This is a preliminary version of an article published in
Proc. of the International Workshop on Educational Multimedia and Multimedia Education on ACM Multimedia (EMME), pp. 11-18, Augsburg, Germany, September 2007
by Fleming Lampi, Stephan Kopf, Manuel Benz, Wolfgang Effelsberg

An Automatic Cameraman in a Lecture Recording System

Fleming Lampi, Stephan Kopf, Manuel Benz, Wolfgang Effelsberg
Department of Computer Science IV, University of Mannheim
A5,6 68159 Mannheim, Germany
{ lampi | kopf | effelsberg }@informatik.uni-mannheim.de, mabenz@rumms.uni-mannheim.de

ABSTRACT
We describe the design and implementation of an automatic cameraman for lecture recording. A major problem with traditional lecture recordings is that they tend to be boring for the students, especially if only the slides and the audio of the lecturer are presented. In a first step, we determine the tasks a real cameraman would have, in particular with respect to liveliness of the video. We then adapt these tasks to a computer system and show in detail how they can be implemented. In a second step, we describe how our algorithms support the virtual director system into which the automatic cameraman is integrated. We conclude that lecture recordings can be much more lively and interesting using our approach.

Categories and Subject Descriptors
K.3.1 [**Computers and Education**]: Computer Uses in Education - *Distance Learning*

General Terms
Algorithms, Documentation, Design.

Keywords
Educational Multimedia Application, Camera Module, Automation, Camera-Team, Video Production Rules, Lecture Recording and Lecture Transmission.

1. INTRODUCTION
Lecture recordings are very common in the past because students can prepare for exams independent from time, they enrich and amend e-learning materials and they are easy to achieve [11]. But in many cases they easily become boring, completely independent how fascinating the original session was, by only recording the slides and the lecturers audio.

Television has pushed our expectations by the quality we watch everyday. Although students preparing for their exams are highly motivated, it would be really helpful to support their learning from recorded lectures by applying video production or cinematographic rules during the recording, e.g., how long must a shot stay at least, which order of long shots, medium shots and close-ups changes the meaning of a scene in which way, how to record a dialogue, etc. Such rules are used to make it easier for the spectator to follow up the thread and to keep focused.

Especially in times when universities have to save money it is far too expensive to hire a real camera team for lecture recordings. In some cases it is possible to use staff of the institute to replace a camera team, but even then it is very unlikely to get the same result as an experienced team concerning quality.

By implementing a real-time system it is possible not only to record but also to broadcast lectures and it can cooperate with interactive tools for lectures [18, 19].

In Section 2, we present related work concerning video recording from video surveillance over meeting recordings to other lecture recordings and make the differences to our approach clear. Section 3 describes the scenario of a live TV production concerning the necessary staff. Coming from the whole team we focus on the cameraman and his duties, reducing it to the tasks he has to perform during the recording. After this, in Section 4, these tasks are mapped and adapted to a computer system and the algorithms are presented. Their results are also shown there. Section 5 presents the conclusion and an outlook on future work.

2. RELATED WORK
Automated camera recording is used in several fields, some of which are similar to our work. A major research focus is on video surveillance [4]. Further research was done on indexing lecture recordings and recognizing the transitions between shots [13]. Those approaches are not directly related to our work. However, somewhat closer to our area are meeting recordings, often done by 360° camera and microphone sets [15, 2]. Even if these approaches are transferred to larger auditions, e.g., [6], the results seem more like video surveillance rather than lecture recordings, due to the position of the camera.

For recording presentations or lectures there are two general approaches. The first approach uses a high resolution, wide angle camera to record the whole scene and uses the image or a subset of it in standard resolution to visualize active parts of the presentation [17, 12]. Because this approach uses the zoom only it is not applicable to our goal to simulate a camera-team.

Much closer to our approach is the second way using pan and tilt operations and image processing for framing and following the lecturer. A sample application is “AutoAuditorium” [1], which shows a basic level of automatic presentation recording without any video production rules implemented. More advanced is the system developed by Microsoft Research [16] which has been improved meanwhile [22, 23]. It uses multiple cameras and implements basic video production rules. A video director module based on a finite state machine (FSM) is available which can be configured by a scripting language for cinematography rules.

In spite of these correlations to our approach it differs in many ways; it solely uses image processing to determine the image framing and to track the lecturer while we additionally use an indoor positioning system. We are able to identify the absolute positions of all tracked persons. Thereby, we are able to implement more sophisticated video production rules, e.g., two tracked persons may be framed in such a way that they face each other not only in reality but at least while switching between their shots. Such constellations were deferred by the other authors to further research, e.g., on eye gaze orientation detection. In particular, our implementation of the cinematography or video production rules differs. By using a scripting language, their rules are simply rewritten in a note form and therefore stand for fixed durations of the shots and predetermined transitions coming from the fixed weights for alternative transition targets given in the script. For the recording of real-time applications, similar basic rules have been proposed by [5].

3. SCENARIO

In live TV productions a huge staff is necessary to cover all parts, but for lecture recordings in universities it is neither possible nor useful to have such a large staff. For example there is no need for make-up artists or set constructors. Furthermore, caused by the relatively well known workflow of a lecture a camera team would be sufficient. That is the reason why we focus on the camera team in the following.

3.1 A Real Camera Team

A real camera team for live studio production still consists of several people. As an example here is a list of who may belong to the team, as used at Südwest Rundfunk (SWR), German television: director, editor, taped recordings operator, inserts operator, lighting cameraman, cameramen, lighting technicians, iris operator, audio engineer and final signal controller.

Although each single role is important for a show produced at a high quality level, there are cheaper productions which try to reduce the number of people in the team. From the viewpoint of a cameraman only the director, the editor, the lighting cameraman and all other cameramen are necessary to work as a team to produce a live show. In smaller productions the director and the editor may be the same person. If there is no iris operator, every cameraman has to adjust its iris on his own, which may lead to images using completely different exposures when comparing the different camera signals.

Staying at the viewpoint of a cameraman we will now have a closer look on the duties of a cameraman.

3.2 Duties of Cameramen

As we have seen above, a cameraman has to perfectly work in a team, which starts from the planning phase and leads up to post-production. The cameraman mainly steps in into the recording phase, which begins not exactly when the red light is on, but several minutes earlier. So the duties of cameramen are divided in three parts, a) before the show, b) during the show and c) during a shot.

3.2.1 Before the show

At first there is a meeting to review the storyboard, whereby the director goes through all details of the show and makes the important points clear to the lighting cameraman and all other cameramen. Figure 1 shows a part out of such a storyboard.

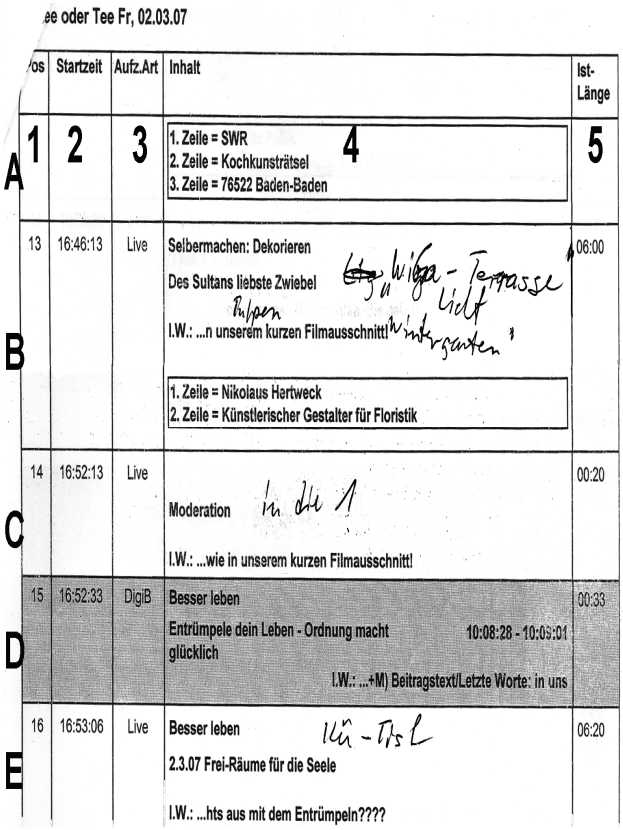

Figure 1. Storyboard example of a German TV show

We describe the content of Figure 1 in the following. Column 1 contains a sequential number, column 2 the starting time in hh:mm:ss notation. The third column describes the origin of the signal where "DigiB" stands for DigiBeta, a video recording format of Sony Corporation, and "Live" stands for live recording. In column 4, the content to broadcast of each step is described. The last column 5 contains the duration of its step in mm:ss notation. We describe the content of column 4 in detail now. It contains at first the title of the part. If its origin is a video tape the corresponding SMPTE time code is shown additionally as in row D. The text parts written in rectangles shown in rows A and B are used for the inserts to publish an address for a lottery in row A or to introduce a person in row B. The abbreviation "l.W.:" stands for "last words" and are the last words of this part. It is the signal for the director to switch the correct signal "on air". All handwritten annotations are additional information given in the meeting by the director to enable the cameramen to do their job. The cameraman gets his orders in three steps: Basic information out of the storyboard, additional information by the director in the meeting and at last live during the show using the intercom.

The next location is the studio. The position of each cameraman is crucial. As described by [20] the "line of action" must never be

crossed. Figure 2 shows a little example from top view to clarify how important this line is:

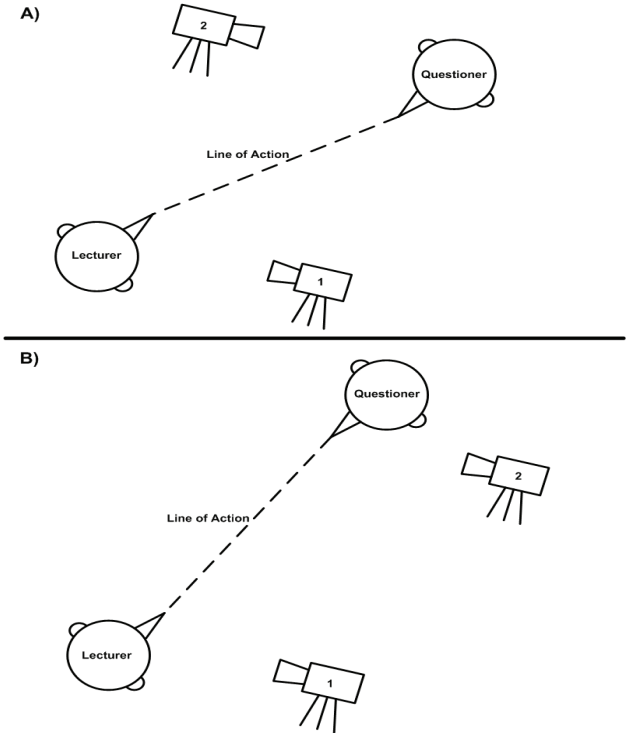

Figure 2. A) Mistake, B) Correct concerning "Line of Action"

We argue that the lecturer and the questioner in the audience are discussing. In case A), camera 1 will show the lecturer viewing from the left edge to the right and camera 2 will show the questioner viewing from the left edge to the right, too. If these shots are shown one after the other the effect will be very confusing to the spectator, because they are discussing but do not face each other. The reason is that both cameras are on different sides of the "line of action". In case B), camera 1 will still show the lecturer viewing from the left edge to the right but camera 2 shows the questioner viewing from the right edge to the left. So the spectator gets the impression that both are facing each other while discussing. Therefore the correct place of a cameraman is very important.

Now the cameraman set up his camera and makes it ready first for the dress rehearsal and second for the show.

3.2.2 During the show

As done in the rehearsal the show commences, during the whole show the cameramen use headsets to communicate with the editor in the central control room. There is only one intercom for all participants and everyone is able to speak at the same time. Therefore, it is necessary to be extremely disciplined so that everyone is able to understand the single person who is speaking.

Using this intercom the cameraman gets his orders from the editor and director. These orders include information about "who is on air", "who will be on air next" and "which detail or framing a certain cameraman should show". Sometimes the cameraman informs the control room, for example, if caused by some technical reason he is unable to fulfil a requested shot or if he has got an idea of an extraordinary detail or framing, which he wants to show.

Included in this conversation are commands like: "Camera 1 please frame person A in a way that he looks from the left edge into the image." Another cameraman gets the command for the counterpart: "Camera 3 please frame person D in a way that she looks from the right edge into the image." Now the editor is able to switch between these two shots as long as these two people are talking to each other. For the spectator in front of the TV-set it looks like those two people are facing each other while talking, even if there are hundreds of miles between them.

So throughout the show there is a continuous communication to optimize the aesthetic aspects of the recording.

3.2.3 During the shot

To get into more detail we will now have a closer look on the shot itself and on the work of the cameraman. This part focuses on the technical work of a cameraman and not so much on the artistic, aesthetic part.

At first, the cameraman has to bring the requested image into the sight of the camera by moving, panning and tilting. Next, in case there is no iris operator, the cameraman has to control the iris himself. He is constantly adjusting it to achieve an always similar exposed image, even if the illuminating varies from one part of the studio to another. It is very important that the cameraman focuses the main parts of the chosen image. By zooming in or out before or even while being on air the image gets its "final touch".

This complex process which needs a lot of experience for live production is repeated for every single shot during the show.

4. REALISATION

Let us think of a minimal camera team which allows us to record lectures in a reasonable way. At first there is the need of a cameraman – for each different point of view. We need one for a *long shot* as an overview of the scenario, one for the lecturer which is able to follow him and his gestures and one for the slides to record them clearly readable. At last we need one cameraman for the audience in case questions are posed. To coordinate all these cameramen a director is needed.

In addition there is the need to record the audio of the lecturer, some sounds from a computer, e.g., while playing simulations, and we need the audio of the questioner. Therefore, we need audio engineers to record the necessary audio.

Not belonging to a normal camera team but still needed for our lecture recordings are sensor tools which offer additional information to the director. These tools are, e.g., an indoor positioning system and a question manager. While the indoor positioning system supports the necessary coordinates of the targeted objects, the question manager recognizes requests to pose a question and manages them in a way the lecturer can reasonably handle them. Figure 3 shows an overview over the whole system as it is planned. The grey shaded parts a not yet implemented completely.

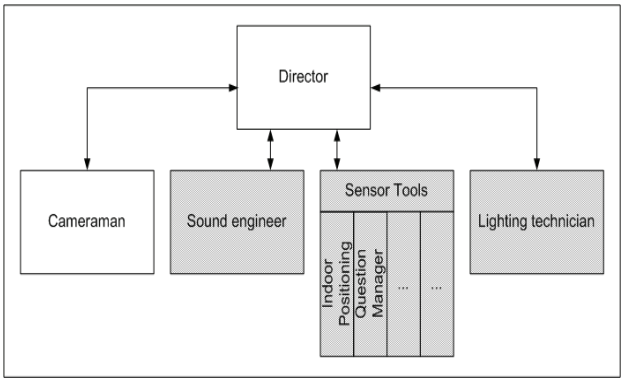

Figure 3. Overview over the Automated Lecture Recording System

The director module is already realized as it was presented in [9, 10]. A prototype of the indoor positioning system as our preliminary work shows that indoor positioning systems based on 802.11 offers a position accuracy that is suitable for automated lecture recording system [7]. Especially, if the scenario of a lecture hall is taken into account the position algorithm can be optimized to deliver even better results [8]. In this paper, we are focusing only on the cameraman.

4.1 Functions of a Cameraman

As we have seen in Section 3.2, a cameraman has a lot of duties throughout a show. In order to implement a reasonable subset of functions we have to identify which are definitely necessary and which may be deferred to a later stage of realisation. In the following, we step through the workflow of a cameraman as described above.

In the setup phase the cameraman has to be located to a certain point in the lecture hall. From there he is able to frame his starting view, calibrating his white-balance and adjusts the exposure of the camera. In addition the cameraman is told whom or what to show by the director.

During the show the cameraman will get new coordinates for the next shot, he will get information whether the targeted person should be located on a certain side of the image. If he encounters a reasonable amount of changes in the image, e.g. motion, the cameraman offers this image to the director.

For every single shot the cameraman has to repeatedly focus on the important part of an image, bringing it into the right position in the image and selecting the appropriate zoom. If lighting conditions have changed he also has to adjust the iris. At the end of a show the director often wants the cameraman to show a final shot, which may be identical to the starting view or a certain detail.

We have now reduced the complex functionality of a cameraman to a reasonable minimum, sufficient for lecture recordings. At least this minimum of functions has to be adapted to the computer system.

4.2 Adapt Camerawork for the Computer

Like a real cameraman we set up the virtual cameraman on a certain point, to get to know its coordinates. Again it is important not to cross “the line of action” while positioning the cameras. By setting these locations of the cameras carefully it is easy to define which camera should shift a person to the left edge and which camera should shift a person to the right edge, in case of a discussion or a dialogue.

Using this coordinate data the cameraman registers at the director module, via standard LAN, which assigns an ID and sends back the data of its starting view. This starting view is focussed on by panning and tilting and the cameraman calibrates the white-balance. Up to now, we are still using the presets of the camera for white-balance. The last step of the setup phase is the adjustment of the exposure.

During the show the tasks are repeated. The cameraman will always pan and tilt to the new coordinates, sets the zoom and the focus and adjusts the exposure. We argue that the work of a cameraman can be regarded as a kind of control-loop, even if in reality some of these functions can be done in parallel. Therefore it is possible to write the job of cameraman as flow chart, e.g., shown in figure 4.

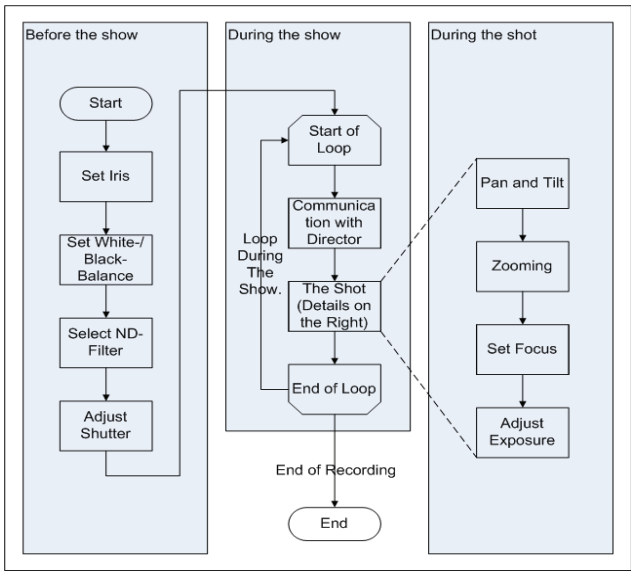

Figure 4. Camerawork as a flowchart

Coming from a complex bundle of tasks a cameraman has to do we reduced them to a reasonable minimum in the previous section and abstracted them as a flowchart. In a first instance we only need to implement the functionality of setting the iris, pan and tilt the camera to given coordinates and to adjust the zoom to a reasonable image as well as the communication between director and cameraman.

In the following section we focus on the algorithms we have implemented.

4.3 Algorithms

4.3.1 Before the Show

Before the show is started, the cameraman has to be initialized. Therefore, the configuration is read from an XML-file. It contains its identifier and IP address, the identifier and IP address of the director, the capabilities of the camera and parameters for the algorithms concerning the camera. Communication between cameraman and director is established by using the identifiers and IP

addresses. The capabilities of the camera are of special interest. They describe whether the camera is able to pan, tilt and zoom, whether it is able to adjust the shutter, grey filters or white-balance. In addition, they describe which protocol is used to control the camera, e.g., hypertext transfer protocol (HTTP) over a network connection or commands over a serial port communication. At last it contains some parameters necessary for the algorithms, e.g., number of rectangles to segment the image.

In the next step the cameraman registers himself at the director and the whole communication is done by XML based messages. This gives the advantage to transmit multiple parameters at once, which may be necessary if, e.g., a new position is transmitted and the cameraman is advised to put the target on a certain side of the image, due to his position and the "line of action". The director can now send position coordinates to the cameraman, which are interpreted taking its own position and its old viewing angle into account and transferred to the device coordinates of the camera. This does not apply if the camera is used for the slides and is therefore immovable.

To complete the initializing phase the automatic cameraman selects the correct iris value, sets the white balance and adjusts the shutter. For setting the white balance we are using the pre-set for indoor lighting up to now. The shutter is set to 60 Hz as a standard. We avoid any interference with TFT displays, because we record only the lecturer and the audience with these cameras. For screen recording as mentioned before we use the Camtasia recorder where no shutter has to be set.

4.3.2 During the Show

During the show the communication between cameraman and director is kept up so coordinates, framing information, etc. are sent from the director. The cameraman himself sends command acknowledgements and status information, e.g., still moving, finished focusing, etc., as well as change rate alerts to the director. These change rate alerts may have different origins.

Concerning the Camtasia recorder those alerts may stem from annotations or changes of the slides, so switching and annotating them can be identified easily. Concerning the other cameras the change rate alerts may stem from an active or moving lecturer or from an agitated audience. In each case the automatic director is able to rate the alerts taking the amount of change and its source into account. These values are cumulated into the calculation of the probabilistic values which shot to show next. All possible transitions from the active shot to the next shot get a fixed start value. By having access to a shot history, input from sensors and the feedback of the cameraman a transition gets its value more or less decreased. Finally the transition with the highest value is selected.

4.3.3 During the Shot

During the shot the cameraman uses the M-JPEG stream to capture single images for analysis because the frame rate we use can be up to 25 frames per second. Even fast changes can be analysed based on this frame rate. If the quality is more important than speed, it will use the image grabbing functionality which delivers a bitmap image. For each bitmap image it takes up to 500 milliseconds to deliver it, i.e., a frame rate of up to 2 frames per second is possible. The quality of the bitmap image is better than the JPEG image because no lossy compression algorithm is used. The main advantage of the video servers used with our cameras is that image grabbing of bitmaps, M-JPEG streams and MPEG-4 streams run simultaneously and independently. The director can use the MPEG-4 stream without any dependencies. Concerning the Camtasia recorder compromises have to be made. It provides only one video stream on which cameraman and director have to rely, so we are still testing if we need a unicast to multicast stream converter or whether two unicast streams will work well due to the small frame rate of about 7 frames per second.

4.3.4 Implementation

In our implementation we are using different camera models, an Axis camera (214PTZ), a Canon Camera (VC-C4) coupled with an Axis Video Server (241S) and a Techsmith Camtasia Recorder in streaming mode to capture the slides. The Axis technology has got the advantage to stream MPEG-4 streams simultaneously to M-JPEG streams both up to 25 or 30 frames per second respectively and is still able to provide bitmaps on demand. The Camtasia recorder on the other hand provides fewer frames per second which is absolutely sufficient for screen capturing and can be used as a camera device for many other applications. While the MPEG streams will be routed to the director, the M-JPEG streams and still images are used by the automatic cameraman to do his job.

We now focus on the implementation of the needed functionality, starting with the task of setting the iris. The automatic iris algorithm included in the camera is often not useful especially in a backlight situation.

Therefore, we developed our own algorithm taking some more aspects into account. At first we choose an iris value showing a maximum number of details in the image, based on edge detection and counting the number of detected edge pixels. This is a good measure, because details will get lost if the picture is too bright or too dark. In a next step, we take into account that we will record people in most cases. Therefore, we are searching for parts of the image containing skin colour, similar to the approaches of [3, 14, 21]. Now we are going to maximize the number of details only in the segments containing skin colour, neglecting all other parts by adjusting the iris value. As a result we achieve good perceptible persons in our recordings. Figure 5 shows the result of a standard automatic iris on the left and the result of our algorithm on the right. Using this technique can correct backlight or dawning light situations.

Figure 5. Comparison of Iris Algorithms: Automatic approach based on camera hardware (left) and automatic camera calibration (right).

The algorithms for detecting skin colour and counting pixels of edges are used to adjust the iris in the setup phase and during the shot to improve the image. The only difference is that during the

shot it is a concurrent process for fine adjustments while in the setup phase it is the main task, which exhaustively determines the minimal, maximal and optimal values of the iris. Segmenting the image into rectangles and selecting those segments which show most likely skin, combine it with the results of the edge detection algorithm, we can estimate which part of the image belongs to the background and which does not. By joining the selected segments to regions, we can assume which segments of an image represent a person. The regions showing a person are used to calculate a centre, which enables us to shift this person to the left or right border of the recorded image.

If more than one person is shown in the image, the centre is between them, which is sufficient for us. We have to use other techniques to decide which person is of interest to us. But this part is shifted to our future work at the moment.

At last we need an algorithm to react on gestures and movements of the person of interest. In such a case it depends whether the person is still completely visible in the image. We must react if parts, e.g., the hands of the person get out of the image or if the person is going to leave the image. In the first case the virtual cameraman detects derived from the motion direction of skin coloured segments that these regions will probably leave the image. This leads to a zoom out of the camera until those segments reliably stay inside the image. In the latter case the person additionally moves around which leads to a change of the coordinates of the indoor positioning system. Usually, there is a short delay to detect this by the indoor positioning system. It makes the cameraman zooming out a little until the indoor positioning detects the new coordinates and lets the camera track the moving lecturer by panning and tilting. In general, it is important to avoid too many camera movements. Instead of panning left and right continuously will choose a zoom out operation.

The segmenting approach used for our iris algorithm is also useful to control the zoom out. We have to detect and select again rapidly moving skin coloured segments. Therefore, we use a motion detection algorithm. We have to zoom out until the selected segments do not disappear any more in a series of images. Adding a little border space to this range leads to a good presumption for the final zoom.

For the correct movement of the camera we need another approach, in which we convert the coordinates given by the indoor positioning system into the camera device context. Using these results the coarse pan and tilt is done. For fine adjustment the image is segmented into a matrix of rectangles, combining skin coloured segments with movement detection of adjacent frames. The regions belonging together are determined in a robust way. The centre of gravity is calculated and is used for fine positioning.

4.4 Functionality

To implement the complex tasks of iris setting, adjusting zoom and steering the camera we needed to implement some basic algorithms and combine them in an appropriate way. Therefore, we use conversion algorithms for Cartesian and Polar coordinates, and transfer algorithms into the device context of the camera for steering it.

The algorithms for image processing, besides standards like: saving, resolution adaptation, cropping and grey scale converting, consist of:

- grabbing of a still image as bitmap or JPEG,
- calculating histograms of grey scale images,
- noise reduction,
- calculation of differences between two images,
- edge detection,
- image segmentation into rectangles and interpreting these segments again as images,
- calculation of percentage and absolute value of a certain detail, e.g., edges in images,
- skin colour detection,
- motion detection to detect people in an image as nobody will be absolutely immobile.

In addition we select segments based on the results of the algorithms named above and the combinations of their results. For example, selecting a segment can be done by the number of pixels being parts of edges, having skin like colour, or belong to a part of the image showing movements, etc. The segments are joined to regions if they share an edge. Regions may be differentiated by the location of the segments and by the number of segments. The centre of gravity of a region or of more than one regions is calculated.

For testing purposes we have implemented additional algorithms to mark segments, regions, the centre of gravity or to set certain segments to black. Figure 6 shows an example of the skin colour detection: the original sample, the sample with rectangle segments marked and the sample with all selected segments joined to regions and all unselected segments set to black (from left to right).

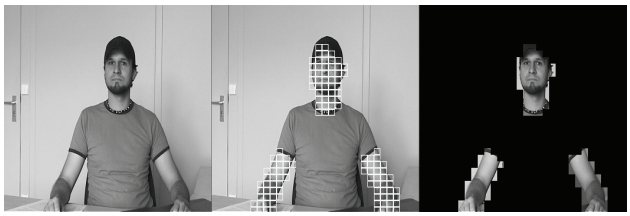

Figure 6. Skin Color Detection Example

To identify persons we use in addition to skin color detection a movement detection algorithm. An example is shown in Figure 7.

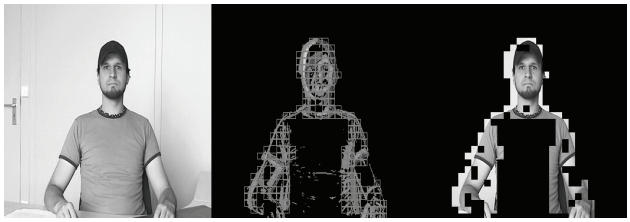

Figure 7. Movement Detection Example

On the left, there is the sample image, in the middle there is the minimal movement as a difference image of the sample image and its following image joined with the marked segments. On the right is the filtered image.

By combining the results of both algorithms the detected regions give us a good estimation about the location of a person. By cal-

culating the centre of gravity, which would be in the middle of the chest of the shown person in this example, we are able to know where on the screen a person is located.

As presented above, the algorithms used for analysis on the images work well for our purposes. To make sure the analysis will work during the recording we sample the images used for analysing parallel to normal streaming operation. Therefore, all adjustments for focus, iris, position and zoom are steered based on these results without disturbing the normal video stream. However, the control-loop is implemented as a framework for our analysing algorithms to realise a virtual cameraman, by fulfilling similar jobs compared to a real cameraman in his work.

5. CONCLUSION

Starting from the work of a real cameraman, we have determined the camera functions desirable for lecture recording. These functions were mapped to a control-loop approach to create an automatic cameraman. We then showed that based on indoor positioning algorithms and image processing algorithms, all the necessary information can be obtained to steer an automatic cameraman. The information we use is shared through a network with the virtual director and other virtual cameramen so that more complex control processes can be implemented.

The director module is based on the system presented in [10]. It is the master of the virtual cameraman. It receives the information sent by the cameraman and incorporates it into its own weighing algorithm to choose the next shot for the recording. The director module also decides how to do the framing and how to react on other sensor inputs.

The status of our project is work in progress. At the moment we are in the last steps of development and in the first steps of testing. So far, we got very robust results, as shown in the figures above. Already noticeable is that the modular approach of the cameraman and of the whole system allows implementing even complex video production rules easily, which is absolutely necessary to overcome uniformed lecture recordings. In the upcoming fall semester we will test the system in a graduate course on Multimedia Technology to adjust and improve the settings.

In the near future we plan to change the focusing operation from the camera-implemented auto-focus to an algorithm based on the distance calculated by the given coordinates, on the specifications of the camera optic and on the selected zoom level. We also plan to optimize the motion detection for new purposes, e.g., detection of annotation, detection of the change of slides and detection of gesticulating persons. We also intend to implement the detection of different persons shown in the image; this is important to improve the overall framing of a person. The communication protocol between the director module and the virtual cameramen will be extended to better support the video production rules.

In the longer-term future we intend to add a sound engineer module for the lecturer and the questioners and a lighting technician module. The director module will be improved to not only switch between different camera streams using hard cuts, but also use cross fades, picture-in-picture effects, etc. for more lively transitions.

Finally a thorough empirical evaluation of the entire system with students is planned, both from a technical and a pedagogical point of view.

6. ACKNOWLEDGMENTS

We would like to thank Adin Hassa, Burkard Kreisel and their entire team at Südwest Rundfunk (SWR) Baden-Baden for letting us take a look behind the scenes of live TV production.

7. REFERENCES

[1] Bianchi, M. H. AutoAuditorium: A fully automatic, multi-camera system to televise auditorium presentations, *Proceedings of the Joint DARPA/NIST workshop on smart spaces technology*, Gaithersburg, MD, USA, 30-31 July 1998.

[2] Cutler, R., Rui, Y., Gupta, A., Cadiz, J.J. Distributed Meetings: A Meeting Capture and Broadcasting System, *Proceedings of ACM Multimedia 2002*, Juan-les-Pins, France, 2002, 503-512.

[3] Graf, H. P., Cosatto, E., Gibbon, D., Kocheisen, M., Petajan, E. Multimodal system for locating heads and faces, *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, Killington, VT, USA, 1996, 88-93.

[4] Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., Merkl, H., Pankanti, S. *Smart video surveillance: Exploring the concept of multiscale spatiotemporal tracking,* IEEE Signal Processing Magazine, Vol. 22, No. 2, 03/2005, 38-51.

[5] He, L., Cohen, M. F., Salesin, D. H. The virtual cinematographer: A paradigm for automatic real-time camera control and directing, *Proceedings of ACM SIGGRAPH: 23. International Conference on Computer Graphics and Interactive Training 1996*, 217-224.

[6] Huang, Q., Cui, Y., Samarasekera, S. Content based active video data acquisition via automated cameramen, *Proceedings of IEEE International Conference on Image Processing (ICIP 1998)*, 1998, 808-812.

[7] King, Th., Haenselmann, Th., Kopf, S., Effelsberg, W. *Key Factors for Position Errors in 802.11-based Positioning Systems*, Technical Report TR-2007-003, Department for Mathematics and Computer Science, University of Mannheim, March 2007, http://www.informatik.uni-mannheim.de/pi4/publications/King2007b.pdf, last visited 05/29/2007.

[8] King, Th., Kopf, S., Effelsberg, W. Position detection of students in lecture halls using the Chi-Square-Adaptation-Test. (In German: Positionserkennung von Studierenden in Hörsälen mit dem Chi-Quadrat-Anpassungstest), *Proceedings of the 3. GI/ITG KuVS Fachgespräch "Ortsbezogene Anwendungen und Dienste" 2006*, Berlin, Germany, September 2006, 44-48.

[9] Lampi, F., Kopf, S., Effelsberg, W. Automatic FSM-Based Video Directors for Lecture Recording, *Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2007)*, Vancouver, BC, Canada, 2007, to appear.

[10] Lampi, F., Scheele, N., Effelsberg, W. Automatic Camera Control for Lecture Recordings, *Proceedings of the World*

Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2006), Orlando, FL, USA, 2006, 854-860.

[11] Lauer, T., Ottmann, Th. Means and Methods in Automatic Courseware Production: Experience and Technical Challenges, *Proceedings of E-Learn 2002*, Montreal, Canada, 2002, 553-560.

[12] Liu, Q., Kimber, D., Foote, J., Wilcox, L., Boreczky, J. FLYSPEC: A multi-user video camera system with hybrid human and automatic control, *Proceedings of ACM Multimedia 2002*, Juan-les-Pins, France, 2002, 484-492.

[13] Mukhopadhyay, S., Smith, B. Passive capture and Structuring of Lectures, *Proceedings of ACM Multimedia 1999*, Vol.: 1, Orlando, FL, USA, 1999, 477-487.

[14] Oliver, N., Berard, F., Pentland, A. LAFER: Lips and face tracker, *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, USA, 1996, 123-129

[15] Rui, Y., Gupta, A., Cadiz, J. J. Viewing meetings captured by an omni-directional camera, *Proceedings of ACM CHI 2001*, Seattle, WA, USA, 2001, 450-457.

[16] Rui, Y., Gupta, A., Grudin, J., He, L. *Automating lecture capture and broadcast: Technology and videography*. ACM Multimedia Systems Journal Vol.10, No.1, 2004, 3-15.

[17] Rui, Y., He, L., Gupta, A., Liu, Q. Building an intelligent camera management system, *Proceedings of ACM Multimedia 2001*, Ottawa, Canada, 2001, 2-11.

[18] Scheele, N., Mauve, M., Effelsberg, W., Wessels, A., Horz, H., Fries, St. The Interactive Lecture - A new Teaching Paradigm Based on Ubiquitous Computing, *Poster Proceeding of the CSCL 2003*, 2003, Bergen, Norway, 135-137.

[19] Scheele, N., Seitz, C., Effelsberg, W., Wessels, A. Mobile devices in Interactive Lectures, *Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 04)*, 2004. Lugano, Switzerland, 154-161.

[20] Thompson, R. *Grammar of the shot,* Elsevier Focal Press, Oxford, 2nd edition, 2002.

[21] Yang, J., Waibel, A. A real-time face tracker, *Proceedings of IEEE Workshop on applications of Computer Vision (WACV)*, Sarasota, FL, USA, 1996, 142-147

[22] Zhang, C., Crawford, J., Rui, Y., He, L. An Automated End-to-End Lecture Capturing and Broadcasting System, *Proceedings of ACM Multimedia 2005*, Singapore, 2005, 808-809.

[23] Zhang, C., Rui, Y., Crawford, J., He, L. *An Automated End-to-End Lecture Capturing and Broadcasting System*, Technical Report MSR-TR-2005-128, Microsoft Research, September 2005, ftp://ftp.research.microsoft.com/pub/tr/TR-2005-128.pdf, last visited 05/09/2007.