

# Application of Audio and Video Processing Methods for Language Research and Documentation: The AVATeCH Project

Przemyslaw Lenkiewicz<sup>1</sup>(✉), Sebastian Drude<sup>1</sup>,  
Anna Lenkiewicz<sup>1</sup>, Binyam Gebrekidan Gebre<sup>1</sup>, Stefano Masneri<sup>3</sup>,  
Oliver Schreer<sup>3</sup>, Jochen Schwenninger<sup>2</sup>, and Rolf Bardeli<sup>2</sup>

<sup>1</sup> Max Planck Institute for Psycholinguistics, Wundtlaan 1,  
6525 XD Nijmegen, The Netherlands  
{Przemyslaw.Lenkiewicz,Sebastian.Drude,  
Anna.Lenkiewicz,BinyamGebrekidan.Gebre}@mpi.nl  
<http://www.mpi.nl/>

<sup>2</sup> Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS,  
Schloss Birlinghoven, 53757 Sankt Augustin, Germany  
{Jochen.Schwenninger,Rolf.Bardeli}@iais.fraunhofer.de  
<http://www.iais.fraunhofer.de/>

<sup>3</sup> Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute,  
Einsteinufer 37, 10587 Berlin, Germany  
{Stefano.Masneri,Oliver.Schreer}@hhi.fraunhofer.de  
<http://www.hhi.fraunhofer.de/>

**Abstract.** Evolution and changes of all modern languages is a well-known fact. However, recently it is reaching dynamics never seen before, which results in loss of the vast amount of information encoded in every language. In order to preserve such rich heritage, and to carry out linguistic research, properly annotated recordings of world languages are necessary. Since creating those annotations is a very laborious task, reaching times 100 longer than the length of the annotated media, innovative video processing algorithms are needed, in order to improve the efficiency and quality of annotation process. This is the scope of the AVATeCH project presented in this article.

**Keywords:** Language research · Audio and video processing · Automated annotation

## 1 Introduction

Languages and cultures have always been evolving due to many well-understood historical factors. However, in recent decades the dynamics of those changes

---

AVATeCH is a joint project of Max Planck and Fraunhofer Institutes, started in 2009 and funded by MPG and FhG. Some of the research leading to these results has received funding from the European Commissions 7th Framework Program under grant agreement n 238405 (CLARA).

got an enormous speedup, partially due to the globalization process. As a consequence UNESCO has reported that currently one of the world's languages becomes extinct every two weeks and even major languages, like English or French, are changing. Similarly to biology, we can see a huge decrease in linguistic diversity [1]. Also the cultures are changing rapidly, identity building for young people becomes very difficult and the stability of societies is affected. We are deemed to loosing part of our cultural heritage since every language can be seen as a unique result of evolution resulting in rather different language systems. We also risk losing much of our knowledge about environment, species etc. since this is to a large extent encoded in the semantics of a given language.

During the last decades we recognize an increasing awareness about these threats resulting in a number of world-wide initiatives to document, archive and revitalize languages (DOBES<sup>1</sup>, HRELP<sup>2</sup>, PARADISEC<sup>3</sup>). It is well understood now, that we have the obligation to preserve our material and knowledge about languages for future generations, since they may want to understand their roots. Future generations may also want to return to proper linguistic constructions that are currently blurring or which we currently are loosing.

Furthermore, in the last years it has become clear that making recordings alone is not sufficient to guarantee that future generations will indeed be able to access the data. Recordings without appropriate annotations and metadata can be completely useless for anybody that has no knowledge about their creation and purpose. Manual annotation of such recordings is a very laborious task. Therefore the AVATeCH project has been started with partners from the HHI and IAIS Fraunhofer Institutes, with the goal to develop automated annotation algorithms [2,3]. Their role is twofold: (1) they would allow a significant decrease of time necessary to perform the annotation task; (2) automation of some parts of the process can greatly increase the uniformity of the annotations created worldwide by different researchers, which would contribute to consistency of the available language data.

In this article we describe in detail the algorithms that operate on video recordings and present some results that we could obtain with them.

## 2 Audio Analysis Algorithms

Linguistic researchers are interested in annotating many aspects of the audio track. These aspects can vary greatly, and range from word or phonetic transcriptions to prosody and emotional aspects of speech. The diversity of use cases would require numerous dedicated automatic analysis tools. Instead we have decided to focus on smaller annotation building blocks within AVATeCH, which accept user feedback and can be combined to speed up complex tasks. Here we present two examples of these blocks.

---

<sup>1</sup> <http://www.mpi.nl/dobes>

<sup>2</sup> <http://www.hrelp.org/>

<sup>3</sup> <http://www.paradisec.org.au/>

## 2.1 Automatic Detection of Speech

Here, the focus is on detecting the speech portions of the signal, as they are the most interesting parts of the signal with respect to annotation. Data diversity means unprecedented sound environments. To capture this diversity, we use a model-based or data-driven approach for speech detection instead of an approach in terms of heuristics and thresholds. We rely on the algorithm to learn a model of speech and non-speech from examples of corresponding data. Here the assumption is that non-speech sounds display patterns that are very different from speech, and can be effectively represented by a probabilistic model.

User feedback can be used to improve the models. In the user feedback scenario, the user is providing examples of speech or nonspeech segments from the file that is currently being annotated. Non-adapted baseline models can be adapted with a specific file currently annotated, and used for annotation of similar files.

In the first evaluation the models were adapted in a supervised manner, with data known to be either speech or non-speech. In a second experiment the models were used to classify the speech and non-speech segments in a file, and then use these automatic annotations for adaptation. For evaluation we used the metric defined in [4].

## 2.2 Language-Independent Alignment

The problem of aligning a written transcript to an audio recording of speech, i.e., finding out which of the words in the transcript is spoken at which time, is typically addressed by employing an HMM-based speech recognizer. In a process called forced alignment, the HMMs in the speech recognizer are concatenated in a way determined by the transcript and then a Viterbi decoding step will yield an alignment.

However, when dealing with many different low resource languages, this approach is not viable because it is prohibitive to train speech recognizers for each of these languages. Beyond trying to patch the mainstream approach and reuse speech recognizers for other languages by finding sufficiently good phoneme mappings between the languages, we are looking for a completely language independent approach.

Our approach is divided into two steps. First, we find specific points in the audio — so called anchor points — which are aligned based on structural properties of text and audio. Then, portions of text between those anchor points are aligned based on estimates of the rate of speech and by pause lengths in the audio.

For the first step, we use a mapping between punctuation marks in the text and pauses in the audio as basic anchor points. To this end, punctuation marks are extracted from the text and candidates for speech pauses are extracted from the audio. Based on the type of punctuation mark, the order of punctuation marks and pauses, and the length of the pauses, the algorithm finds a basic mapping between punctuation and pauses. The user can then (but need not)

manually correct the punctuation-to-pause alignment by selecting from a list of a fixed number of most probable pauses for each punctuation mark.

In addition to punctuation, we also make use of repetitions in the text. This is enabled by the user marking one occurrence of each repeated word in the audio. Then, our algorithm finds repetitions of these words in the audio and matches them to the text positions. For the alignment of non-anchor words, the rate of speech is estimated based on the length of the transcription of words between two anchors. The length of detected pauses is also incorporated in the rate of speech estimation.

These rather simple cues already give pretty good alignment results with considerably less manual effort than completely manual alignment.

### 3 Video Analysis Algorithms

The main principle that led the development of video analysis algorithm was to reduce the time needed to perform the annotation process and, when possible, make it completely automatic. The creation of robust and efficient algorithms was mandatory, due to the huge size of the video database of the MPI and the great diversity of the content. These two constraints defined the main guidelines in the creation of new algorithms and in the adaptation of existing ones to this specific problem. All the algorithms are designed to work without user interaction (except for the initial setup of some parameters) to allow batch processing on multiple videos. The implementation is done using a highly modular structure, so that future automatic annotators can be easily integrated in the current framework, using as input the results provided by the previous detectors.

The video analysis framework consists on a set of pre-processing algorithms that provide important information for the core module, the hand and head tracking analysis. In the next subsections, the different modules are described in short.

#### 3.1 Skin Color Estimation

Human body part tracking mainly belongs to skin colour, which is a unique and very relevant feature of human body. Due to the peculiarities of the dataset in the underlying application scenario, there is no unique set of skin color parameters which can achieve good results in the entire dataset and therefore typical approaches that make use of a training set to collect the parameters for skin detection on the entire dataset cannot be applied [5,6].

Automatic skin colour estimation is performed before any detection and tracking of human body parts will start. It is mainly based on an optimization scheme for a selected range of U and V values in the YUV domain that belongs to human skin colour. Some candidate frames of the video are selected where individual regions are moving assuming that human body parts correlate with moving regions. Based on the selected frames, an optimization scheme is applied that performs a parameter variation in the YUV domain aiming at compact

pixel clusters that correlate with the moving region. The resulting parameters define a subset in the YUV color space and all the pixels in the image within this subset are marked as skin.

### 3.2 Face Detection

In this module, a software library from Fraunhofer IIS is used, the so-called Shore Library<sup>4</sup> [7]. This face detector allows robust frontal and profile face detection and tracking for a large variety of faces. This library is commercially used in many image and video annotation tools as well as in security. Due to a common agreement between Fraunhofer HHI and Fraunhofer IIS, the library can be used within our framework for research purposes. The face detector offers the position and size of any detected face in the image, the left and right eye position, frontal or profile view as well as a rating how likely it is a face. All this information is further exploited for human body part tracking.

### 3.3 Global Motion Detection and Background Estimation

Accurate motion analysis allows distinguishing between different types of video content and it can be used to segment a video in order to select only the parts which are relevant for the researchers. For example, the presence of zooms and motion inside of a scene are usually the most interesting, while shots containing just panning and a low amount of internal motion are of little interest and can be usually discarded without further analysis. For each frame in the video a motion vector map is computed using the Hybrid Recursive Matching (HRM) algorithm [8]. For each vector both the absolute value (i.e. the speed, calculated as L2 norm) and the phase (i.e. the orientation) are then computed. To detect the direction of global motion an 8-bins histogram of the phase of the motion vectors is also computed. Frames are considered candidates for global motion analysis when a significant amount of motion vectors occurs. Pan, tilt and zoom can be detected as well by exploiting the phase histogram of the motion vectors. The approach used for zoom detection is similar to [9] and is based on the idea that when a zoom happens, the majority of motion vectors point to (or come from) the center of the frame, with phases that range evenly between  $[0, 2\pi]$  and absolute values that decrease nearing the center of the image. If no global motion is detected for a particular frame but there is nonetheless a significant amount of motion in the image, the frame is then marked as having motion inside the scene.

The detection and tracking of human body parts can be performed quite robustly if the moving foreground object is detected and tracking is just performed on the skin colour regions of the foreground object. Hence the global motion analysis, i.e. the detection of zoom, pan or tilt, helps to identify when the camera is not moving, in order to start adaptive background estimation. The adaptively estimated background then supports the detection and tracking

<sup>4</sup> <http://www.iis.fraunhofer.de/bf/bsy/fue/isyst>

of human body parts. The detection and tracking framework assumes a static camera and constant lighting; the background subtraction is then performed by subtracting the median image, pre-computed using a set of frames created sampling the input video once a second.

### 3.4 Hands and Head Tracking

The core algorithm works at first by segmenting the image in skin vs. non-skin pixels, using the information provided by the skin color estimator. The adaptive background estimation module supports the robustness of segmentation as skin colored background does not harm the segmentation and clustering. The subsequent step in the detection process involves the search of seed points where the hands and heads regions most likely occur. Histograms along the horizontal and vertical directions compute the number of pixels with luminance and color values within the desired interval; the pixels where a maximum occur in both the directions are selected as seed points (Fig. 1, left). A region growing algorithm is then applied to the seed points in order to cluster together all the skin pixels in the neighborhood. Each region is approximated by an ellipse (Fig. 1, right). The tracking is performed by analyzing the change in position and orientation of the ellipses along the timeline, assigning labels based on position of the regions in the current and previous frames. The face detection module further supports the robust assignment of the head cluster.

On the frame level we get the following information:

- The x,y-position of each cluster for each person detected in the video;
- The size of the cluster defined by the minor and major axis of the ellipse;
- The orientation of the cluster defined by the orientation of the ellipse;
- The speed of a cluster, i.e. the change in position between succeeding frames.



**Fig. 1.** Histogram of skin color points (left) and examples of ellipses approximating the skin areas in given image (right) (Color figure online)

### 3.5 Higher Level Analysis and Annotation

The tracking information described in the previous section is then exploited to cluster the frame based information and to achieve a more compact annotation for longer temporal segments. Beside an analysis of the motion of hands, relational and directional information is of interest by researchers in order to detect specific motion patterns and gestures.

**Hand Motion Analysis.** The hand motion analysis is performed on a temporal segment in which a hand is moving a significant amount. Within the resulting temporal segment, three different properties are calculated. The speed of the hand movement, the area covered by the moving hand in the considered temporal segment and the major orientation of the movement. The threshold defining the start and end of a movement is set by the user amongst three possible values, depending on how much detail the user requires. The lowest threshold allows the detection of the majority of the movements but can also lead to an over-segmentation of the detected movements. Conversely the high threshold may not detect all the occurring movements but avoids the risk of over-segmentation. In order to make it possible for the user to compare movements detected in different videos the values of speed and area covered by the movement are computed as frame and resolution independent values.

On the temporal segment level, we get the following information:

- The movement of a cluster, defined by a threshold for the speed;
- The area covered by the cluster during a motion;
- The orientation of a movement;
- The average speed during a movement.

**Directional Analysis of Hand Motion.** If the same temporal segment is considered, where a hand is moving, then, a directional analysis of the motion is performed. The analysis of the direction is limited to the four major directions, UP, DOWN, LEFT and RIGHT. Although only temporal segments with a certain amount of motion are considered, several sub-segments may occur, which contain only very small undirected motion. These sub-segments are annotated with the label “SMALL\_MOTION”.

**Relational Analysis of Body Parts.** In addition to the motion and direction of the hands, the relative position of body parts is annotated as well. The following relations are investigated, if, and only if, the specific body parts have been labeled: (1) Left hand to head relation; (2) Right hand to head relation; (3) Left hand to right hand relation.

The position of left and right hand is defined relative to the head. Hence the image is divided in a grid and depending on the position of the hand a horizontal and vertical grid coordinate is calculated.

In Fig. 2, an example is given for some sample frames of a testing video. The timeline below shows the different kind of annotations over time. The sample

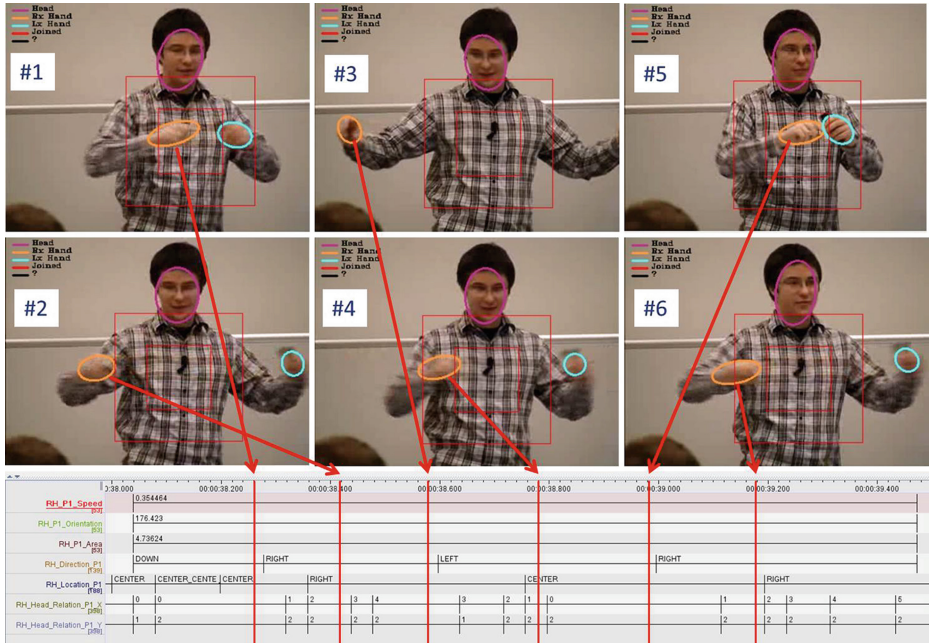


Fig. 2. Example for annotation of human body part movements

frames relate to different time instants. All the annotations relate to the position and movement of the right hand of the person, the orange ellipse (this is the left hand from the observer’s view). The first three annotations are single values for the whole motion segment i.e. the speed, the orientation of movement and the area covered by the moving hand. After that, the direction of movement, the location of the right hand in gesture space and the relation between right hand and head in X and Y-direction is shown. It can be seen that the different annotations relate very well to the analysed video.

## 4 Advanced Machine Learning-Based Annotation Tools

### 4.1 Media Query Language for Automated Annotation

The process of coding linguistic recordings differs depending on the research question that needs to be answered and the experience and preferences of the annotating person. Even with the support of semi-automated annotation tools more specific annotation still may be required. To influence the unification of coding schema the Media Query Language (MQL) has been designed and partially developed at the Max Planck Institute for Psycholinguistics. The task of this language is to search through the automatically detected annotation segments to find those indicating the desired strategies for multimodal communication, using body movement and addition of a semantics decoded by the researcher.



The MQL implementation and design involves description of rules of syntax in a form of a context-free grammar rules resulting in: a Look-Ahead Left-to-Right (LALR) parser, an implementation of the semantic action to the nodes of the AST of the parser, integration with the AVATeCH algorithms.

The design assumes decomposition of the movement according to its kinetic properties followed by assembly of atomic parts of the movement in order to result as kinetically complicated pattern. In the terminology of the MQL the atomic part of the movement is called a pattern. The pattern is a movement primitive associated only to one body part and by design decision is not carrying any semantic meaning. The combination of patterns called an action and it is the part of the MQL where semantics are added. It is created according to the research theory, enabling the possibility of direct comparison with instances of the movement present in the recording.

The relation between MQL elements can be spatial and temporal. Motion can be also described with specific property of the body part relocation over given number of frames. All elements can be saved in user-customizable libraries, which may be exchanged e.g. between colleagues or team members. As the definition of pattern and action depends on user's preferences, the requirements of flexibility for data coding is met.

The MQL typically works on the output of video-analyzing algorithms. The applied pre-processing algorithm increases the feature space  $R^n$  in a way that for each detected body part  $m$  described with  $n$  values calculation is performed changing the dimension of vector describing one frame to  $R^{(n-1+l)m+1}$ , where  $l$  is the number of added features. As the result of pre-processing the vectors describing frames for each body part contain movement classified according to the kinetic properties of the movement with consideration of a threshold determining the movement. The threshold is user-customizable and it is used as the system sensitivity adjustment. Other values present in the frame vector are calculations of total and partial distance and average and local speed. The presided calculations allow user to retrieve the data and formulate theory on a precise information. The principle goal of MQL is to allows research validation and replication of results on different data.

## 4.2 Speaker Diarization Using Gesturing Activity

Speaker diarization – the task of determining *who spoke when?* in an audio/video recording – has a number of applications in document structuring of multimedia data. Many of these applications come in the form of speaker indexing (used for video navigation and retrieval), speaker model adaptation (used for enhancing speaker recognition) and speaker attributed speech-to-text transcription (used for speech translation or message summarization).

Early research in speaker diarization focused on using audio signals [10] and since recently on audio-video signals [11]. This work performs speaker diarization using only video signals. The basis of our work is the hypothesis that *the gesturer is the speaker*. The hypothesis arose from the observation that although a speaker may not be gesturing for the whole duration of the speech, a gesturer is

mostly speaking. Theoretical and empirical evidence for the tight linkage between gesture and speech is also documented in the literature [12].

The synchronization of gesture and speech, which is one of the ways gesture and speech are linked, implies that identifying *who is gesturing* is indicative of *who is speaking*. Based on this idea, we proposed a vision-only diarization algorithm, the details of which are given in [13]. The basic idea is that the algorithm tests for gesture occurrence in regions of the video where there is optical flow. Significant optical flow are associated with particular regions and these regions are generally the same regions occupied by the speakers and not by the listeners.

We evaluated our algorithm on the 13 video recordings of the Augmented Multi-Party Interaction (AMI) meetings<sup>5</sup>. The AMI corpus consists of annotated audio-visual data of four individuals engaged in a meeting. We used a subset of the IDIAP meetings (IN10XX and IS1009x) totaling 8.9 video hours. The best performance obtained, measured in terms of diarization error rate, is 15%. The result shows that gesture can indeed be used for speaker diarization. Future work will consider combining gesture with audio for more accurate and robust performance.

## 5 Experiments and Results

For the purpose of analyzing the effectiveness of our solutions we have defined the measure as the difference between the times it takes to create annotation to given media manually and with our algorithms. This value is not easily calculated, as this time depends on factors like: the purpose of the recording and contents of the media; what exactly from the contents needs to be analyzed and annotated; the person performing the annotation process and their expertise.

We have created a scenario in which researchers had to perform annotation tasks to answer different linguistic research questions. The tasks have been chosen to represent a very common set of actions taken by researcher annotating their recordings and they included: (1) labelling the speech segments of all speakers in an audio recording; (2) labeling all gestures of a person in a video recording. For each experiment the tasks have been first performed manually by several researchers and the time necessary to carry them out was measured and averaged. As the next step, the same tasks have been performed with the help of AVATeCH recognizers. In each experiment after the automated annotations have been created, a human annotator has verified them in order to perform necessary corrections, as the automated annotations were not flawless. The time to preform these corrections is measured and included in the total time of the automated annotation scenario.

Experiments on labeling speech have been performed using two parts of a field recording, representing typical noise and quality conditions for such recordings. One part was annotated with the help of recognizers, the other one without them.

<sup>5</sup> <https://www.idiap.ch/dataset/ami/>

The two used parts were slightly different in annotation density: the experiment with the manual one resulted with 557 annotation blocks and the recognizers one with 457 blocks. However, the average annotation speed expressed in the number of created annotation blocks per minute is independent of this difference. As it can be seen from table below - the annotation process is almost twice faster for the scenario with recognizers.

	Length	Blocks	Density	Time	Speed
Manual	40.55	557	13.57	1800	0.31
AVATech	40.07	457	11.405	780	0.58

The column labeled Length shows the length of annotated media, Blocks is the number of annotation blocks created in the final annotation, Density describes the number of annotation blocks per minute of recording to give an idea about the final annotation that was created, Time shows the amount of time necessary to perform the full annotation and Speed describes the speed of the annotation process in number of created annotation blocks per minute.

The experiments on gesture labeling were performed on two videos. The annotations to be created have been first consulted with 2 researchers and then performed by an assistant. Four annotation experiments were performed, 2 for each video with and without the recognizers. As it can be seen in table below the resulting efficiency for gesture segmentation with recognizers has increased by 46 % and 43 % in these two experiments. The length of Video 1 was 1 min and 24 s, the length of Video 2 was 46 s.

	Video 1	Video 2
Manually	30 min 00 s	13 min 10 s
AVATech	16 min 11 s	7 min 30 s

## 6 Conclusions and Future Work

The specification and implementation of the above-described video processing recognizers has been performed in a very close contact with linguist researchers and according to the needs they have specified. After testing the relative effectiveness of our methods and witnessing the significant decrease of time necessary for annotations, we can say that our goals have been chosen correctly and our methods have proven very useful.

The successful work started in AVATech is continued with the extended group of partners in the AUVIS<sup>6</sup> project. As our next steps we are planning to

<sup>6</sup> [http://tla.mpi.nl/projects\\_info/auvis](http://tla.mpi.nl/projects_info/auvis)

fully develop the possibility of detecting and tracking the hands in the videos and also work together with new partners to develop new recognizers that would create new types of annotations for different research questions.

## References

1. Crystal, D.: *Language Death*. Cambridge University Press, Cambridge (2000)
2. Lenkiewicz, P., Gebre, B.G., Schreer, O., Masneri, S., Schneider, D., Tschöpel, S.: Avatech automated annotation through audio and video analysis. In: Choukri, K., Declerck, T., Doğan, M.U., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S. (eds.) *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)* (N. C. C. Chair), Istanbul, Turkey. European Language Resources Association (ELRA), May 2012
3. Lenkiewicz, P., Uytvanck, D.V., Wittenburg, P., Drude, S.: Towards automated annotation of audio and video recordings by application of advanced web-services. In: *INTERSPEECH, ISCA* (2012)
4. Ajmera, J., Bourlard, H., Lapidot, I., McCowan, I.: Unknown-multiple speaker clustering using hmm. In: *INTERSPEECH, Citeseer* (2002)
5. Terrillon, J.-C., Shirazi, M., Fukamachi, H., Akamatsu, S.: Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In: *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 54–61 (2000)
6. Vezhnevets, V., Sazonov, V., Andreeva, A.: A survey on pixel-based skin color detection techniques. In: *Proceedings of the GRAPHICON-2003*, pp. 85–92 (2003)
7. Kueblbeck, C., Ernst, A.: Face detection and tracking in video sequences using the modified census transformation. *J. Image Vis. Comput.* **24**(6), 564–572 (2006)
8. Atzpadin, N., Kauff, P., Schreer, O.: Stereo analysis by hybrid recursive matching for real-time immersive video conferencing. *IEEE Trans. Circuits Syst. Video Technol.* **14**(3), 321–334 (2004)
9. Dumitras, A., Haskell, B.G.: A look-ahead method for pan and zoom detection in video sequences using block-based motion vectors in polar coordinates. In: *Proceedings of the ISCAS*, vol. 3, pp. 853–856 (2004)
10. Tranter, S., Reynolds, D.: An overview of automatic speaker diarization systems. *IEEE Trans. Audio, Speech, Lang. Process.* **14**(5), 1557–1565 (2006)
11. Anguera Miro, X., Bozonnet, S., Evans, N., Fredouille, C., Friedland, G., Vinyals, O.: Speaker diarization: a review of recent research. *IEEE Trans. Audio, Speech, Lang. Process.* **20**(2), 356–370 (2012)
12. McNeill, D.: So you think gestures are nonverbal? *Psychol. Rev.* **92**, 350–371 (1985)
13. Gebre, B.G., Wittenburg, P., Heskes, T.: The gesturer is the speaker. In: *ICASSP 2013* (2013)