This paper proposes a method for automatic recording of audience during the lectures based on the results from conventional studies of teaching process improvements. In the presented approach we propose the audience analysis features which are later used for proper camera settings in the lecture hall with the multiple cameras installed. We focus on the analysis of student activities during the lectures in order to select the regions with different features representing the cues related to the estimated lecture attendance levels. The experimental results of the application of the proposed system together with user comments are presented in order to evaluate the validity of our approach.

Keywords: automatic video recording, lecture archives, multiple cameras, observing students, faculty development

1. Preface

Recently the e-learning methods are introduced increasingly into education and automatic video recording for distance lectures or lecture archives is being more extensively studied [1–5]. Most of these studies aim to acquire videos for students at distant lecture and for users of the lecture archive to understand the content efficiently.

In addition to the acquisition of videos for students, the use of automatic lecture recording system includes the possible application in order to improve lectures based on later review of archives by teachers to improve their teaching techniques. Lecture observation for teacher’s education, such as faculty development (FD), very often includes recording of teacher-students interaction during the classes at universities. The lectures are observed to find possible improvements by analyzing and evaluating the activity of students in relationship to the learning process [6–8]. Many approaches in FD to improve the lectures propose the enhancement of student’s attention directly or suggest changes in teaching strategies. Our study concerns the improvements of lectures based on the observation of videos captured during the classes by showing interaction between teachers and students. Our approach helps teachers to find weak points in their performances, which they usually do not notice during their lectures. We propose the utilization of automatic videos archiving system for such purpose.

The implementation of our approach focuses on acquiring of proper student videos as reference to educate the teachers. Conventional studies on lecture observation have shown that such approach is useful to improve the lectures by analyzing the students’ responses to presented content by the teachers. This study utilizes the two following features related to the efficient lectures according to FD studies, which we implement in video capturing system. The first feature is based on the common finding that an experienced teacher can determine how well a lecture is understood and whether students pay attention by correlating the following two sets of audience cues. The first set includes active listening, note-taking and inattention, and the second one includes student feedback and satisfaction level. The observation of audience reactions is useful to improve the content and teaching techniques [6]. Automatic lecture archiving enables the teacher later review of student attention at different points during the lecture. We defined this feature as the audience response range. The second feature is based on the notion that good teachers proceed while observing the audience at different levels [9] such as individual, groups, and the audience as a whole. The captured videos must provide such information, and thus should use different focuses to catch these levels. We defined this feature as the focus on the audience.

The main goal of this research is the video acquiring of students using automatic lecture recording with multiple...
cameras installed in a lecture hall in order to capture the audience response range and the focus on the audience.

The next section discusses the video features for evaluation of the audience response range. Based on this discussion, Section 3 proposes how to select and capture the focus on the audience in videos. Section 4 reports the results of experiments using our technique and Section 5 presents conclusions and future work guidelines.

2. Video Features

2.1. Estimation of Student Location

To record videos for teaching improvement purposes, as described in Section 1, the audience location information, (e.g. where students sit in a lecture hall) is necessary. Conventional automatic recording applications tend to focus mainly on the teacher. The students are recorded only sometimes, as a whole audience using a camera in front of them [4, 10]. Such approach may disturb the view of individual students in an image. This method does not provide the optimal solution to estimate the focus on the audience.

We propose the use of the camera with a fish-eye lens mounted on the ceiling in the lecture hall in order to estimate the focus on the audience robustly and efficiently. The students’ seats are usually arranged in a graduated auditorium style (Fig.1). This enables the audience to view and listen to the teacher clearly. The ceiling mounted camera with the fish-eye lens can thus capture the videos without occlusion. The example is shown in Fig.2.

Figure 3 shows a rectangular area called a cell which corresponds to each student’s seat, enabling attendees to be grasped at a glance (Fig.2). The following discussion assumes that students’ seats are arranged in a horizontal 2-dimensional (2D) grid. The x and y axes of 2D coordinates are regarded as rows and columns, e.g. a seat at row $x_i$ and column $y_i$ is represented by $(x_i, y_i)$. The cell corresponding to $(x_i, y_i)$ in the camera image is represented by $C(x_i, y_i)$. $x_i$ and $y_i$ are positive integers including zero. Even if the seats are not arranged in such a way, the same discussion applies if each seating position can be described with adjacent relationship features.

For a student occupying the seat $(x_i, y_i)$, we introduce a binary variable $s(x_i, y_i)$, which is true or false. This variable is used to represent the student’s location. The binary value of the above variable is determined based on the background differences $\Delta b(x_i, y_i)$ inside the cell $C(x_i, y_i)$. $\Delta b(x_i, y_i)$ is obtained by dividing the number $n_b(x_i, y_i)$ of pixels exceeding the threshold of which an absolute value is in $C(x_i, y_i)$ by the total number $n(x_i, y_i)$ of pixels in $C(x_i, y_i)$. If this value exceeds threshold $T_b$, a student is present in cell $C(x_i, y_i)$, otherwise the location is marked as empty.

\[
s(x_i, y_i) = \begin{cases} 
\text{true} & \text{for } \Delta b(x_i, y_i) \geq T_b \\
\text{false} & \text{otherwise}
\end{cases}
\]  

(1)
2.2. Cluster and Scale

Consider a cluster of students including one or more of them, which becomes the focus on the audience based on the location obtained in Section 2.1. The number of students included in this cluster is defined as a scale in our approach. To find the cluster we search for the adjacency of seats which are occupied by students. The chessboard distance (near-eight-distance) \( d(x_i, y_i, x_j, y_j) \) is shown in the following expression in order to evaluate adjacency between seat locations \((x_i, y_i)\) and \((x_j, y_j)\).

\[
d(x_i, y_i, x_j, y_j) = \max(x_j - x_i, y_j - y_i) \ 
\]  

(2)

The cluster of students is defined based on the distance \( d \) between each student’s location hierarchically. First, we define a set \( G_1 \) which contains the cells occupied by students as follows (see Fig.4 with single student example):

\[
G_1 = \{(x_i, y_i)|s(x_i, y_i) = \text{true}\} \ 
\]

(3)

The set \( G_1 \) contains \( l \) elements \( g_1^l \). Next, we compose the set \( G_2 \) which contains the clusters of the neighboring student pairs. The grouping procedure continues up to set \( G_k \) with the constrain of finding the set containing the largest clusters of neighboring students which distance \( d \) to the previous cluster \( g_{k-1}^l \) is expressed as follows:

\[
\min_{(x,y) \in g_{k-1}^l} d(x_i, y_i, x_j, y_j) = 1 \ 
\]

(4)

Finally, the set \( G \) is composed from the subsets \( G_1 \) to \( G_k \) as follows:

\[
G = G_1 \cup G_2 \cup \ldots \cup G_k \ 
\]

(5)

A cluster grouping procedure example is shown in Fig.4. The cells occupied by students are depicted as shaded circles with letters A to F. In this example, there are 12 clusters all together which composed the subsets \( G_1 \) to \( G_4 \).

The above cluster grouping procedure enables later proper camera zoom setting since all groups of neighboring students are marked properly because they are adjacent in the clusters. The number of students in the cluster represents the scale captured in the recorded video.

2.3. Size of Cluster Movement

It is necessary to acquire information on the activity of the individual students when acquiring the video for lecture improvement. The amount of movement derived from differences between the consecutive video frames is used as one of the features of human activity in motion recognition methods. Our approach also incorporates such technique and evaluates the activity for each cluster using inter-frame differences. Inter-frame differences for a single cell \( C(x_i, y_i) \) are represented by \( \Delta_f(x_i, y_i) \). \( \Delta_f(x_i, y_i) \) is obtained by dividing the absolute value of the difference between the number of pixels \( n_f(x_i, y_i) \) exceeding the threshold in consecutive frames by the total number of pixels \( h(x_i, y_i) \) in the cell. The value \( \Delta_f(x_i, y_i) \) is used to calculate the average activity \( D(g_k^l) \) inside of each cluster \( g_k^l \) obtained in Section 2.2. The average cluster activity \( D(g_k^l) \) is obtained as follows:

\[
D(g_k^l) = \frac{1}{k} \sum_{(x,y) \in g_k^l} \Delta_f(x_i, y_i) \ 
\]

(6)

3. Selection of a Region

3.1. The Focus on the Audience and Audience Response Range

We consider the acquisition of videos with the focus on the audience and audience response range using the multiple cameras installed in a lecture hall incorporating the methods described in Section 2. The focus on the audience includes the concept that the attendees are captured a single individuals in some cases or group of students in certain students situations. It means that there is no selection bias about the number of students captured in video frame throughout the video recorded by each camera. To implement this recording concept, it is necessary to select a cluster, in which the distribution of students’ number in time is uniform. This enables to obtain the proper audience response range from video recorded throughout a lecture. This means recording in which multiple cameras capture as many possible audience responses as possible. The above assumption enables us to select the clusters as regions with the different average cluster activity recorded by multiple cameras at the same time.

To implement the cluster selection satisfying the above requirements, we propose the score function which de-
3.2. Score Function for Region Selection

3.2.1. Implementation of Movement Range

We order the clusters \( g'_l \) based on the average cluster activity \( D(g'_l) \). All clusters have activity values within an interval \([D_{\text{min}}, D_{\text{max}}]\) which represent a distribution range of the cluster activities in a lecture hall. To capture the variety of activities with such a distribution range without failure by multiple cameras, interval \([D_{\text{min}}, D_{\text{max}}]\) is equally spaced by the number of cameras in order to determine a score function for selecting regions. Let's assume that the number of cameras to be \( c \) and the score function for the \( m \)th camera which can capture the selected cluster \( g'_l \) is \( P_m(g'_l) \). This score function is represented by the following expression, where \( \alpha^m_m \) is a positive integer defined in Section 3.2.2:

\[
P_m(g'_l) = \begin{cases} 
\alpha^m_m, & \text{for } D(g'_l) \in [D_0, D_1] \\
0, & \text{otherwise}
\end{cases}
\]

\( (7) \)

where

\[
D_0 = D_{\text{min}} + \frac{m - 1}{c} (D_{\text{max}} - D_{\text{min}})
\]

\[
D_1 = D_{\text{min}} + \frac{m}{c} (D_{\text{max}} - D_{\text{min}})
\]

The example of ordering of the clusters according to the average cluster activity is presented in Fig.5. The horizontal axis indicates the cluster arrangement and the vertical one the average cluster activity values as described in Section 2.3. In this example, three cameras were used and the interval \([D_{\text{min}}, D_{\text{max}}]\) was divided into three ranges. Every camera installed in the lecture hall should select a cluster within defined above interval. In this example, the score function is determined so that cameras \#0, \#1, and \#2 may have a positive selection score for the 25 clusters within an interval of \([D_{\text{min}}, D_{\text{min}} + 1/3(D_{\text{max}} - D_{\text{min}})]\). The 24 clusters within an interval of \([D_{\text{min}} + 1/3(D_{\text{max}} - D_{\text{min}}), \text{ and the three clusters within an interval of } D_{\text{min}} + 2/3(D_{\text{max}} - D_{\text{min}})\] respectively. Such ordering causes each camera to capture the cluster with preselected amount of activity range.

3.2.2. Implementation of the Focus on the Audience

After selection of one cluster with average cluster activity within an interval allocated by the score function described above for each camera, the time distribution of the level of a cluster selected as a region must be uniform to satisfy the focus on the audience. In other words, the number of students in a recorded student cluster must be selected without selection bias throughout recording.

It is, thus, necessary to determine the score function \( P_m(g'_l) \) so that the selection of clusters becomes uniform inside the region possible to capture by each camera as described in Section 3.2.1.

In order to define this score function distribution, the clusters selected for the \( m \)th camera are classified based on the focus of the audience.

Figure 6 shows clusters allocated to a single camera and there are six clusters including one student, four clusters including two students and two clusters including three students. Assuming the number of degree of focus as \( \sigma(m) \) and the number of clusters with a focus of \( k \) as \( v_m(k) \), \( \alpha^m_m \) of expression (7) which is responsible for selection of camera for cluster is defined as follows:

\[
\alpha^m_m = \frac{1}{\sigma(m) \cdot v_m(k)}
\]

\( (8) \)

Figure 6 shows an example where for \( \sigma(m) = 3 \), \( v_1(1) = 6 \) and the score function value is \( P_m(g'_1) = 1/18 \), for \( v_2(2) = 4 \) the score function value is \( P_m(g'_2) = 1/12 \) and for \( v_3(3) = 2 \) the score function value is \( P_m(g'_3) = 1/6 \), respectively. The score for clusters with a focus of \( k = 1, 2, 3 \) is \( 1/3 \).

4. Experiments

The proper implementation of defined score function in the above sections let us properly capture the ongoing students’ activity during the lecture with regard to propose measures such as the focus on the audience and the audience response range. The experiments conducted in a real lecture hall at our university let us determine what kind of videos can be recorded for teaching improvement purposes.

Figure 7 shows the lecture hall used for these experiments. The layout of the room is the same as the drawing
Automatic Video Recording of Lecture’s Audience with Activity Analysis and Equalization of Scale for Students Observation

![Graph showing cluster activity and ID of clusters sorted by activity](image)

(a) number of students in a cluster
(b) number of clusters having the same number of students
(c) selection probability

**Fig. 6.** Example of implementing the focus on the audience

![Camera and lectern setup](image)

**Fig. 8.** Results of estimated student locations

![Graph showing activity of student cluster over time](image)

**Fig. 9.** Size of selected cluster movement

Fig. 7. Full view of lecture hall used for experiments

in Fig. 1.

Two cameras were installed there as shown in Fig. 8. The lecture recording time was set for 90 minutes which is a general lecture’s time interval.

4.1. Camera Operation

We applied the proposed method for the central seating area of the lecture hall in Fig. 8. The students were seating in locations indicated by shading in Fig. 8. The presence of seating students was detected using the proposed method as described in Section 2.1 and indicated on Fig. 8 with bold line squares. We found that the locations were detected almost accurately. Based on the proposed method, we selected shooting regions for each camera and later recorded them.

**Figure 9** shows the variations of the amount of average cluster activity during 20 minutes and about 30 minutes after the start of the lecture. The horizontal axis is the elapsed time (in minutes) from the start of the lecture, while the vertical axis is the amount of the average cluster activity. There are shown two plots of variations of the average cluster activity captured by two cameras. Camera #1 captured the clusters with students characterized by small activity values (solid line in Fig. 9). The camera #2 captured the higher activity clusters, respectively (dotted line). The above procedures shows that using our method we can capture the audience during the lecture according to preselected audience response ranges.

**Figure 10** shows the captured changes in the number of students recorded in videos in a selected cluster during 20 minutes of a lecture. The horizontal axis is the elapsed time (in minutes) from the start of the lecture and the vertical axis is the number of recorded students. Cameras
#1 (solid line) and #2 (dotted line) captured frames with a small and large number of students, respectively. The focus on the audience was also estimated.

Figures 11 and 12 present the locations’ history at which two cameras were setting their optic axes. The rectangulars enclosed within thick dotted lines correspond to the seats in Fig.8. Cameras recorded an area based on the location of the seat in which students were presumed to be present. However, our method does not deal with individuals equally, so there is selection bias during recording. There might be a few students who are infrequently recorded.

Figure 13 shows an example of a selected cluster. The time slots A, B, and C in Fig.13 correspond to the same time slots in Figs.9 and 10. The value indicated for each seating location is the activity of each student at that time. Since two cameras are used, two clusters are selected each time. Thus, two clusters are enclosed with solid and dotted lines to show the activity in each cluster. The cluster indicated by dotted line was selected by camera #1 and the cluster indicated by solid line was selected by camera #2.

The use of a relatively simple algorithm allows us the real-time automatic recording. Each camera selects a cluster as its region with frequency of a maximum 10Hz or more. The clusters are selected at intervals of about ten seconds during actual recording.

Figure 14 shows shots of cluster frames recorded by two cameras. For example, camera #1 captured a student apparently not attending the lecture actively at time slot A with little movement. Camera #2 captured more active student looking at a teacher or moving his arms. Videos recorded by each camera at different times captured different numbers of students and reflected the focus on the audience.

4.2. Evaluation for Teaching Improvement

As discussed in Section 1, the automatically recorded lectures and stored in lecture archives may be a valuable support for teachers to enhance their teaching skills. The teachers may access the archives to watch the videos and search for possible mistakes they do. We applied the proposed system to four lectures and determined whether teachers could evaluate their performances. Our intention was to check whether the videos acquired by the proposed system would be useful in automatic recording of lectures for teaching improvement.

For the evaluation of our system we chose four lectures and two videos were recorded for each. We asked the four professors responsible for these lectures to review the videos for 10 minutes, that is, one 5 minutes video segment starting 10 minutes after the beginning of the lecture and one 5 minutes video segment starting 60 minutes after the beginning of the lecture, together with the recorded audio. We obtained the following responses from the professors based on the following 6 points questionnaire:

- **Question 1.** Could you see in the video the facial expression of students which you didn’t see during the lecture?
- **Question 2.** Could you see the facial expressions and responses of students while explaining complex
### Table 1. Result of questionnaire

<table>
<thead>
<tr>
<th>Question</th>
<th>A Economics</th>
<th>B Physics</th>
<th>C Information</th>
<th>D Law</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.75</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4.00</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4.00</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4.75</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.75</td>
</tr>
</tbody>
</table>

A. 2003/10/15 11:01:06 (32 min after lecture started)

- **magnitude of cluster movement**: 0.000 for camera #1, 0.286 for camera #2

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
<th>0.028</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.028</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.059</td>
<td>0.540</td>
<td>0.054</td>
<td>0.000</td>
<td>0.014</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.271</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.160</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

B. 2003/10/15 11:05:04 (36 min after lecture started)

- **magnitude of cluster movement**: 0.038 for camera #1, 0.116 for camera #2

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
<th>0.015</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.015</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.000</td>
<td>0.195</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.047</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.116</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.034</td>
<td>0.000</td>
<td>0.085</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

C. 2003/10/15 11:10:59 (42 min after lecture started)

- **magnitude of cluster movement**: 0.000 for camera #1, 0.156 for camera #2

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
<th>0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Fig. 13. Example of selected clusters**

- Question 3. Could you observe all the students?
- Question 4. Could you observe the behavior of students in detail?
- Question 5. Do you think that proposed shootings algorithm let you capture individual students, groups and the whole audience comfortably?
- Question 6. Could you obtain information unavailable to you during your lecture?
- Evaluation 5. Absolutely

**Fig. 14. Recording based on selected cluster**

- Evaluation 4. Satisfactory
- Evaluation 3. Not sure
- Evaluation 2. Not satisfactory
- Evaluation 1. Absolutely not

The results of the questionnaire are shown in Table 1. The average value of 4 or higher was given to all questions. The answers for questions 2, 3, and 5 were evaluated as low. Regarding question 2, in this experiment we asked teachers to watch videos and listen to audio, but since the proposed method did not record students based
on their movement types, their facial expressions, activity and the presence of responses to presented material. For questions 3 and 5, "all students" were interpreted as the whole audience present in the lecture hall. Additionally, our method does not capture the whole audience all the time so the students' captured locations could be not uniform. This may cause that some students in the audience might be captured very seldom. These factors could cause the low evaluation.

We also obtained the free comments from the professors which were as follows: Teacher A did not perceive not attending students during his lecture. Teacher B could not remember particular students, but he could remember the audience as a whole. Our videos enabled him to observe the attitude and let him look at individuals or groups. Such videos, according to him, would be useful for his future lectures evaluation. Teacher C could gain a lot from presented videos since he could see some of the students whom he could not perceive during his actual lecture and additionally, he noticed some of the students listening to him eagerly without giving active responses. Teacher D had the impression that there were significant differences in attending a lecture depending on individuals.

We found from this description that each teacher has discovered new information that could be not perceived during a lecture based the analysis of the amount of the student activity or the degree of attention.

5. Conclusions

The intension of our research was to develop an automatic lecture recording system which could have to improve teaching skills such as used in FD. We focused on the practical implementation of the technique for acquiring the videos useful from the FD point of view. We propose the use of multiple cameras installed in a lecture hall in order to record the defined in this paper audience response range and the focus on the audience. Our approach is based on the conventional FD studies which involve the observational behavior analysis for teaching skills improving purposes. The proposed method captures the activity of students in different group sizes using multiple cameras and lets each camera to select regions covering the clusters of students with different sizes for further recording. Our approach was implemented with determination of the score function for selecting the regions for capturing, based on information about the location and activity of students obtained from the camera with a fish-eye lens installed on the ceiling in the lecture hall. Our method was applied to the actual lectures at university and it was confirmed that videos recorded according to the defined score function provided the audience reaction range and the focus on the audience satisfactory. It was also confirmed by questionnaires to teachers that they could gain new information about their own lecture styles.

Our future work will include an attempt to introduce a more detailed motion feature extraction from videos for evaluating the audience activity in addition to the proposed in this paper the average cluster activity. An attempt has been made to recognize activity from motion features. Incorporation of such more advance features should lead to more precise estimation of the focus on the audience and the audience response range. The image resolution problem will always remain for large lecture halls even with application of high resolution cameras. We will attempt to enhance the motion recognition methods from lower resolution videos. The another issue would be the consideration of additional constrains for choosing captured regions of the lecture hall. In our research we considered only the two constrains which are the audience response range and the focus on the audience, which according to FD findings are the most important features for teachers skills improvement purposes. We also consider the introduction of the method to record all the students equally or the method to select the cameras in order to record students from various directions to improve the quality of the final video. The above requirements could be implemented within our framework, but additional studies should be made to determine the details and usefulness for teaching skills improvement.

Acknowledgements

We thank to Associate Professor Shinichi Mizokami of the Higher Education Research and Development Center of Kyoto University and Research Associate Masayuki Murakami at Kyoto University of Foreign Studies for their useful comments during the process of writing this paper.

References:


188 Journal of Advanced Computational Intelligence and Intelligent Informatics Vol.8 No.2, 2004
Automatic Video Recording of Lecture's Audience with Activity Analysis and Equalization of Scale for Students Observation

Name: Satoshi Nishiguchi
Affiliation: Research Associate, Graduate School of Law, Kyoto University

Address: Yoshidainomatsucho, Sakyoku, Kyoto 606-8501, Japan
Brief Biographical History: 2000- Research Associate, Kyoto University
Membership in Learned Societies: • The Institute of Electronics, Information and Communication Engineers (IEICE)
• The Japanese Society for Artificial Intelligence (JSIAI)
• The Institute of Electrical Engineers of Japan (IEEE)

Name: Koh Kakusho
Affiliation: Associate Professor, Academic Center for Computing and Media Studies, Kyoto University

Address: Yoshidainomatsucho, Sakyoku, Kyoto 606-8501, Japan
Brief Biographical History: 1992- Visiting Scholar, Stanford University
1994- Research Associate, Osaka University
1997- Associate Professor, Kyoto University
Membership in Learned Societies: • The Institute of Electrical and Electronics Engineers (IEEE)
• Association for Computing Machinery (ACM)
• Information Processing Society of Japan (IPSJ)

Name: Yoshinari Kameda
Affiliation: Assistant professor, Institute of Engineering Mechanics and Systems, University of Tsukuba

Address: 3F306, University of Tsukuba, 1-1 Tennoudai, Tsukuba, Ibaraki 305-8573, Japan
Brief Biographical History: 1996-2003 Assistant professor, Kyoto University
2001-2002 Visiting Scientist, Massachusetts Institute of Technology
2003- Assistant professor, University of Tsukuba
Membership in Learned Societies: • The Institute of Electronics, Information and Communication Engineers (IEICE)
• Information Processing Society of Japan (IPSJ)
• The Institute of Electrical and Electronics Engineers (IEEE)

Name: Michihiko Minoh
Affiliation: Professor, Academic Center for Computing and Media Studies, Kyoto University

Address: Yoshidainomatsucho, Sakyoku, Kyoto 606-8501, Japan
Brief Biographical History: 1983- Research Associate, Kyoto University
1987-1988 Visiting Researcher, University of Massachusetts
1989- Associate Professor, Kyoto University
1995- Professor, Kyoto University
Membership in Learned Societies: • The Institute of Electrical and Electronics Engineers (IEEE)
• Association for Computing Machinery (ACM)
• The Institute of Electronics, Information and Communication Engineers (IEICE)