Colour-based Object Tracking in Surveillance Application

Tang Sze Ling, Liang Kim Meng, Lim Mei Kuan, Zulaikha Kadim and Ahmed A. Baha'a Al-Deen

Abstract—**The motivation of the proposed method is to resolve typical tracking's challenge which is object's occlusion in the scene. In this paper, we proposed a method to track the objects consistently in real-time unconstraint surveillance application. The proposed method capable to detect moving objects, track the objects appear in the scene and provide consistent identifier for tracked objects. We describe the characteristic of the motion tracker which based on colour as the key feature to compare the object's similarity. Initially, we produce a motion map which delineates the foreground and background. Then, the motion map is segmented and analyzed in order to remove noise and connect into motion regions. Subsequently, motion tracking is performed based on colour-based feature matching. The colour information is extracted cluster-by-cluster to compare the object's similarity across the image sequences. We ensure the efficiency of the method with assembly an assumption that only entire objects are allowed to be track. In order to support the assumption, we utilized wide-view camera to avoid capturing partial objects.**

***Index Terms*—Colour-based tracking, Occlusion, Group merging and splitting.**

I. Introduction

Typical surveillance system requires human concentration to monitor behaviors over complex scenes. This type of observation task is not appropriate for human performance, as it requires an over long period and careful concentration. As a result, the motivation to develop automated vision-based monitoring system is clear which relieve human to process risk detection and analysis. Object tracking plays important roles in surveillance system. In order to achieve the surveillance application efficiently, the object must be tracked before recognition.

The major contribution of object tracking is in determining the position of an object of interest moving in the scene at each moment in time. In other words, object tracking described as tracking the motion of an object of interest by consistently assigning tags to the object throughout consecutive video frames of a scene. Highlight here, the challenges will not be occur if the objects appeared in the scene is continuously noticeable without occluded with others objects in scene (i.e. the shape, size and motion not diverge over time). Yet, in real-time unconstraint environment, the statement aforementioned will not occur especially human as the object of interest. For instance, the human profile varies while moving with inconsistent motions. Besides, occlusion with objects or structures in the scene directs to discontinuity in the observation of objects induced the complexity of object tracking.

Object occlusion can be categorized into two major categories: 1) *inter-object occlusion* and 2) *occlusion of objects* [1]. *Inter-object occlusion* happens when group of objects appeared in the scene where objects are blocked by other objects in the group. The difficulty is to identify the objects in the group and determine the position in the region. *Occlusion of objects* occurs when the object is not observable or disappeared for an amount of time blocked by structures or objects (stationary or dynamic) in the scene. For instance, a person walks behind a pillar. Hence, a decision required concludes either wait for the object's reappearance or the object has leaved the scene.

Due to the issues aforementioned, we proposed a method aim to resolve problems with: 1) identify the objects when individual merge into or split from the group and 2) identify the objects after disappeared for an amount of time. Object tracking essentially to be complex yet attempt to reduce cost of computation. To tradeoff these inverse proportionate conditions, we propose a colour-based object tracking which require low computation cost and capable to track and tag objects with acceptable complexity. In the paper presented by L. S. Boon [2] has claims the advantages of colour in tracking applications. From the experimental results, colour has the ability of discriminate various objects; therefore it provides very useful information especially during occlusion. Yet, further analysis required in order to accurately identifying objects in the scene.

This paper is organized as follow: Section 2 briefly review the conventional object tracking methods. In section 3, we describe the proposed method in detailed which consist of the techniques and the chronological processes of object tracking. Then, we demonstrate the experimental results and followed by the discussion in section 4. Finally, we conclude this investigation in section 5.

II. Overview

We survey the techniques for tracking objects from object detection and method related to object tracking, specifically approaches that perform colour-based tracking and handle occlusions.

As mentioned in the previous section, an object required to be tracked before recognition, yet objects also compulsory to be detected before tracked. Tracking objects can be complex

Manuscript received December 30, 2009.
Tang Sze Ling is with MIMOS Berhad, Malaysia (e-mail: sze.ling@mimos.my).
Liang Kim Meng is with MIMOS Berhad, Malaysia (e-mail: liang.kimmeng@mimos.my).
Lim Mei Kuan is with MIMOS Berhad, Malaysia (e-mail: lim.mkuan@mimos.my).
Zulaikha Kadim is with MIMOS Berhad, Malaysia (e-mail: zulaikha.kadim@mimos.my).
Ahmed A. Baha'a Al-Deen is with MIMOS Berhad, Malaysia (e-mail: ahmed.bahaa@mimos.my).

due to numerous issues. In fact, some of the issues are propagated from object detection. For example, the scene illumination changes, undesired moving objects, and shadows. R. Cucchiara et al. [3] proposed Sakbot system which is a robust and efficient detection techniques based on statistical and knowledge-based background up and use HSV colour information for shadow suppression. The method capable to deal with luminance condition changes (e.g. lights, day hours and shadows), limited and high frequency camera motions (e.g. vibrations and wind), background changes (e.g. still objects) and various moving object's speed.

Object tracking defined as the problem to estimate the trajectory of the object of interest moving in the scene. Conventional tracking process consists of establishment of correspondences of the image formation between consecutive frames based on features (e.g. colour, position, shape, velocity) and involves matching between frame using pixels, points, lines or blobs based on their motion. The assortment of blobs composed a representation of the human. In the early generation, Pfinder [4] is a well-known method which tracks the single entire human body in the scene without occlusion. This method modeled pixel colour disparity using multivariate Gaussian. S. J. McKenna et al. [5 and 6] then performed a tracking at three levels of abstraction (i.e. regions, people and groups) to tracked people through mutual occlusions as they form groups and separate from one another. Colour information (i.e. colour histogram and Gaussian mixture model) is used to disambiguate occlusions and to provide estimation of depth ordering and position during occlusion. T. Boult et al. [7] presented a system which monitoring non-cooperative and camouflaged targets, is suggested for visual surveillance domain especially controlled outdoor environment (e.g. parking lots and university campuses) with low contrasts targets moving in changing environments with high occlusion. W4 [8] employ projection histogram to locate human body part (i.e. head, hands, feet and torso) and distinguished single person and group. In addition, each person in a group is tracked by tracking the head of that person. Hydra [9] essentially is an extension of W4 which developed by University of Maryland. Yet, both approaches not using colour cues for tracking. A. J. Lipton et al. [10], using shape and colour information to detect and track multiple people and vehicles in a cluttered scene and monitor activities over a large area and extended periods of time. However, these methods required complicated calculation or expensive computational power, thus we proposed a rule-based object tracking to identify persons appear in the scene based on colour information of the moving objects.

III. Method

Our proposed algorithm aims to assign consistent identifier to each object appears in the scene when 1) individual merge into or split from the group and 2) after disappeared for an amount of time. It involves several essential practices in order to encounter the lowest possibility of false tracking and tagging [12]. Generally, the structural design of our proposed method shown in Fig. 1:

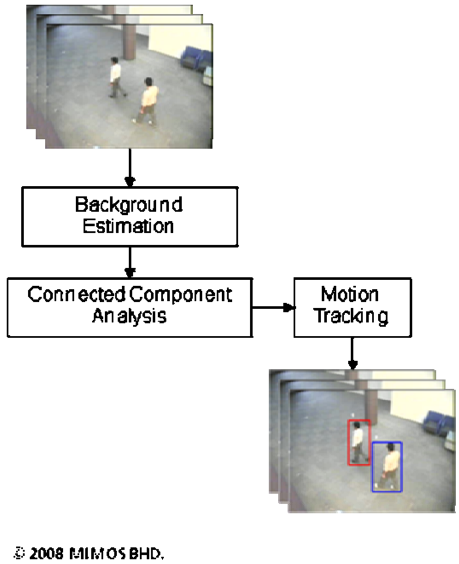

Fig. 1: Overview architecture of proposed method.

A. *Background Estimation*

Background subtraction delineates the foreground from background in the images. Initially, this method estimates the background from a certain number of early frames without any object inside (refer to Fig. 2). Subsequently, the following image frames are compared to the estimated background and calculates the difference in pixel's intensity. The drastic changes in pixel's intensity indicate that the pixel is a motion pixel. The background subtraction step generates a binary map where the non-zero pixels are corresponding to motion pixels. We employed Sakbot [3] algorithm in order to solve the problems which mentioned in section 2 (i.e. the scene illumination changes, undesired moving objects, and shadows) specifically to detect shadow whereby the shadows can affected the tracking process.

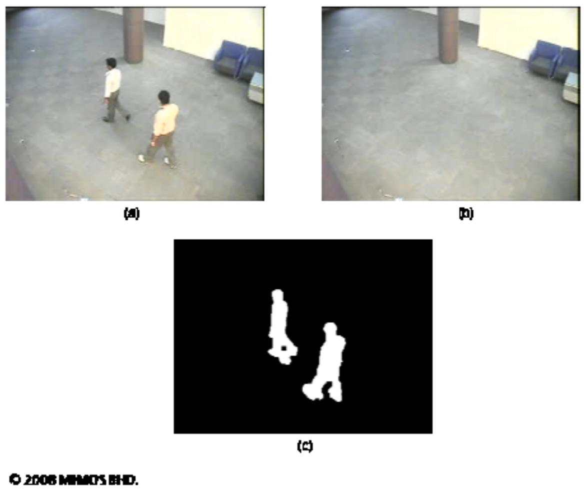

Fig. 2: (a) Current frame with Moving objects.
(b) Background image of the scene.
(c) Binary image from Background Estimation

B. *Connected Component Analysis*

Then a post processing step is applied on the binary image to remove noises (present as small group of motion pixels) and label groups of connected motion pixels as motion blobs using connected component analysis. Connected component analysis used to connect homogenous regions in an image, in our context a binary image [10]. The key idea of connected component analysis is to attach the adjacent foreground's pixel (i.e. white pixels) in order to construct a region. Thus, the final map from this section called motion map with connected motion region is called as motion blob.

C. *Motion Tracking*

The key feature of proposed method is the colour information of each object is extracted cluster-by-cluster. Each cluster has its own weightage for comparison. Contrast to some conventional methods which computed only an enclosed region of objects without part-based analysis, our proposed method directly improved the accuracy of tracking object with discontinuity observation (i.e. reappearance after occlusion) in the scene.

The proposed method tracked the object and tagged as a motion block which includes detecting a lateral view of motion blocks in a current frame to identify the occlusion of the motion blocks happened in the current frame. The colour information is extracted from the motion blocks in the current frame to categorize matching colour information between motion blocks in the current frame and previous frames. Subsequently, a tag is assigned to the motion blocks in the current frame. In our context, feature-based motion tracking is investigated which encountered colour as the main feature in the tracking process. Basically, the properties of the motion blocks include colour, distance, area, velocity, speed, etc. of the objects. Yet, the main feature extracted for similarity comparison is colour information. The colour information is extracted based on luminance and chrominance measures. This enables extraction of colour information from monochromatic and coloured objects during the day as well as at night. The following sub-sections explained the work flow of colour-based object tracking in details and illustrated in Figure 4.

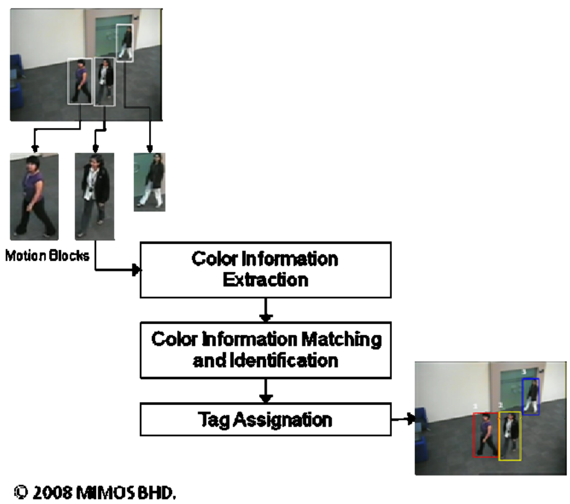

Fig. 4: The work flow of colour-based motion tracking component.

1) *Colour Information Extraction* – In this sub-task, each motion block in the current frame is segmented into areas of almost similar colour known as clusters (refer to Fig.5). For each cluster of the motion block, colour information is then derived. This colour information is known as cluster colour information which consists of 3 colour channel, i.e. Red, Green and Blue (RGB). The cluster colour information is computed using colour quantization and it consists of a fixed number of square bins in a 3D colour cube. The number of square bins is based on the requirement of the automated system in terms of the level of accuracy required to identify matching colour information between the motion blocks in the current frame and all motion blocks in the previous frames. In the following section, the term of motion blocks in the previous frames includes all motion blocks in the previous frame and any motion blocks that had left the scene. Figure 5 shows a flowchart that illustrates the steps of colour information extraction.

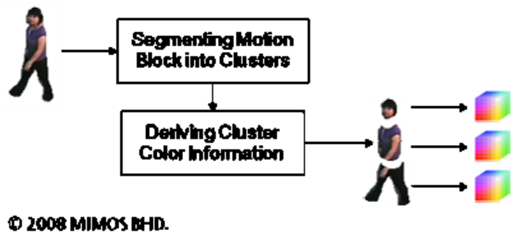

Fig. 5: The flowchart of colour information extraction.

2) *Colour Information Matching and Identification* - The processor is programmed to identify matching colour information between motion blocks in the current frame and all motion blocks in the previous frames using weighted cluster-based matching. Weighted cluster-based matching begins with comparing the cluster colour information of a cluster of the motion block in the current frame with the cluster colour information of clusters in all motion blocks in the previous frames. This is repeated for every cluster of the motion block in the current frame. For each comparison made, the processor computes a respective comparison score. The comparison score for each of the clusters of the motion block in the current frame is stored in the data storage system. The processor then identifies the highest comparison score of each cluster in the current frame.

Prior to computing an average comparison score of the motion blocks in the current frame, the processor assigns a predetermined weight for each cluster of the motion block in the current frame. The predetermined weight is assigned based on the location of the cluster in the motion block. The predetermined weight assigned for each cluster of the motion block in the current frame is stored in the data storage system.

Fig. 6 illustrates the steps of computing the average comparison score of the motion blocks in the current frame by comparing the cluster colour information of all clusters of the motion blocks in the current frame with the cluster colour information of all clusters of the motion blocks in the previous frames. The motion block in the current frame is segmented into three clusters. The corresponding motion block in the previous frame is also segmented into three clusters.

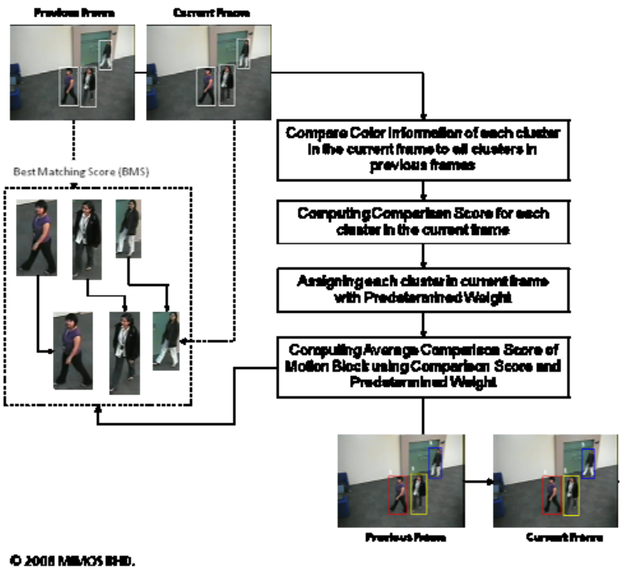

Fig. 6: The steps to identify matched colour information.

The cluster colour information of the first cluster of the motion block in the current frame is compared with the cluster colour information of all three clusters of the motion block in the previous frame. The processor computes a comparison score for each of the three comparisons made. This is repeated for second cluster and the third cluster of the motion block in the current frame, wherein the cluster colour information of the second cluster and the third cluster of the motion block in the current frame are respectively compared with the cluster colour information of all three clusters of the motion block in the previous frame. The comparison scores for each of the three clusters of the motion block in the current frame are stored in the data storage system.

Based on the computed comparison scores, the processor then identifies the highest comparison score for the first cluster of the motion block in the current frame and repeated for the second cluster and the third cluster of the motion block in the current frame, respectively. Next, the average of comparison score of the motion block in the current frame using the highest comparison scores of the clusters of the motion block in the current frame and the predetermine weight assigned for the clusters of the motion block in the current frame.

3) Tag Assignation - Once the average comparison score of the motion block in the current frame is computed, the processor then assigns a tag to the motion blocks in the current frame. Figure 7 presents a flowchart that illustrates the steps of tagging the motion blocks in the current frame.

The processor tags the motion blocks in the current frame with either a tag similar to that of the previous frames, including tags of motion blocks that had left the scene or a new tag. The decision to retain a tag or assign a new tag is dependent on the average comparison score computed for the motion block in the current frame and all motion blocks in the previous frame.

If a motion block in the previous frames is tagged as *N*, and the motion block in the current frame has an average comparison score that is higher than a predetermined threshold of the motion block in the previous frames, then the motion block in the current frame will be assigned with the same tag, *N*. On the other hand, if a motion block in the previous frames is tagged as *N*, and the motion block in the current frame has an average comparison score that is lower than the predetermined threshold of the motion block in the previous frames, then the motion block in the current frame will be assigned with a new tag, as depicted in Fig. 7.

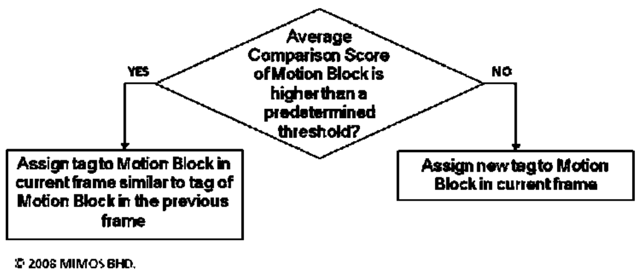

Fig. 7: The flowchart of tag assignation in the current frame.

IV. Results & Discussion

We show the experimental results with 3 image sequences consists of 4 persons, 3 persons and 2 persons to illustrate different conditions which is 1) identify objects when merge into or split from group and 2) identify reappear objects after an amount of time. We ensure the efficiency of the method with assembly an assumption that only entire objects are allowed to be track. In order to support the assumption, we utilized wide-view camera to avoid capturing partial objects. To verify the ability of proposed method, the implemented algorithm must fulfill following requirement:

1. Assign the consistent identifier to similar objects that are wandering in the scene.

2. Able to dispense same identifier to the similar object which had left the scene less than a certain preset number of frames.

3. Identify the occluded objects which merge into group.

4. Identify the individual which split from the group.

Objects reappearance - Fig. 8 shows a number of frame extracted from image sequence captured at outdoor with shelter where natural light source is provided, 4 persons are appearing with lateral view in the scene consequently. Frame 320 to frame 420 demonstrates the new objects appear in scene with new identifier provided. Frame 320 shows the first object appears in the scene and given identifier 1. In frame 350, second object appears with identifier 2. Then the first object disappears in frame 390, but third object appears with second object with identifier 3. In frame 420, a new object appears with identifier 4 in the scene. In the second row of the frames (i.e. frame 600 to frame 715), the objects reappear in the scene after few frames (average 250 frames) enter from the similar position. The objects are identified and labeled with the old identifier. The results from Fig. 8 had evidenced the ability of the proposed method in requirement 1 and 2 abovementioned.

Objects merging and splitting – Fig. 9 shows 2 persons acting in the scene similar to Fig. 8 where natural light source is provided but different views are captured (i.e. back view or front view). In frame 390, both objects are observable with provided different identifiers. Then two objects are occluded or merging into 1 group, yet the occluded object is identified and same identifier is provided in frame 400. While frame 410 and frame 420 demonstrate identification when the object splitting from group. To evaluate the proposed method with different complexity dataset, image sequence shows in Fig. 10 are captured where 3 persons are acting in the scene. Similar results are verified indicates the ability of proposed method with various complexities and established the rest of the requirements, requirement 3 and requirement 4 in aforementioned.

V. Conclusion

The advantages of using colour as feature to achieve object's similarity are robust against the complex, deformed and changeable shape (i.e. different human profiles). In addition, it is also scale and rotation invariant, as well as faster in terms of processing time. From the experimental results, it shows the ability to discriminate various occluded colour objects.

This proposed method provides a potential to the surveillance system which required wide area observation and tracking objects over multiple cameras, for example airport, train station or shopping malls. It is not possible for a single camera to observe the complete area of interest as the structures in the scene constraint the visible areas and the devices resolution is restricted. Therefore, surveillance system of wide areas required multiple cameras for objects tracking. Our future work is to implement our proposed method in multiple camera system in order to measure the robustness of the method in assigning a similar identifier for the similar object moving from one camera to another camera.

Acknowledgment

This work was supported by the MIMOS Berhad, Malaysia.

References

[1] O. Javed and M. Shah, "Tracking and object classification for automated surveillance," In *Proc. 7th European Conf. Computer Vision*, Springer, 2002, pp. 343.
[2] L. S. Boon, "Tracking and surveillance," School of EEE, Nanyang Technological University, Singapore.
[3] R. Cucchiara, C. Grana, G. Neri, M. Piccardi and A. Prati, "The Sakbot system for moving object detection and tracking," *Video-based Surveillance Systems-Computer vision and Distributed Processing*, pp. 145-157, 2001.
[4] C. Wren, A. Azarbayejani, T. Darrell, A. Pentl, "Pfinder: Real-time tracking of the human body," In *IEEE Trans. Pattern Analysis and Machine Intelligent,* vol. 19, no. 7, pp. 780-785.
[5] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld and H. Wechsler, "Tracking group of people," *Comput. Vis. Image Understanding*, vol. 80, no. 1, pp. 42-56, 2000.
[6] S. J. McKenna, S. Jabri, Z. Duric and H. Wechsler, "Tracking interacting people," Dept. of Applied Computing, University of Dundee, Scotland, 1999.
[7] T. E. Boult, R. J. Micheals, X. Gao and M. Eckmann, "Into the woods: Visual surveillance of noncooperative and camouflaged targets in complex outdoor settings," In *Proc IEEE*, IEEE Press, vol. 89, no. 10, 2001, pp. 1382-1402.
[8] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," In *IEEE Trans. Pattern Analysis and Machine Intelligent*, vol. 22, no. 8, 2000, pp. 809-830.
[9] I. Haritaoglu, D. Harwood, and L. S. Davis, "Hydra: Multiple people detection and tracking using silhouettes," In *IEEE International Workshop on Visual Surveillance*, pp. 6-13, June 1999.
[10] A. Lipton, H. Fujiyoshi and R. Patil, "Moving target classification and tracking from real-time video," In *DARPA Image Understanding Workshop*, pp. 129-136, November 1998.
[11] M. Sonka, V. Hlavac and R. Boyle, "*Image processing, Analysis and Machine Vision*," PWS publishing, 2nd Edition, 1999.
[12] *Method and system for tracking and tagging objects*. Malaysia patent office. Pending for PI: 20083070

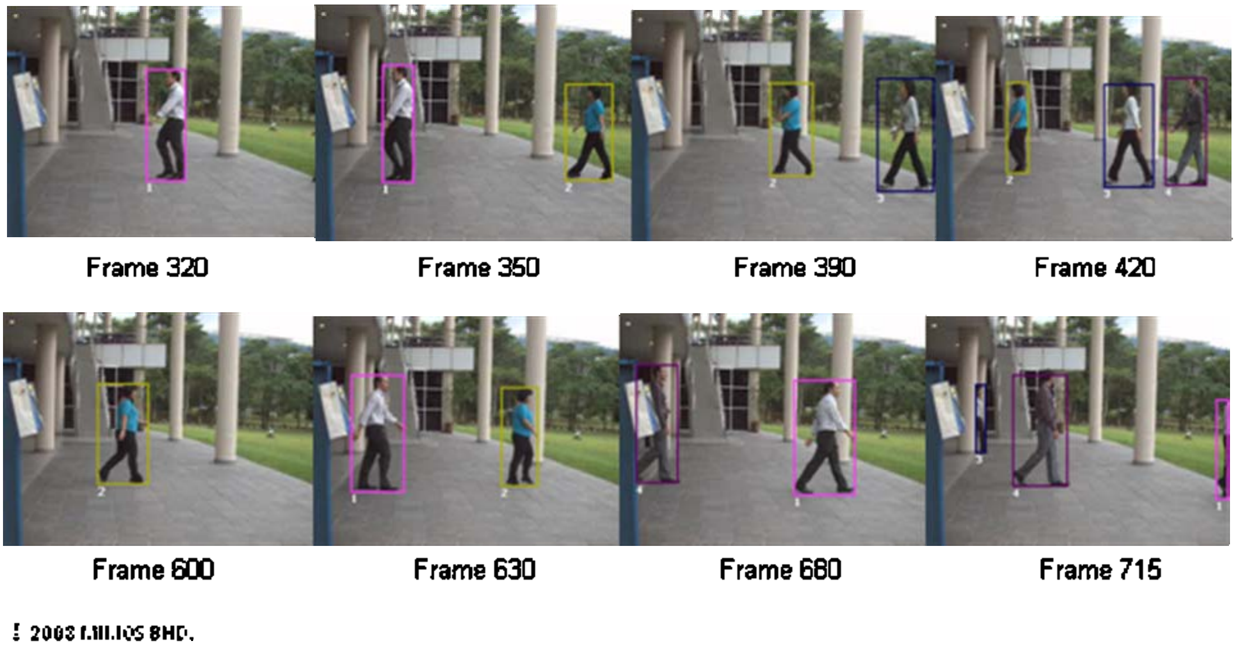

Fig. 8: The experimental results. Frame 320 to frame 420 demonstrates the new identifier assigned for new object appeared in the scene and the consistency of the identifier. Frame 600 to frame 715 illustrates the capability of dispense same identifier to similar objects which had left scene (discontinuity observation) less than a certain preset number of frames.

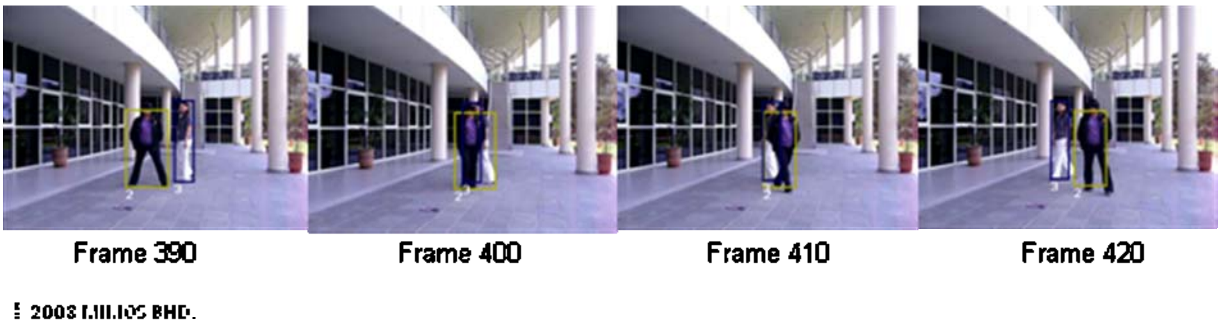

Fig. 9: Occlusion of objects 2 and 3. Each of the object in the group (2 persons) is identified which can be observe in frame 390 to frame 420. Frame 390 to frame 400 performed individual merge into group while frame 410 to frame 420 showed the individual split from the group.

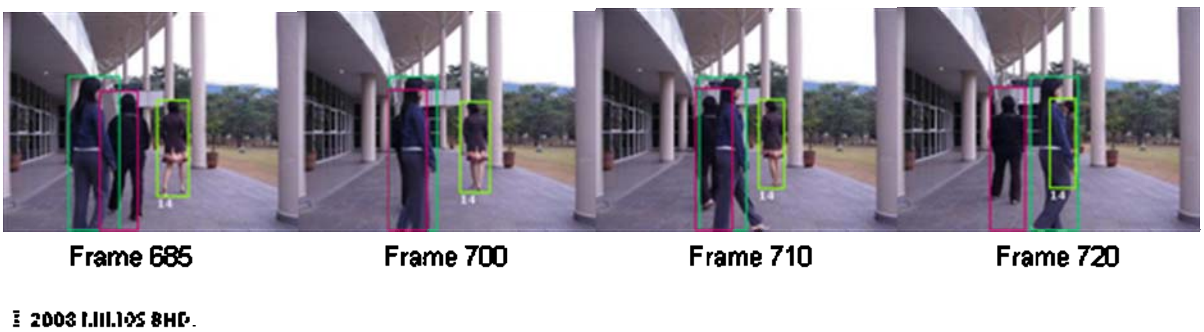

Fig. 10: Occlusion of 3 objects. Each of the object in the group is identified which can be observe in frame 685 to frame 720. Frame 685 to frame 700 performed individual merge into group while frame 710 to frame 720 showed the individual split from the group.