Automatic video stream selection method by on-air microphone detection

Florent Lefevre, Vincent Bombardier, Nicolas Krommenacker, Patrick Charpentier, Bertrand Petat

Automatic video stream selection method by on-air microphone detection

Lefevre Florent * †, Bombardier Vincent *, Krommenacker Nicolas *, Charpentier Patrick * and Petat Bertrand †

*Université de Lorraine
CRAN CNRS UMR 7039,
F-54506 Vandoeuvre-les-nancy
firstname.lastname@univ-lorraine.fr
†CitizenCam
132 rue andré Bisiaux
54320 Maxéville
(flefevre, bpetat)@citizencam.eu

***Abstract*—This article presents an automatic video editing method for video stream selection in a multi-camera environment. The specific context of this study is council meetings recording and broadcasting. In order to offer the best view to spectator our method is based on a speaker detection, to select the right camera. Since no sound information is available, the proposed method is based on the detection of the change in the visual state of the microphones LED in image sequences, in order to automatically and efficiently select the camera where the speaker is. Studies about the suitable size of the used sliding window and about the relevant features' selection for the verification of microphones' activation are also presented. We have selected seven features to effectively train one classifier, which can be used on different cameras. The feasibility of this approach is shown by the experimentation on councils' videos where the proposed method allows a very efficient detection of the speaker in real-time.**

***Index Terms*—Automatic video editing, Speaker Detection, Feature selection, LED detection**

I. Introduction

Automatic video editing allows small events to be available to a much larger audience. Indeed, many events cannot be broadcast because of the cost of the human production crew and equipment . By Automatic video editing, we mean automatic selection of the best viewing angle in a multi-camera system, in order to provide to the spectator the video stream where the action take place. CitizenCam [1] is a French company which offers multi-camera automatic recording solutions in order to retransmit on the web every type of event. Their goal is to reduce costs by automating recording and broadcasting while using IP cameras. This in order to be affordable to the greatest number of people. In the context of council meetings, around ten cameras are filming the potential speakers from the center of the room. Searching in the ten proposed views, the one where the speaker is located is tedious for the spectator. In addition, live broadcasting of an event requires selecting the most relevant view for bandwidth saving reasons. That's why automatic view selection is mandatory in order to improve the user experience.

In [1], the authors describe autonomous camera systems as a system which have to solve three simultaneous problems: find out how the camera should be oriented to capture the action, define how the camera should move to film the subject and finally select the right camera which should be on air, In our case study, we use fixed camera so our problem is to find out if there is a person speaking on each camera to select the most relevant camera.

In literature, speaker detection or localization method generally need audio systems to confirm the presence of a talking person. Some methods [2]–[5] only use microphones array in order to localize the person talking in a room. Some other methods [6]–[11], use both video and audio in order to retrieve the person which are currently speaking. In [12], the authors propose an automating camera management to record lectures. The system automatically select the view where a person is talking: either the professor, or someone from the audience. In order to do that, they use a two-microphone array to estimate sound source location.

In our context, many cameras are filming the different speakers. However the audio systems used in most of the council meetings don't allow the isolation of each microphone, so we can't use this kind of technique. Changing the microphone system will greatly increase costs, making broadcasting impossible.

In order to detect the speakers, we offer to detect the light emitted by the microphone when a person speaks. Indeed, most of the French municipality is equipped with microphone systems where a LED lights up on the microphone when someone speaks. The detection of this light allows the detection of the speaker without changing the actual system. Even if those research have been made for traffic light detection, thus treat a different environment than the one we're interested in, light source's recognition is regularly the base of the work. The offered algorithms rely on the extraction of candidate regions

[1]This work results from a collaboration between CitizenCam and the CRAN.

using color thresholding methods. In this methods, [13], [14] are using the HSV color model, whereas [15], [16] are working on normalized RGB color models.

The first part of this paper explains the methodology used for the speaker detection. We detail the method proposed and make a study of the features used in the microphone verification step. Our second step will be presenting the videos used to validate our method and the results we obtain. Finally, we'll introduce how we can improve this method.

II. Speaker detection method

The method we propose is therefore based on the search for a light source in the image, symbolizing the activation of a microphone. Due to a number of light disturbances, it is necessary to check whether the received light is that of a microphone.One of the objectives being the generalization of the method to be used in different conditions, we will look for the features allowing the best characterization of microphones.

The proposed speaker detection method is divided in four principal parts and is described in the first paragraph. The second defines the choosing principle of the size of the sliding windows for the microphone detection. The third part details the choice of attributes used for the validation step of the speaker presence allowing a generalization of the method for different viewing angles.

A. Method description

The method presented in this paper detects light from the microphone, in order to find the people who is speaking. The principle of our method is shown in fig. 1 where each process is executed sequentially. In order to use this method in every situation, we introduce an initialization step to manually select the limits of the research region and the HSV threshold.

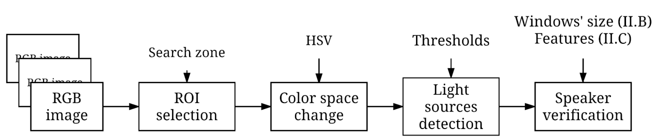

Fig. 1. Speaker detection method

Selection of the Region of Interest (ROI): The first step of our method is the selection of the ROI involving a reduction of computation's time and the risk of wrong detections.
Since microphones are located on the table in front of every eventual speaker, it is appropriate to look for microphones between the table and the top of the head of speakers. Because the microphones are not fixed on the tables, we need to define a large area in case the speakers move the microphones. An example of ROI selection is shown in Fig. 2 from the video "Sight 1", where the research zone measures 1800 x 340 pxl (cf Table I).

Color space selection: The second step of the method is to prepare for the light detection stage. To do this, it is

Fig. 2. Search region (Red rectangle)

necessary to change the color representation. Different color spaces were tested (RGB, HSV [17], CIE L*a*b* [18]). Those color spaces, frequently used in colorimetry, allow an efficient representation of the luminosity in the image, especially with the Intensity component of the HSV color space or the Luminance component of the L*a*b* color space.The choose of the HSV color space was made in line with the selection of relevant features (II-C)

Detection of light sources: In the first place, we want to localize microphones which are currently used. In other words, we want to find light sources emitted by active microphones. The HSV color model and especially the V component allows to efficiently find them. We execute thresholding to get candidate regions, as we can see in the following equation :

$$C(x,y) = \begin{cases} 1 \quad if & \begin{matrix} ((S(x,y) \geq Ts1 \cap S(x,y) \leq Ts2) \\ \cap (V(x,y) \geq Tv1) \cap V(x,y) \leq Tv2)) \end{matrix} \\ \\ 0 & else \end{cases} \tag{1}$$

Where $C(x,y)$ is the result of the thresholding operation, $S(x,y)$ and $V(x,y)$ are the saturation and intensity values. The four threshold values (Ts1,Ts2,Tv1 et Tv2) are empirically defined.

The lighting condition may change during the videos, therefore many reflections may occur, resulting in small lighting sources in the thresholded image (see Fig. 3). We apply a connected component analysis [19] in order to execute a dimensional thresholding. The light sources which are inferior (like reflections) or superior (like lighting) to the light source from a microphone are ignored.

Speaker verification: In order to check that the light sources are from microphones' LED, we use a classification tree [20] and a sliding windows techniques to separate zones that contain one active microphone from another. For each light sources, a windows of size 19x19 (see II-B), is swept across the candidate regions. The features calculated (see II-C) in each window location are then tested with a classifier. The

TABLE I
CHARACTERISTICS OF VIDEOS USED

Name	# images	Research zone	Area of microphone	pixel per mm	Properties
Sight 1	9 957	1800 x 340	345 pxl	0.71	Front view in close-up
Sight 2	22 459	830 x 230	190 pxl	0.23	Front view, wide shot
Sight 4	15 737	1000 x 300	120 pxl	0.21	Side view, wide shot
Sight 6	32 508	1000 x 220	210 pxl	0.30	Side view, wide shot

Fig. 3. Thresholded image (from Fig. 2): Microphone on the center, reflection on the right

classification tree is created during the initialization step of our system. Using a classification tree was a choice made thanks to the possibility of interpreting the causal connection, unlike methods like neural network [21], KNN [22] or SVM [23].

B. Windows' size selection

The selection of the size of the window for feature calculation is an important step for the characterization of the microphone state. In order to determine the best size to use, three windows were tested: a small one of 3x3 pixels, a medium one of 9x9 pixels and a big one of 19x19 pixels, as shown in fig. 4.

Fig. 4. The three windows' size

In order to check the windows size, a specific image database has been created. For each size of windows, 500 images from active microphones and 500 images from other parts were extracted from each the four videos explained in Table I. Classification trees were trained using the features (presented in II-C), extracted from these 4000 images. We use a k-fold cross-validation [24] with k=3 in order to prevent overfitting. Each trained classifier was next tested on each video, which will be presented in detail in III-A. The table II sum up the obtained results in term of accuracy.

The use of a small window doesn't permit to separate efficiently the two classes in every situation. Indeed, a small window doesn't represent all the structure of the microphone, that may explains the results obtained with images from sight 4 and sight 6. The medium and the big windows permit a better separation of those classes. The windows' size 19x19

TABLE II
INFLUENCE OF THE WINDOWS' SIZE ON CLASSIFICATION

Size:	3x3	9x9	19x19
Sight 1	100%	100%	100%
Sight 2	97.9%	99.6%	100%
Sight 4	88.6%	95.6%	98.2%
Sight 6	87.7%	96.8%	98.1%

are used for the microphone's presence check, thanks of its stability in orientation changes.

C. Feature selection

A preliminary step in the microphone's LED detection is to find the right measure to characterize them in the right way. In order to find those features, a great number of statistical measures were calculated for the three components of the tested color spaces. The most relevant were selected in order to determine those having the biggest impact on the detection of scenes where there is an active microphone from those with none. The following features were calculated in every components histograms:

- The mean (m), mode (mo), variance (σ^2) and standard deviation (σ).
- The Khi-2 (K2).
- The third central moment (skewness - sk), a measure of the asymmetry, as well as his third root ($\sqrt[3]{sk}$) for scaling reasons.
- The fourth central moment (kurtosis - ku), a measure of the flattening coefficient, as well as his fourth root ($\sqrt[4]{ku}$).

The calculation of these 81 features may be extremely time-consuming. Several feature selection methods were used in order to only keep those having a great discriminatory power. We can identify three types of features:

- Complementary: those which, combined, allow a better differentiation of classes
- Redundant: those which bring identical informations
- Antagonistic: those which bring contradictory informations as for the separation of the classes.

The removal of antagonistic features allows to obtain a greater recognition rate. The suppression of redundant features allows to decrease computational cost.
Images from test sequences have been extracted (2 images per second) and reduced zones (19 x 19 pixels) around the microphones were selected, as well as images derived from zones with inactive microphones. Every reduced images was then annotated "positive" or "negative". In feature selection, filters methods are used when a large number of samples

TABLE III
FIVE BEST FEATURES PER METHOD AND VIDEO.

Methods:	RELIEFF	SFS	SFFS	SBS	SBFS
Sight 1	$\sqrt[4]{ku}$V; mV; σ^2H; mS; σ^2V	mV; mH; moH; $\sqrt[3]{sk}$H; mS	mV;mH; moH; $\sqrt[3]{sk}$H; mS	σH; σ^2H; mV; mS; σ^2V	σH; σ^2H; $\sqrt[4]{ku}$V; moV; σ^2V
Sight 2	mH; skH; $\sqrt[4]{ku}$V; mV; moH	mV; moH; σ^2H; skH; moS	mV; mS; σ^2H; skH; moS	mH; σ^2H; $\sqrt[4]{ku}$V; moV; mV	mH; skH; moS; σ^2V; mV
Sight 4	mH; σ^2H; skH; moS; σ^2V	mH; $\sqrt[4]{ku}$V; mS; σ^2V; $\sqrt[4]{ku}$H	mH; $\sqrt[4]{ku}$V; mS; σ^2V; $\sqrt[4]{ku}$H	σ^2V; skH; mS; σ^2S; $\sqrt[4]{ku}$H	σ^2H; skH; σS; σ^2V; $\sqrt[3]{sk}$V
Sight 6	moV; σ^2H; skH; mS; $\sqrt[4]{ku}$V	mH; σ^2H; skH; σV; $\sqrt[4]{ku}$V	mH; σ^2S; skH; σV; $\sqrt[4]{ku}$V	mH; $\sqrt[4]{ku}$H; mS; mV; kuV	mH; kuH; mV; σ^2V; $\sqrt[4]{ku}$V

and features to be selected are available [25]. Since we have a reduced number of samples, we can use "wrapper" [26] methods. Moreover, these feature selection methods include a classification step, needing a labeled database. They provide more suitable selected feature set. The following reference methods are used:

- ReliefF [27]
- SFS : Sequential Feature Selection [28]
- SBS : Sequential Backward Selection [29]
- SFFS : Sequential Forward Floating Selection [30]
- SBFS: Sequential Backward Floating Selection [30]

For each dataset, by combining the results given by the algorithms of selection, a subset of 10 features was extracted. This subset gathers the most discriminating features for the separation of "switched on microphone" classes (positive images) and "rubbish" classes (negative images). The table III shows the five best features selected by each method for each video from our databases. The features from the HSV color space are those which characterize efficiently the microphones' LED. The interest of using many methods is that the features selected are confirmed by all of them, as shown in table III.

The figure 5 summarizes the percentage of appearance of each characteristic, video dataset and algorithms of selection altogether. In other words, it is the number of occurrences of each characteristic in the results given by the 5 algorithms of selections of characteristics applied on the images extracted from the 4 videos of the dataset. The results show that the features in the HSV color space, especially the Hue are important for the characterization of active microphones. The low number of selected features of component L is due to the fact that the information is redundant with that of component V.

The seven recurring features to each picture library have been selected[2]:

- Mean, variance and third central moment of the H component
- Mean of the S component,
- Mean, variance and the V component's Kurtosis fourth root.

[2]Features in green in the figure 5.

These are the features that best differentiate an on-air microphone from an off microphone. These results show the importance of taking into account the average and dispersion of the microphone's hue and luminance for class separation.

These characteristics are thus used during the creation and during the use of the decisional tree for microphone's presence check.

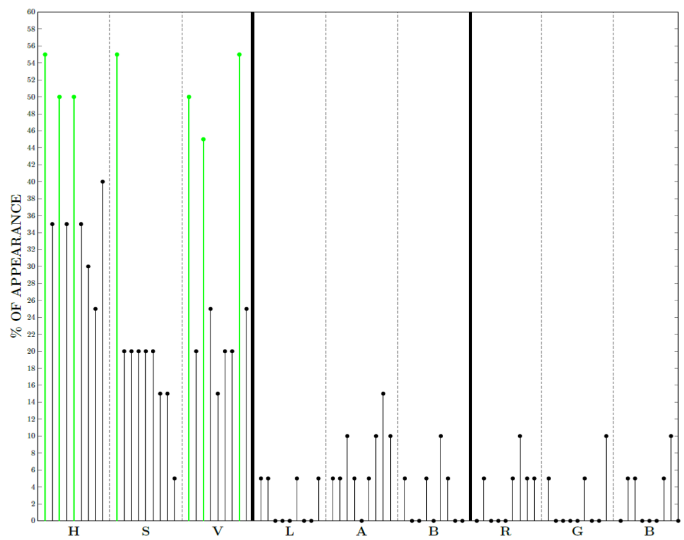

Fig. 5. Percentage of appearance of the features from the 5 features selection algorithms apply on the 4 datasets. In green are represented the selected features.

III. VALIDATION

The evaluation of the proposed method was performed on a computer equipped with an Intel Core I7-5557U processor (Base Frequency 3.1 GHz) and 8GB RAM. The method was implemented in Python, using functions from OpenCV library.[3]

A. Videos database

The evaluation of the method was performed on videos from a city council, that was filmed in Villers-Lès-Nancy[4]

[3]opencv.org
[4]villerslesnancy.citizencam.fr

in France. Table I presents different videos that were used. Original images have a size of 1920x1080 pixels for the video ”Sight 1” and 1080x720 pixels for the others. Since videos 2, 4 and 6 have similar characteristics, we can group them under the name DB2. The size of the zone corresponding to the region of interest is manually defined during the initialization step. Calculating times are really impacted by the definition of these zones. The microphone's area is the average number of pixels in the reduced area (19x19 pixels) around the microphone: 361 pixels maximum. The properties are each sighting's features. These test sequences were manually annotated to serve as the ground truth in order to estimate the efficiency of the method.

B. Classifiers

As we can see in I, there are two different types of views. The first one,“Sight 1”, is a camera directed at the mayor, with a higher resolution. The other cameras film advisors and have lower resolution. That's why we use two different decision trees trained in an initialization step. The first one is trained from 500 images from the camera 1 and is then applied one the entire sequence of “Sight 1”. The second one is trained from 500 images from the camera 2 and applied on the 3 others sequences: sight 2 , 4 and 6 (DB2).

C. Results

The table IV sums up the results obtained by using our method. The results are obtained thanks to the confusion matrix and are expressed in terms of precision, recall and accuracy [31].

TABLE IV
RESULTS

Name	Precision	Recall	Accuracy	P.T.I. (ms)
Sight 1	100 %	97.65 %	98.20%	33
Sight 2	98.47 %	94.27%	99.18%	19
Sight 4	99.72 %	100 %	99.93%	12
Sight 6	100 %	99.63%	99.89%	27

The processing time per image (P. T. I.) expresses the average time (in milliseconds) needed to detect the presence of an active microphone in an image. This time depends on the search area defined during initialization.

The results obtained show the performance of the proposed method on the selected videos. The precision and the accuracy are in a 99% range and the recall about 98%. These results confirm the effectiveness of the features' selection presented in part II-C. The processing time per image shows the ability to make it work in real time. In addition, the results obtained on DB2 show that it is possible to generate a general model, from one camera, for processing multiple cameras. The majority of false positives obtained are caused by light disturbance (reflections, smartphones,...). A longer learning time could maybe reduce those errors. False negatives are caused by a total occlusion of the microphone (when the speaker holds the microphone by its LED).

The presented results are obtained by using the proposed detection method on each image, without consideration of the video sequence's dynamics. In the context of automatic editing, detection and loss of detection under 300ms are ignored. By doing so, occlusions and false positives are not felt by the user.

IV. CONCLUSION

The automatic video selection method offered is reliable in a council context and the consideration of the dynamics of the scene must improve the obtained results. we have selected seven features (mean, variance and third central moment of the H component, mean of the S component, as well as the mean, variance and the V component's Kurtosis fourth root) that make it possible to obtain a microphone model that can be used for different cameras but have close viewing characteristics. This method is compatible with a real-time utilization and the selection of the stream of interest is then possible. Each time a new person speaks, we can change the stream being relayed in order to show this new speaker.
However we're planning several solutions to improve it.

An upper-body detector should allow to look for microphones only where the eventual speakers are. Instead of a global research zone, we could obtain one zone for each speaker. It would result in a diminution of the computational cost and a decrease of false detection

Taking into account the phenomena of scale should make it possible to generalize for all shots taken within a council, or even different installations.

The use of microphones's LED as an informative medium [32] is also considered. The microphone will become an active component of the system, communicating an identification signal, owned by each speaker, that will be captured by the cameras. It will allow the verification of the detection as well as the identification of the speaker.

REFERENCES

[1] J. Chen and P. Carr, “Autonomous Camera Systems: A Survey,” in *Workshop on Intelligent Cinematography and Editing*, pp. 18–22, 2014.

[2] T. Yamada, S. Nakamura, and K. Shikano, “Robust speech recognition with speaker localization by a microphone array,” in *, Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings*, vol. 3, pp. 1317–1320 vol.3, Oct. 1996.

[3] L. J. Rodríguez, M. Peñagarikano, and G. Bordel, “A Simple But Effective Approach to Speaker Tracking in Broadcast News,” in *Pattern Recognition and Image Analysis*, Lecture Notes in Computer Science, pp. 48–55, Springer, Berlin, Heidelberg, June 2007.

[4] N. Madhu and R. Martin, “A scalable framework for multiple speaker localization and tracking,” in *In Proceedings of the International Workshop for Acoustic Echo Cancellation and Noise Control (IWAENC 2008*, 2008.

[5] N. Madhu and R. Martin, “A Versatile Framework for Speaker Separation Using a Model-Based Speaker Localization Approach,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 1900–1912, Sept. 2011.

[6] R. Cutler and L. Davis, “Look who's talking: Speaker detection using video and audio correlation,” in *2000 IEEE International Conference on Multimedia and Expo.*, vol. 3, pp. 1589–1592 vol.3, 2000.

[7] E. D'Arca, N. M. Robertson, and J. R. Hopgood, "Look who's talking: Detecting the dominant speaker in a cluttered scenario," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1532–1536, May 2014.

[8] K. Nakadai, K. Hidai, H. G. Okuno, and H. Kitano, "Real-time speaker localization and speech separation by audio-visual integration," in *Proceedings 2002 IEEE International Conference on Robotics and Automation*, vol. 1, pp. 1043–1049 vol.1, 2002.

[9] G. Lathoud, J.-M. Odobez, and D. Gatica-Perez, "AV16.3: An Audio-Visual Corpus for Speaker Localization and Tracking," in *Machine Learning for Multimodal Interaction*, Lecture Notes in Computer Science, pp. 182–195, Springer, Berlin, Heidelberg, June 2004.

[10] C. Busso, S. Hernanz, C.-W. Chu, S.-i. Kwon, S. Lee, P. G. Georgiou, I. Cohen, and S. Narayanan, "Smart room: Participant and speaker localization and identification," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 2, pp. 1117–1120, Mar. 2005.

[11] H. Vajaria, T. Islam, S. Sarkar, R. Sankar, and R. Kasturi, "Audio Segmentation and Speaker Localization in Meeting Videos," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 2, pp. 1150–1153, 2006.

[12] Q. Liu, Y. Rui, A. Gupta, and J. J. Cadiz, "Automating camera management for lecture room environments," (Seattle), pp. 442–449, Jan. 2001.

[13] H. Tae-Hyun, J. In-Hak, and C. Seong-Ik, "Detection of Traffic Lights for Vision-Based Car Navigation System," in *Advances in Image and Video Technology*, pp. 682–691, Springer, Berlin, Heidelberg, Dec. 2006.

[14] Y.-C. Chung, J.-M. Wang, and S.-W. Chen, "A vision-based traffic light detection system at intersections," *Journal of Taiwan Normal University: Mathematics, Sciences & Technology*, vol. 47, no. 1, pp. 67–86, 2002.

[15] M. Diaz-Cabrera, P. Cerri, and P. Medici, "Robust real-time traffic light detection and distance estimation using a single camera," *Expert Systems with Applications*, vol. 42, no. 8, pp. 3911–3923, 2015.

[16] M. Omachi and S. Omachi, "Detection of traffic light using structural information," in *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*, pp. 809–812, Oct. 2010.

[17] G. H. Joblove and D. Greenberg, "Color Spaces for Computer Graphics," in *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '78, (New York, NY, USA), pp. 20–25, ACM, 1978.

[18] I. 11664-4, "ISO 11664-4: 1976 L* a* b* Colour Space," *Joint ISO/CIE Standard, ISO*, pp. 11664–4, 2008.

[19] C. Fiorio and J. Gustedt, "Two linear time Union-Find strategies for image processing," *Theoretical Computer Science*, vol. 154, pp. 165–181, Feb. 1996.

[20] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, Mar. 1986.

[21] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 23–38, Jan. 1998.

[22] H. Zhang, A. C. Berg, M. Maire, and J. Malik, "SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 2126–2136, 2006.

[23] E. Osuna, R. Freund, and F. Girosit, "Training support vector machines: An application to face detection," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 130–136, June 1997.

[24] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence*, vol. 2, pp. 1137–1143, 1995.

[25] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, no. 3, pp. 1157–1182, 2003.

[26] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, pp. 273–324, Dec. 1997.

[27] K. Kira and L. A. Rendell, "The Feature Selection Problem: Traditional Methods and a New Algorithm.," in *Proceedings Tenth National Conference on Artificial Intelligence*, pp. 129–134, Jan. 1992.

[28] A. W. Whitney, "A Direct Method of Nonparametric Measurement Selection," *IEEE Transactions on Computers*, vol. C-20, pp. 1100–1103, Sept. 1971.

[29] T. Marill and D. Green, "On the effectiveness of receptors in recognition systems," *IEEE Transactions on Information Theory*, vol. 9, pp. 11–17, Jan. 1963.

[30] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, pp. 1119–1125, Nov. 1994.

[31] J. Davis and M. Goadrich, "The Relationship Between Precision-Recall and ROC Curves," in *Proceedings of the 23rd International Conference on Machine Learning, ACM*, vol. 06, June 2006.

[32] C. Danakis, M. Afgani, G. Povey, I. Underwood, and H. Haas, "Using a CMOS camera sensor for visible light communication," in *2012 IEEE Globecom Workshops*, pp. 1244–1248, Dec. 2012.