-tasks viz ST_1 through ST_n, each of which can be undertaken by one or more of the m different sensor types viz $SensorType_1$ through $SensorType_m$ as shown in figure 2. Each sensor type has one or more member sensors which can be represented as $Sensor_1$ through $Sensor_o$. It is possible to come up with a set $STCap_i$ for each sub-task that contains the list of all sensors which can handle the i^{th} sub-task. We must note that at any instant, there must be atleast one sensor capable of handling each sub-task i.e. $\forall STCap_i : STCap_i \neq \emptyset$. Our approach for allocating the sensors to the various sub-tasks is as follows.

Step 1: Choose the most restrictive sub-task p which can be undertaken by the least number of sensors i.e. $STCap_p : \forall STCap_i, \cap(STCap_p) \leq \cap(STCap_i)$

Step 2: If $STCap_p$ has only one sensor as its member, Allocate it.

Step 3: If it has more than one member, Allocate from the sensor-type with the highest availability.

Step 4: Remove the allocated sub-task and the sensor from the allocation pool.

Step 5: If number of sub-tasks in pool $\neq 0$, Go to Step 1

Step 6: After all sub-tasks have been allocated the minimum one sensors, allow the 'redundant' sensors to also perform tasks per their capabilities.

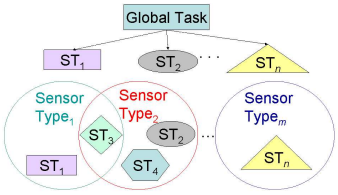

Fig. 2. Sub-tasks and their allocation

Now, let us understand this allocation in our surveillance setup which has two sub-tasks: S_Obj localization and S_Obj focusing. The S_Obj focusing can be undertaken by the two cameras while the S_Obj localization can be done by two cameras as well as the motion-sensor grid. The better suited of the two cameras is allocated the S_Obj focusing task. Thus for our experimental setup, the camera which can obtain better frontal facial images of the intruder is allocated the focusing task. Similarly, the sensor (camera or the motion-sensor grid) which provides first information about a 'new' intruder handles the localization task (further discussed in section 2.2).

2.2 Handling asynschronous data

In multimedia surveillance, the system may encounter asynchronous data about similar context coming from multiple sensors at different time intervals. To understand this, let us look at a scenario as shown in figure 3 in which two types of sensors are providing similar type of information. The first sensor starts giving information at $t = 0$ and continues to provide information at the interval of 10

time-units. On the other hand sensor 2 starts providing information from $t = 15$ and continues to do so every 15 time-units.

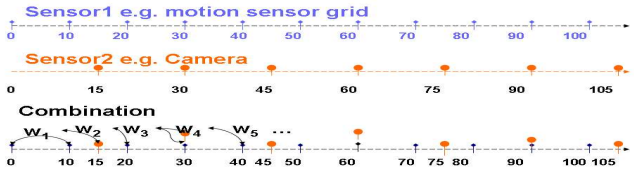

Fig. 3. Coopetition approach for using asynchronous data

In our surveillance setup a similar situation arises when motion sensor grid and cameras both compete to provide localization information for a new S_Obj. Whichever sensor provides the information about the first S_Obj for the first time clearly 'wins' the competition for the first case. For subsequent cases, the sensors could either be providing recurrent information about an already focused S_Obj or it could be the information about a genuinely new S_Obj. Finding whether the S_Obj is old or new is important for the system to react accordingly as normally the initial overhead for the system to react to a new object are significantly more than the effort required for per-frame tracking of an old object. Hence we compare each new information with the last 'winning' entry to see if it is recurrent (competition 'lost') or new (competition 'won'). Only the information from the 'winning' sensor is allowed to trigger the response process for handling a 'new' S_Obj. This decision process has been demonstrated in figure 3 with w_1, w_2 etc. representing the winning entries. The recurrent information on the other hand is simply added to the position information vector which is maintained for each individual S_Obj currently being focused and used for per-frame tracking.

2.3 Handling multiple simultaneous events

A multimedia surveillance system must decide the order for undertaking the various sub-tasks. Certain sub-tasks might always be required to be undertaken before others. For example, in our scenario, localization needs to be done before the facial images can be captured. Hence the system undertakes such tasks first. However, when dealing with multiple simultaneous surveillance events, the system may encounter multiple sub-tasks which can be executed in any arbitrary order. In such situations, the system must choose the execution order so as to maximize the global output function. Such a global output function can be represented as $\sum_{i=1}^{n} w_i \times ST_i$, where w_i is the weight/importance assigned to the i^{th} sub-task. This signifies that in case of a conflict i.e. if undertaking only one of the z remaining sub-tasks is possible, we must choose the one with the highest importance. We represent the conflicting z sub-tasks as CST_1 though CST_z and assume that they are arranged in order of their importance in a descending

order. Let us represent our set of sub-tasks which can be actually be completed as $CanDo$. Hence our combined strategy for undertaking the various tasks using a 'greedy' algorithm is as follows:

$CanDo \leftarrow CST_1$

For: r= 1 to z

if (CST_r can still be completed after finishing CST_{r+1})

    $CanDo \leftarrow CanDo \bigcup CST_{r+1}$

else

    Do nothing;

Next r

We encounter such a conflicting scenario in our implemented scenario when multiple intruders enter the monitored space simultaneously. Using a predictive methodology the system constantly keeps track of the $ETEx$ (Expected Time of Exit) for each S_Obj. If the system realizes that intruder r is going to stay inside the monitored premises even if it 'takes time off' to focus on intruder $r+1$ first, then it would do so. Else, it would continue to focus on intruder r. The estimate on how long the intruder shall stay inside the monitored premises can be undertaken using Kalman filter approach by keeping track of position and velocity vectors of the various intruders as described in [7].

3 Experimental Results

We conduct two sets of experiments to verify the suitability of our proposed approaches. In first experiment we consider a single door enclosed environment setup like that commonly found in ATM lobbies or museum sub-sections. The system task is to obtain at least 3 frontal facial images of 200px by 200px resolution (which suffices for most face identification/expression recognition algorithms[8]) for each intruder and then continue to obtain more images if there are no other intruders. The setup consists of two active cameras, one placed directly above the entrance and the other directly above the principal artifact e.g. ATM machine. Essentially, the setup is similar to that described in [7] but now we also have an additional motion-sensor grid covering the entire premises to provide additional localization information.

In this experiment, we compare the performance of 'only cooperation', 'only competition', 'coopetition without MPC' and 'Coopetition with MPC' approaches for heterogeneous sensors and also see how they relate to the results obtained using only cameras [7]. In 'only cooperation' interaction approach sensors try to help each other in better performing their sub-tasks e.g. by passing S_Obj localization information. However there is no differentiation between sensors based on their worthiness and inappropriate sensors may also be passed information and allocated critical roles e.g. face capturing sub-task may be allocated to the camera in opposite direction with the face. In 'only competition' approach, the sensors do not help other sensors in performing their sub-tasks. So, in this case the cameras need to perform both S_Obj localization and focusing on their own. In 'coopetition' based approach the sensors help each other, but do so only if

the other sensor is worthy of performing the allocated sub-task. Relating to our setup, the S_Obj location is transferred only to the appropriate camera(s) which use this information to focus on the S_Obj and obtain facial images better. In 'coopetition with MPC' approach the sensors have the additional advantage of using a predictive methodology to order and perform their sub-tasks.

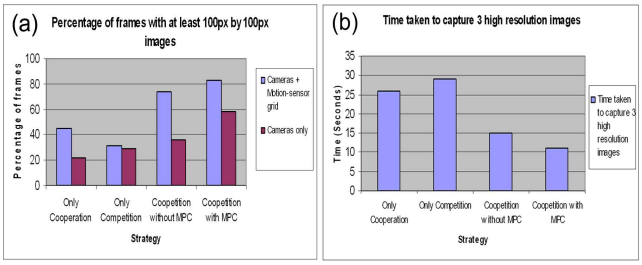

Fig. 4. Comparison of interaction strategies

For this experiment, the volunteer intruders were asked to enter the monitored premises for a one minute duration and intentionally avoid getting their facial images captured by the cameras. Twenty rounds of this experiment were conducted for each strategy and we compare the average data based on the percentage of frames for which atleast 100px by 100px images were obtained (figure 4a) and the time taken to capture three 200px by 200px facial images (figure 4b). Hence in all, we used data from 80 one minute rounds using two cameras each working at 768px by 576px with 4 fps for obtaining these results.

The presented figure 4a also shows the image capturing results obtained for similar setup with cameras alone [7]. We notice consistent increase in face capturing capability with the addition of motion-sensor grid. This corroborates well with the additional localization information available from motion-sensor grid which is faster (no need to continuously pan) and more robust (non-frontal face situations can also be handled). We notice that the 'coopetitive with MPC' strategy significantly outperforms the other strategies for heterogeneous sensors too and is able to obtain appropriate facial images for 82% of frames.

Figure 4b shows that the average time taken to obtain three 200px by 200px frontal facial images is also the least (11 sec) for 'coopetitive with MPC' approach. This is due to appropriate allocation of sensors to the sub-tasks, ability to obtain localization information from other sensors and performing forward state estimation for intruder's trajectory. As one each of these features is not present in 'only cooperation', 'only competition' and 'coopetitive(without MPC)' approaches respectively, the 'coopetitive with MPC' approach works better.

Being convinced about the superiority of 'coopetitive' interaction approach over 'only competition' and 'only cooperation' approaches, in the second experiment we closely compare the two variants of 'coopetition' with an aim to

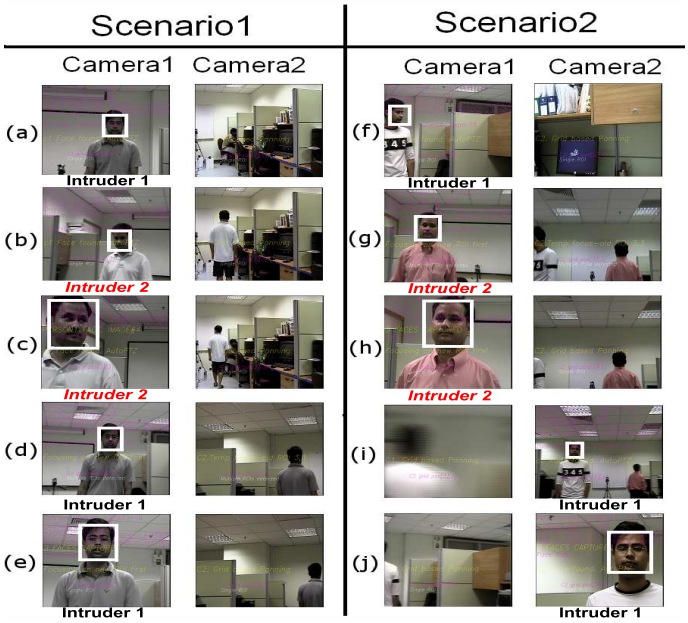

Fig. 5. Two example scenarios with multiple simultaneous events

verify the gains obtained by using a predictive strategy for handling multiple simultaneous events. While the proposed methodology poses no limitation on the number of such simultaneous events, we use a two intruder base case scenario to demonstrate its effectiveness. We consider the scenario of a walkway leading to an important establishment e.g. control room for a nuclear plant etc. with a system goal of obtaining three high resolution(200px by 200px) frontal facial images of two intruders simultaneously walking across the walkway with differing speeds. The system uses a predictive methodology to decide which intruder shall exit the monitored walkway first and then plan the order of focusing on them as to maximize the probability of capturing appropriate images for both of them. We conducted 20 rounds of experiments and found that the 'coopetitive with MPC' approach which uses a predictive strategy was able to obtain appropriate images for both intruders 85% of times as compared to 'coopetitive without MPC' approach which could do so only 65% of times.

Two example scenarios have been shown in figure 5 with time-instances labelled (a) through (j). Scenario1 shows how the system handles two intruders simultaneously walking towards the important room (direction from camera2 towards camera1). The camera1 notices the intruder1 first(5a) but then realizes that intruder2 is going to exit much faster and it has high probability of capturing intruder1 even after completing intruder2 focusing task. Thus it focuses on intruder2(5b). After obtaining three images of intruder2(5c) it goes back to

focus on intruder1($5d$) and captures it's three high resolution images($5e$). Scenario2 shows a similar case where intruder1 is detected first($5f$) but intruder2 is focused($5g$) as it is exiting faster and it's 3 facial images are obtained($5h$). However, this time around intruder1 changes his direction and camera2 is required to focus on it($5i$) and obtain three high resolution images($5j$).

4 Conclusions

In this paper we have extended the 'coopetitive' interaction strategy to multimedia surveillance systems which can handle multiple simultaneous surveillance events. The major issues handled in this extension were of sensor to sub-task allocation, asynchronous data handling and handling of multiple simultaneous sub-tasks. Results obtained from two sets of experiments have demonstrated that 'coopetition with MPC' strategy does work well with heterogeneous systems too. It can also handle multiple simultaneous events and continues to significantly outperform other related strategies like 'only cooperation' and 'only competition'.

We realize that the currently adopted methods for sub task allocation and sceduling are 'greedy' and hence not universally applicable. We intend to explore globally optimal solutions in future work. Further experimentation with larger number of sensors which are also of different type (e.g. infra-red camera, audio sensors, pressure sensors etc.) shall also be undertaken as part of our future work. We also intend to work on handling the complexities of distributed coordination mechanisms which have been ignored for the current work.

References

1. R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade. Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE*, 89(10):1456 – 1477, 2001.
2. J. E. Doran, S. Franklin, N. R. Jennings, and T. J. Norman. On cooperation in multi-agent systems. *The Knowledge Engineering Review*, 12(3):309–314, 1997.
3. J. Garcia, J. Carbo, and J. M. Molina. Agent-based coordination of cameras. *International Journal of Computer Science and Applications*, 2(1):33–37, 2005.
4. B. Hu, J. Liu, and X. Jin. From local behaviors to global performance in a multi-agent system. In *IEEE/ACM conference on Intelligent Agent Technology*, 2004.
5. K.-Y. Lam and C. K. H. Chiu. Adaptive visual object surveillance with continuously moving panning camera. In *ACM International workshop on Video Surveillance and Sensor Networks*, pages 29–38, 2004.
6. R. R. Murphy, C. L. Lisetti, R. Tardif, L. Irish, and A. Gage. Emotion-based control of cooperating heterogeneous mobile robots. *IEEE Transactions on Robotics and Control*, 18(5):744–757, 2002.
7. V. K. Singh and P. K. Atrey. Coopetitive visual surveillance using model predictive control. In *ACM International workshop on Video Surveillance and Sensor Networks*, pages 149–158, 2005.
8. Y.-L. Tian, L. Brown, A. Hampapur, S. Pankanti, A. W. Senior, and R. M. Bolle. Adaptive visual object surveillance with continuously moving panning camera. In *IEEE workshop on performance evaluation of tracking and surveillance*, 2003.