A Smart Meeting Management System With Video Based Seat Detection

Lijun Cao
Center for Research on
Intelligent Perception and
Computing (CRIPC), National
Laboratory of Pattern
Recognition (NLPR), Institute
of Automation, Chinese
Academy of Sciences (CASIA)
lijun.cao@nlpr.ia.ac.cn

Weihua Chen
Center for Research on
Intelligent Perception and
Computing (CRIPC), National
Laboratory of Pattern
Recognition (NLPR), Institute
of Automation, Chinese
Academy of Sciences (CASIA)
weihua.chen@nlpr.ia.ac.cn

Xu Zhang
Center for Research on
Intelligent Perception and
Computing (CRIPC), National
Laboratory of Pattern
Recognition (NLPR), Institute
of Automation, Chinese
Academy of Sciences (CASIA)
xu.zhang@nlpr.ia.ac.cn

Kaiqi Huang
Center for Research on
Intelligent Perception and
Computing (CRIPC), National
Laboratory of Pattern
Recognition (NLPR), Institute
of Automation, Chinese
Academy of Sciences (CASIA)
kqhuang@nlpr.ia.ac.cn

ABSTRACT

Nowadays intelligent meeting management is attracting more and more attention. Compared with most existing smart meeting systems which care more about recording talks during meeting in small meeting room, this paper describes an intelligent meeting management system for large auditorium. A cascade empty seat detection algorithm is embedded in the system to acquire the statuses of each seat. Empty seats are labeled step by step according to the extent of occlusions. Combined with attendee information input in advance, the system could provide people with the seated condition and corresponding information of attendees during the meeting. Experimental results demonstrate that the precision of seat's statuses detection could achieve 99.8%.

Categories and Subject Descriptors

H.4 [**Image Processing and Computer Vision**]: Miscellaneous

General Terms

Algorithms

Keywords

Smart meeting, video analysis, seat detection

1. INTRODUCTION

Meeting is one of the most common activities for the purpose of exchanging information and communicating. While intelligent management of meeting is very important for improving the efficiency of meeting management and building a safe meeting environment. The target of intelligent management of meeting is to understand what has happened, what is happening and what will happen in the meeting room. Thus video analysis, which involves object detection, tracking, action analysis and so on, plays a huge role in the intelligent management of meeting.

By now, a lot of concepts for smart meeting have been proposed, [1, 2, 3, 4]. Video analysis is embedded in many of them to understanding the contents of the scenarios. Detection and tracking methods is used to analysize the attendees' behavior [5, 6] to understand what's happening in a small meeting room. However, in a large auditorium, people always sit on their seats through the whole meeting. There's rare interactions happened. In indoor environment, occlusions and shadows are two main problems for both detection and tracking. To solve that, some previous works use 3D information [7, 8], which is insensitive to shadows and other changes in illumination. Moreover, 3D based methods require strong support of hardware, especially when the number of attendees are large, which cause these methods hardly meet the real-time requirement. Except for 3D methods, Dai *et al.*[9] use motion, color, gradient, raw gray features as visual cues to form a boosted cascade [10] for people detection. They suppose that motion is a significant clue of human presence, which makes it very sensitive to illumination change and other disturbance. In [3], human face

detection is directly involved. However, since face recognition is necessary in this process, the angle of face is crucial. Even though this problem can be solved by constructing a training face database containing certain face angles, face orientation still needs to be estimated during the meeting.
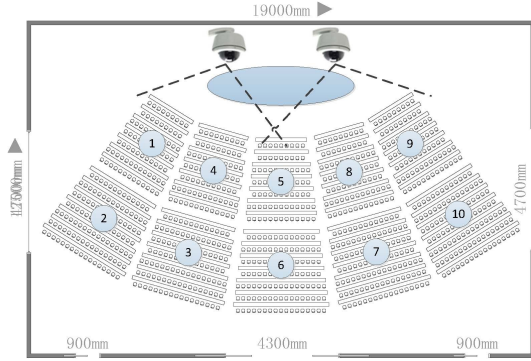

Figure 1: The auditorium, the area is about 30m*30m.

However, these systems care more about people's interactions during meeting, and there are many limitations. Such as, [1] the meeting room is required to be small, so the systems can capture the details of attendees' behaviors through high quality images; [2] the number of attendees should be small enough to keep the system free from sever disturbances and occlusions; [3] the low attendee mobility is necessary to guarantee the effectiveness of systems; [4] these systems emphasize analysis of individuals, a large amount of cameras should be installed to make sure every person has a good view in full field. Thus they are only suitable for small meeting rooms with no more than ten attendees involved.

In a large auditorium, like Figure 1, which could hold thousands of people, it is more important to record the attendees' statuses than their interactions. This paper focuses on the overall situation of the auditorium instead of individuals' behaviors. The situation in the auditorium can be reflected by human presence during the meeting. Considering the accordance between seats and attendees, the problem can be treated as a problem of seat detection. Knowing the attendees' statuses in seats is a direct way to control and predict the progress of the meeting. This will help people take appropriate actions on the meeting management.

In a large auditorium, the task of seat detection has many specific problems. The first one is how to keep a good balance between image resolution and the number of cameras, while the area of the auditorium is big and the number of people is large. Secondly, seat detection is subject to serious noises. The feature of seats should be robust to illumination change, shadows of people, and light spots. Third, external disturbance causes a great interference on results of seat detection. For the problems above, this paper provides some solutions. The video capture device are PTZ cameras, which can switch between several presets. To handle the serious noises, a novel detection algorithm with three steps is proposed. At last, there is a calibration process to deal with all the statistics of all seat information, this process can correct the false results caused by temporary changes of seat situation.

This paper mainly supplies two kinds of support for build-

ing a smart meeting: 1) PTZ cameras are used to collect images in a large auditorium; 2) a three-step detection method is proposed for seat detection to handle occlusion. The outline of the paper is as follows. Firstly the seat detection algorithm is detailed in Section 2. Then experimental results and applications implemented in a real auditorium are shown in Section 3, which show the effectiveness of the proposed seat detection algorithm. At last the work is concluded and future work is given.

2. SEAT DETECTION

Figure 2: Samples of three kinds of seat. From top row to bottom row, the seats are completely empty, occluded empty and non-empty respectively.

In this section, the details of seat detection algorithm are introduced. The task of seat detection can be described as the problem whether the seats are taken or not. However, during a meeting, some seats will be occluded by serious noises, like bags or clothes. In this case, it's hard to detection them as empty seats. To handle this problem, a three-step detection is proposed. It's based on a boosting cascade. The basic idea is that appearances of seats vary a lot according to the levels and positions of occlusion. In fact, all the empty seats show high consistence in color, shape and texture. Even though empty seats are subject to occlusions, occlusions often occur on the side of seats, while the top part of occluded empty seats remain original feature. Once a seat is taken by an attendee, the seat lose almost all of the original appearance in the image. Thus, we divide seats into three type: completely empty ones, occluded empty ones and non-empty ones, as shown in figure 2.

The three steps are coarse retrieving process, fine retrieving process and results calibration. The framework of the proposed boosting cascade of retrieving empty seats is shown in figure 3. Because of high consistence in appearance, the completely empty seats are firstly retrieved to speed up the whole process, which is a coarse retrieving process. Then the occluded empty seats and non-empty seats are analyzed with fine features to make sure the precision of system. That is a fine retrieving process.

2.1 Coarse retrieving process

The coarse retrieving process aims at distinguishing completely empty seats with occluded empty ones and non-empty ones. Considering that all the empty seats show high consistence in appearance, when a seat is taken, its appearance changes greatly. As a result, the substraction between background and foreground increases rapidly, which can be used to obtain the seat status. Firstly background model is constructed. For simplicity, a video captured when there is no people in the auditorium could be used as background
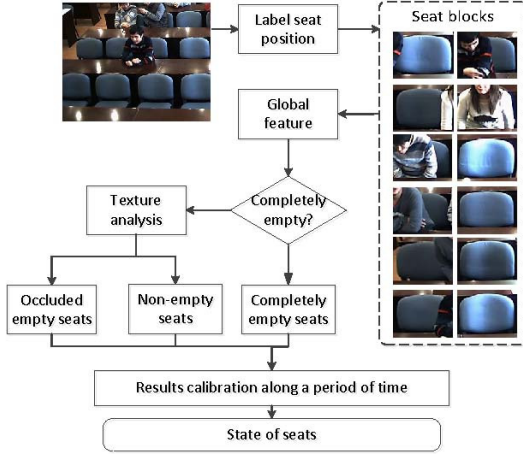
Figure 3: The framework of the proposed cascade method for seat detection.

model. The foreground areas are directly obtained for real-time video. Then background subtraction is employed. To be partly robust to illumination variation, background subtraction is implemented in HSV color space and gradient map. The feature of each seat can be described as:

$$f_i^{hsv} = \|f_{bg}^{hsv} - f_{fg}^{hsv}\|_1 \\ f_i^{grd} = \|f_{bg}^{grd} - f_{fg}^{grd}\|_1 \tag{1}$$

Finally the seat status S_i is obtained by following:

$$S_i = \begin{cases} 0 & \text{if } f_i^{hsv} + f_i^{grd} > \theta \\ 1 & \text{else} \end{cases} \tag{2}$$

where θ is a threshold for status S_i.

2.2 Fine retrieving process

The fine retrieving process aims at distinguishing occluded empty seats and non-empty seats. This is achieved by fine texture and contour analysis of seats.

HOG feature is proved to be effective in many applications, like object detection. Considering that position of occlusion differs when the seat is occluded and occupied by people, a simplified HOG feature(S-HOG) is proposed [11]. The process of extracting S-HOG is described in figure 4.
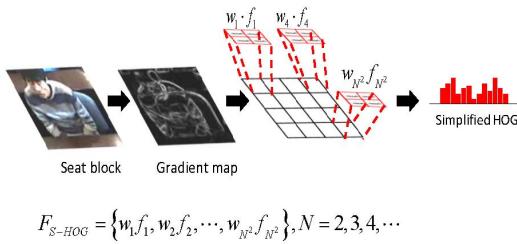

$$F_{S-HOG} = \left\{ w_1 f_1, w_2 f_2, \cdots, w_{N^2} f_{N^2} \right\}, N = 2, 3, 4, \cdots$$

Figure 4: The process of extracting S-HOG feature.

The gradient map is firstly calculated for each seat block, and then divided into cells with 8*8 pixels. For each cell a histogram of oriented gradient is computed as done in the process of extracting HOG feature. Each cell of the seat block represents different parts of the seat. When the seat

is occluded by people or clothes, the contour of seats' appearance changes. According to the consistence of people's posture, some cells of a seat are prone to be occluded while some are not. Therefore, a weight is assigned to each cell to represent the influence of occlusion on distinguishing the statuses of seats. The weights of each cell can be learned in the the following way: The HOG features of all cells are extracted to classify the occluded empty seats and non-empty seats. Different precision of different cell classifiers reflect the influence of occlusion in the according position, and is treated as the weight of the cell. Thus the S-HOG feature can be described as

$$f_{S-HOG} = \{w_1 f_1, w_2 f_2, \cdots, w_{M*N} f_{M*N}\} \tag{3}$$

where M, N is the number of cells in x, y direction, f_i represent HOG feature of cells, and w_i mean weights of cells respectively.

By default we use a linear SVM to trained the classifier for detection.

2.3 Results calibration by time

Through the coarse-to-fine empty seat retrieving process, each seat in the auditorium is assigned a label to represent its status. However, due to external disturbance there may be some errors. These errors are caused by some temporary interferences. For example, people walk across empty seats or people pass by an empty seat. Therefore, results calibration aims at correcting temporary false judgement of the seat. In this paper we adopt median filter of detection results in a period of time for calibration.

3. EXPERIMENTAL RESULTS AND APPLICATIONS

3.1 PTZ camera setting

Considering that the auditorium could hold thousands of people and the area of of the scene is large, camera setting should make sure that all the seats in the auditorium are covered and acquired images are of high resolution while as few as cameras are used. In the work of [12, 13, 14], they adopt omni-direction camera which could cover 360° to achieve high coverage. The problem of this scheme is that occlusions become more severe as the distance between the seat and the camera raises. So it results in difficulties in the task of seat detection.

Generally, the best camera angle to avoid occlusion is to install cameras above those seats. As long as the distance between the camera and seats are long enough, deformation of seats and the occlusion problem can be overlook. We divide the auditorium into several regions and then set up 5-6 presets for each PTZ camera, as shown in figure 1. Each preset is in charge of one region.

3.2 Seat detection results

Seat detection algorithm is used in a large auditorium which could hold more than a thousand people. Seat blocks extracted from two meeting areas collected as training database, which contains 1982 empty seat blocks and 2946 non-empty seat blocks. Video of another meeting about one hour is captured as test database, 77 classical images are collected from this database. The ground truth of a random region in the auditorium is labeled to validate the effectiveness of the
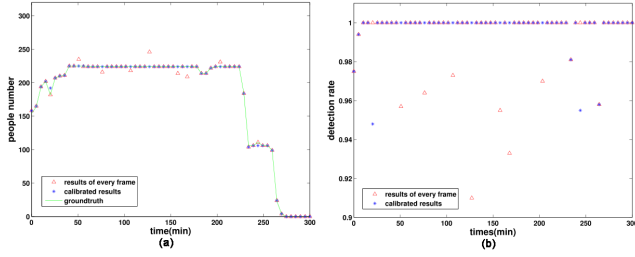
Figure 5: Results along time: (a) number of people (b) detection precision.

proposed algorithm. The region covers 64 seats and each seat block is 80 * 100 pixels in average.

In this paper the recall ratio of empty seats is adopted to evaluate the proposed detection algorithm. Considering that the coarse retrieving process and fine retrieving process aims at retrieving different type of empty seats, recall ratio of completely empty seats and occluded empty seats are obtained respectively for these two processes. Of all the 77*64=4928 seat blocks, there are 1259 completely empty seats, 735 occluded empty seats and 2934 non-empty seats. The results of seat detection can be shown in table 1.

Table 1: Detection results in coarse and fine retrieving processes.

	empty	occluded empty	false positive
coarse	100%	39.3%	0
fine	0	58.8%	0.38%

The coarse retrieving process aims at retrieving the completely empty seats. In this process, background subtraction is implemented in HSV color space to make sure the feature is robust to illumination change. And for some people whose clothes are the same color with seats, background subtraction of gradient map is used to capture the change of contour for seat blocks. Thus the coarse retrieving process recall nearly all the completely empty seats, as the results in table 1. For the remain seats, the fine retrieving process is employed. The results are also satisfied in the recall of occluded empty seats. For few errors after these two processes, a calibration module is employed to filter disturbance. Experimental results show that before the calibration the precision of seat detection reaches 99.73%, and after calibration the precision reaches 99.87%. The final results of the system can be shown in figure 5 (a) and (b).

3.3 Application of Seat Detection

A system based on seat detection is constructed for smart meeting. The system, combined with attendees information input in advance, provides people with the overall situation of the auditorium during meeting. The system could satisfy people's demanding for intelligent management in two aspects: information display automatically and quick query. Since that the management of auditorium involves management of all the attendees, both overall situation of the auditorium and details of each seat should be considered.
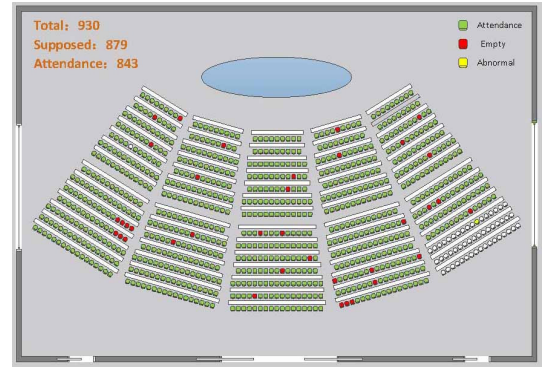
3.3.1 Global situation of the auditorium

Figure 6: Electronic map of the auditorium, in which different kind of seats are denoted with different colors. It could display overall situation of the auditorium in realtime.

To cover the large auditorium, the system uses several PTZ cameras to scanning every region in the auditorium. Then we combine regions together to get the over situation of the auditorium. As shown in figure 6, electronic map is formed as a global description. In the electronic map, each seat of the auditorium is described as a rectangle, and the statuses of the seats are represented by different colors. Once the seats are taken by new attendees or absences, colors of the according seats are changed, thus the overall situation of the auditorium is reflected directly.

As times go on, the history data of electronic map is saved. Electronic map of any time can be queried. Meanwhile, the statistic in a period of time are formed for the convenience of looking up the change of overall situation. Query of a particular seat or attendee results quick locating on the exact position of the auditorium, as shown in figure 7.
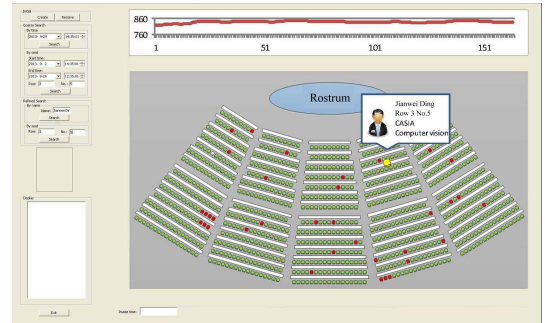

Figure 7: Query interface of the system.

3.3.2 Local situation of the auditorium

Local situation display is to see the details of regions in the auditorium, as shown in figure 8. The information of each attendee in the view could be acquired, including name, company, time of being present and so on. The present history of each attendee and each seat are saved to form statistics.

4. CONCLUSION

In this paper, a real-time intelligent meeting management system focusing on seats' statuses is proposed to monitor the
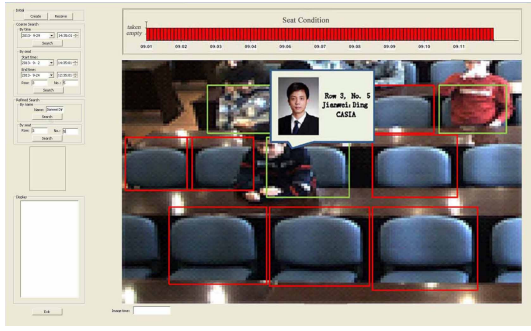
Figure 8: Local query interface of the system.

large auditorium and avoids shortage from other behavior-analysis systems. It adopts PTZ cameras instead of traditional ones, which could cover the whole auditorium with less costs. In detection, through a coarse-to-fine empty seat retrieving method, the status of each seat can be acquired accurately without interference of occlusion and serious noises. Combined with the accordance of seats and attendees, the system could provide users with the condition of each seat and each attendance during the whole meeting. Experimental results on a real meeting validate the effectiveness of the proposed seat detection system.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Z. Yu and Y. Nakamura, "Smart meeting systems: A survey of state-of-the-art and open issues," *ACM Computing Surveys (CSUR)*, vol. 42, no. 2, p. 8, 2010.

[2] M. Bett, R. Gross, H. Yu, X. Zhu, Y. Pan, J. Yang, and A. Waibel, "Multimodal meeting tracker," in *Proceedings of RIAO*, 2000.

[3] I. Mikic, K. Huang, and M. Trivedi, "Activity monitoring and summarization for an intelligent meeting room," in *Human Motion, 2000. Proceedings. Workshop on*, pp. 107–112, IEEE, 2000.

[4] D.-S. Lee, B. Erol, J. Graham, J. J. Hull, and N. Murata, "Portable meeting recorder," in *Proceedings of the tenth ACM international conference on Multimedia*, pp. 493–502, ACM, 2002.

[5] P. Dai, L. Tao, and G. Xu, "Audio-visual fused online context analysis toward smart meeting room," in *Ubiquitous Intelligence and Computing*, pp. 868–877, Springer, 2007.

[6] K. S. Huang and M. M. Trivedi, "Video arrays for real-time tracking of person, head, and face in an intelligent room," *Machine vision and applications*, vol. 14, no. 2, pp. 103–111, 2003.

[7] T. Svoboda, H. Hug, and L. Van Gool, "Viroom̦low cost synchronized multicamera system and its self-calibration," in *Pattern Recognition*, pp. 515–522, Springer, 2002.

[8] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer, "Multi-camera multi-person tracking for easyliving," in *Visual Surveillance, 2000. Proceedings. Third IEEE International Workshop on*, pp. 3–10, IEEE, 2000.

[9] P. Dai, L. Tao, and G. Xu, "Dynamic context driven human detection and tracking in meeting scenarios," in *Proc. 2nd Intl. Conf. on Computer Vision Theory and Applications*, 2007.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–511, IEEE, 2001.

[11] H. Liang, J. Wu, and K. Huang, "People in seats counting via seat detection for meeting surveillance," in *Pattern Recognition*, pp. 202–210, Springer, 2012.

[12] Z. Yu, M. Ozeki, Y. Fujii, and Y. Nakamura, "Towards smart meeting: enabling technologies and a real-world application," in *Proceedings of the 9th international conference on Multimodal interfaces*, pp. 86–93, ACM, 2007.

[13] J. S. Garofolo, C. D. Laprun, M. Michel, V. M. Stanford, and E. Tabassi, "The nist meeting room pilot corpus," in *Proc. of Language Resource and Evaluation Conference*, 2004.

[14] C. Busso, S. Hernanz, C.-W. Chu, S.-i. Kwon, S. Lee, P. G. Georgiou, I. Cohen, and S. Narayanan, "Smart room: participant and speaker localization and identification," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 2, pp. ii–1117, IEEE, 2005.