EG UK Computer Graphics & Visual Computing (2018)
G. Tam and F. Vidal (Editors)

Combining accumulated frame differencing and corner detection for motion detection

N. Algethami and S. Redfern

National University of Ireland Galway

Abstract

Detecting and tracking people in a meeting room is very important for many applications. In order to detect people in a meeting room with no prior knowledge (e.g. background model) and regardless of whether their motion is slow or significant, this paper proposes a coarse-to-fine people detection algorithm by combining a novel motion detection process, namely, adaptive accumulated frame differencing (AAFD) combined with corner features. Firstly, the region of movement is extracted adaptively using AAFD, then motion corner features are extracted. Finally, the minimum area rectangle fitting these corners is found. The proposed algorithm is evaluated using the AMI meeting data set and this indicates promising results for people detection.

CCS Concepts

•***Computing methodologies*** → ***Tracking;*** *Motion capture;*

1. Introduction

Meetings are important events in any organisation. They are essential for knowledge sharing, knowledge creation, information exchange and informed decision making. For a variety of reasons, people may not be able to attend a meeting or may miss important information even if they do attend. Note taking is one of the possible solutions; however, this can be subjective and inaccurate as well as laborious. To address this problem, a meeting recorder is needed to enable future reviewing.

Meeting recording and understanding has attracted much attention for research on a diverse range of technologies. These technologies range from physical capturing to meeting analysis and semantic processing. Meeting capture records different types of data needed for meeting analysis such as video, audio, and text data. Meeting analysis is a low-level processing layer to analyse the captured data from the meeting capture module, while semantic processing is a high-level layer which is responsible for handling semantic manipulation such as browsing. Examples of the application of meeting technologies include structuring and browsing meeting databases [YN10].

For structuring meeting application (meeting analysis), people detection and tracking is necessary. Localising and tracking people plays a fundamental role in meeting analysis. The results of visual tracking can be used in many applications (for example, surveillance applications or as a cue for meeting browser navigation).

In this paper, we describe a novel coarse-to-fine people detection algorithm using a combination of our developed motion detection algorithm, namely, adaptive accumulated frame differencing (AAFD), and Shi-Tomasi corner detection. This coarse-to-fine approach allows robust blunder recovery, since each video frame is considered separately by the AAFD step. The basic idea of our algorithm is that, firstly, the region of movement is extracted using AAFD, and then corner detection is applied on this region. These corners are further processed and only moving corners are considered. Finally, minimum area rectangles fitting them are found. Our goal is to combine techniques that use totally different evidence (i.e. one is temporal while the other is spatial), while working in an environment with no prior knowledge and in which meeting participants may be moving very little (e.g. while seated) or relatively fast (while walking).

2. Related Work

Moving object detection is the first and most important step in video analysis. Any tracking algorithm needs an object detection method applied in every frame or when an object first appears on the video. This is the process of separating foreground objects from the background [BK17]. The object motion is important source of information for detection. In recent years, various approaches for moving object detection and tracking have been proposed, within various application domains. These approaches rely on background subtraction, frame differencing and/or optical flow algorithms.

Optical flow refers to the flow vectors of moving objects, which indicates the speed of movement of pixels in subsequent frames. It indicates velocity and the direction of pixel movements [AAA16] [Far03]. [HGL*15] combined optical flow with three-frame differencing to detect moving objects. Optical flow can detect objects

DOI: 10.2312/cgvc.20181202

with no prior knowledge; however, it does not perform adequately when objects are moving slowly.

Background subtraction is a commonly used technique to detect moving objects [VVV15] [SP18] [YGJ13] [MS18]. This involves firstly building a background model using a set of images, and then calculating the difference between the current frame and the background model, in order to detect the objects of interest. The background must be modelled accurately, with frequent updates to consider changes in the background such as changes in lighting conditions, scene geometry or moving objects (e.g., trees shaken by the wind) [MDP12]. Gaussian mixture modelling (GMM) is one of the most common adaptive background modelling techniques, whereby each pixel is modelled as a mixture of Gaussians and the background model is updated to cope with scene changes [SG99]. However, this technique remains challenging for slow moving objects, and in particular is sensitive to speed changes [Ziv04].

Traditional frame differencing captures the change between two consecutive frames by calculating the absolute difference between the two frames [LD07]. Most of the existing frame differencing methods [ZL01] extract moving objects using a single temporal scale [ZZZ*15]. Therefore, frame differencing can only detect fast moving objects, while slow moving objects will not be detected.

Little work has been done to detect slow-fast moving objects [FC13] [CU12] [GWB*17] [AAA16] [AR17]. [FC13] present a moving object detection algorithm using spatio-temporal information and marked watershed. [CU12] propose an image segmentation algorithm using background subtraction and Expectation Maximization (EM), where each pixel is classified, then the slow moving object is located and its shadow is distinguished from the moving object using modified background subtraction. A combination of background subtraction and frame difference is used to solve the problem of incomplete detection for the moving object [GWB*17]. Cumulative frame differencing (CFD), which involves the summation of successive frame differencing, is presented by [AAA16] to improve the detection of slow moving vehicles. Accumulated frame differencing using temporal window size to detect people in a meeting video is proposed by [AR17].

In many real life scenarios, moving objects differ in their speed and size, therefore a motion detection method which adaptively uses temporal window size to detect objects regardless of whether their motion is significant or small is needed. Existing methods do not have the adaptability to detect and analyze moving objects using different temporal scales. This paper proposes a novel motion detection algorithm which combines our motion detection algorithm (AAFD) and shape features using Shi-Tomasi corner detection. Our approach differs from previous work via two novel features, the first of which is the use of accumulated frame differencing using different temporal window sizes based on analysis of object shape features. For example, a large temporal window size is used to detect slow moving objects and a small window size is used to detect fast moving objects.

[ZZZ*15] propose a spatial temporal detection and tracking algorithm which calculates temporal-spatial windows for each object using octree decomposition of the temporal-spatial domain. It works well on moving platforms with a uniform linear motion. However, when the object is moved in arbitrary directions as well as cases where the object is hidden in the scene are not considered. In a meeting context, meeting participants move in all directions as there is no restriction on their movement. [LD07] present an adaptive accumulated frame differencing to extract objects from head and shoulder video sequences. Firstly, frame difference FD is divided into blocks (8*8 pixels). Then sum of FD within a block is calculated and used as criterion for motion analysis. For each block, FD is accumulated using different a number of frames based on its motion attributes. Finally, thresholding and post processing is applied to segment object. They focus on accurate segmentation of objects from a frontal camera. Our method however considers the detection and tracking of multiple people from overhead cameras.

The second novel contribution in our approach is that we propose a coarse-to-fine motion detection which performs detection and tracking simultaneously - in most of the existing work, the object is firstly detected then it is tracked over subsequent frames. Frame differencing is an important coarse robust step which does not need prior knowledge, therefore, our overall system can recover from blunders. The problem with a lot of techniques is that they assume correctness from one frame to another, and this can go badly when things drift, so we use frame differencing to enable recovery.

3. Combining accumulated frame differencing and corner detection for motion detection

Our contribution at this stage is to implement a robust people detection and tracking algorithm. For this we combine our motion based detection algorithm 'adaptive accumulated frame differencing (AAFD)' technique with a feature based detection algorithm using Shi-Tomasi corner detection [ST94] [HS88]. Frame differencing data is a certain type of evidence describing how much pixels are changing over time, whereas corner detection is completely different evidence which searches for discontinuity in the pixel value or the edges, considered in a single frame. These are totally different forms of evidence and therefore are appropriate to use in combination.

3.1. Outline of combination of AAFD and corner detection technique

Figure 1 provides a block diagram of our approach to detecting and tracking meeting participants. Our contribution at this point is to find the region of movement using our developed AAFD algorithm robustly, since it was proven that it gives a good performance in finding the people in a meeting video sequence given no prior knowledge, whether their motion is fast or slow. Frame differencing is a robust coarse step that helps to recover from blunders. We use standard morphological techniques to extract 'blobs', i.e. regions of connected pixels, which are assumed to indicate coherent moving objects. Various phenomena in meeting videos cause objects to appear as split or merged blobs, and therefore additional processing is required, corner features are detected by applying the Shi-Tomasi corner detector. These corner points are further processed and moving corner points are returned. Finally, the minimum area rectangle fitting these corner points is found.

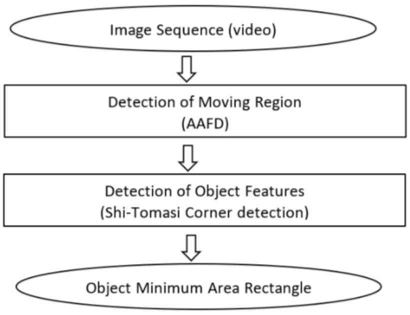


Figure 1: *Block diagram of combination of AAFD and corner detector for people detection.*

3.1.1. Detection of Moving Region (AAFD). The region of movement is extracted using AAFD, where object motion is segmented adaptively based on analysis of object shape features. A large temporal window size is used to accumulate frame differencing data for slow moving objects, while fast moving objects are segmented using a small temporal window. Once the object is detected, the region of movement is extracted around the center of the object.

Outline of the algorithm Our algorithm performs detection of people based on three stages:

1. People detection is applied based on an accumulated frame differencing image using a large temporal window size. Starting with a large window size allows the robust segmentation of all foreground pixels.
2. For each detected blob, motion analysis using the shape features of the blob is applied. Two shape features, fill ratio and blob area, are used to detect 'good' and 'bad' blobs. Fill ratio can be used to assess a blob, whereby good blobs are assumed to be more square and their desirable fill ratio should be closer to 1 than zero. If the fill ratio (area of the blob divided by the area of the blob's bounding box) is smaller than a defined threshold, we conclude the blob was merged or is a 'bad' blob. In our experiments, blob area appears to be a suitable feature for rejecting small blobs (noise) and large blobs (merged objects).
3. Finally, the detection is executed again with a different temporal window size based on the shape feature of the blob. See algorithm 1

compression artefacts Unfortunately, lossy compression artefacts are amplified during AFD image calculation; this noise appears as white blocks in the difference image. This in turn increases the probability of wrong detection of the region of movement. We use a coefficient of motion percent of region of interest (ROI) to reduce these compression artefacts by:

1. Firstly, ROI(x,y) is converted to a binary image:

$$ROI_b(x,y) = \begin{cases} 255, if ROI(x,y) > thresh \\ 0, otherwise \end{cases} \quad (1)$$

2. Secondly,

$$\text{motion percent} = \frac{\text{Number of non zero pixels}}{\text{total number of pixels}} \quad (2)$$

When motion percent is greater than a threshold value, the region of movement is extracted based on AAFD. Otherwise, the region of movement is extracted based on the previous detection (as the ROI has no blobs, only noise). As shown in Figure 3, using motion percent gives robust results to control region of movement extraction based AAFD.

Algorithm 1 Adaptive Accumulated Frame Differencing (AAFD)

1. Apply people detection based on AFD using a window size of 100 frames (Figure 2).
2. Filter blobs by area and only accept area greater than a threshold.
3. For each blob, apply bad blob detection using the shape features of the blob

if $BLOB-FILL-RATIO >= BLOB-FILL-RATIO-CUTOFF$ **then** {//good blob}
  if BLOB-AREA >= GOOD-BLOB-AREA **then**
    Apply second detection using smaller window
  else
    Accept good blob
  end if
else {//blob is not good}
  if $BLOB-AREA >= BAD-BLOB-AREA$ **then**
    Apply second detection using smaller window
  else
    Apply second detection using larger window
  end if
end if

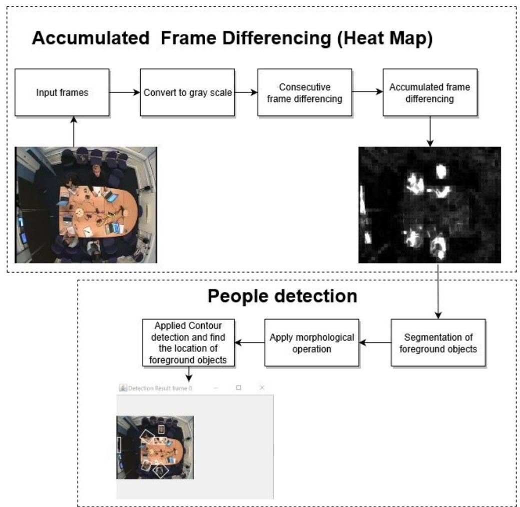


Figure 2: *Accumulated frame differencing algorithm.*

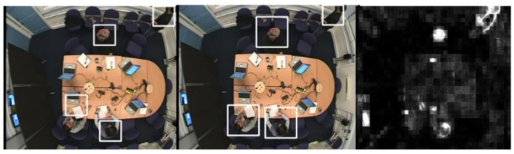


Figure 3: *shows examples of region of movement extraction with/without motion percent (Frame 4921), where the first image is region of movement extraction without using motion percent, the second image is region of movement extraction based on motion percent and the third image is accumulated frame differencing window size = 100. Using motion percent reduces the error resulting from noise of compression artefacts.*

3.1.2. Detection of Object Features (Shi-Tomasi corner detection).

The N strongest corners are found by applying the Shi-Tomasi corner detection method to the region of movement identified by AAFD. Firstly, it calculates the minimal eigenvalue of the covariance matrix of derivatives at every pixels. Then, all corners with the minimal eigenvalue below a threshold value are rejected. After that, the remaining corners are sorted in descending order based on their minimum eigenvalue. The strongest corners are returned, then based on specified minimum distance between corners, nearby corners are rejected and the N strongest corners are returned [ST94] [Od18].

Combining corners with motion As our goal is to track meeting participants, we are interested in the detection of motion corners (corners on foreground objects) and in avoiding erroneous detection (rectangle on background). The issue to be solved is how to relate the detected corner points to motion. Our contribution at this point is to propose a metric to measure the motion of a corner point. We implemented a scoring function to assign a score value for each corner point based on its corner value and AFD data. For all N corner points $C_N(x,y)$, the score S(N) is defined as:

$$S(N) = C_N(x,y) * FD_N(x,y)^x \qquad (3)$$

Where $C_N(x,y)$ is the minimum eigenvalue and $FD_N(x,y)$ is AFD data which ranges from 0 (black) to 255 (white), and x is an empirically determined constant.Figure 4 shows an example of using this score function to reduce background corner points.

To consider only the motion corners, all N corners are sorted in descending order based on their score value. Then, the highest N corners are returned and the minimum area rectangle fitting them is calculated to update the center to extract the next region of interest.

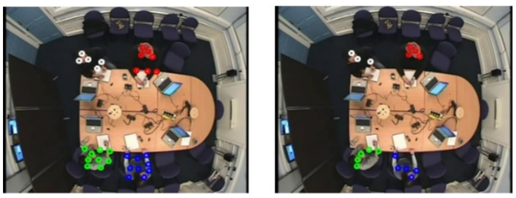


Figure 4: *Sample images to show corner detection without using score function(left) and using score function (right). As shown in Frame 3990, background corner points are relatively reduced using score function.*

3.2. The algorithm of a combination of AAFD and corner detection technique

Algorithm 2 combining accumulated frame differencing and corner detection for people detection

```
for each frame in a video do
    //Moving regions detection
    Extract region of interest (ROI) based on previous centre
    Calculate motion percent from AFD image
    if motionpercent < Thresh then
        extract region of movement based on previous centre[ // there are no blobs only noisy white blocks]
    else
        extract region of movement (ROM) based on the AAFD
    end if
    // Object corner features detection
    Apply corner detection on ROM
    for each corner point do
        Calculate its score
    end for
    Sort corner points based on the score in descending order.
    Return the highest N corners where score != 0
    if (numberofcornerpoints! = 0 then {//good corners are found}
        Find the minimum area rectangle fitting the corner points.
        Update centre based on the centre of detected corners.
        Keep the corner points as previous corners.
        Draw corner points.
    else
        draw previous corner points
    end if
end for
```

3.3. Experimental results on the AMI video corpus

The proposed method was implemented using OpenCV 3.1 and a Java Platform. The AMI video meeting data set was used [MCK*05] to evaluate the performance of our proposed approach for people detection and tracking. Our proposed approach was applied to the ES202a video frames in the AMI meeting corpus captured using the overhead camera. The first section below (3.3.1) discusses evaluation of our AAFD component separately and the second section (3.3.2) illustrates the result of our coarse-to-fine people detection algorithm using AAFD and Shi-Tomasi corner detector.

3.3.1. Evaluation of Adaptive Accumulated Frame Differencing (AAFD).

As depicted in Figure 5, using blob fill ratio and blob area to detect a bad blob and applying detection again using a different window size based on the shape features of the blob reduces detection problems when using a single window AFD algorithm such as blob merging. It segments blobs in large motion robustly using smaller window size and it robustly detects bad blobs. Applying a second detection gives more accurate detection as the ROI (Region of Interest) is reduced compared with the whole frame region. For example, in Figure 5, blob 0, blob 1, blob 3 and blob

4 are classified as good blobs and accepted, with no need for second detection. A second detection is applied on blob 5 as its area is large. As a result, more accurate detection is achieved: see Figure 5

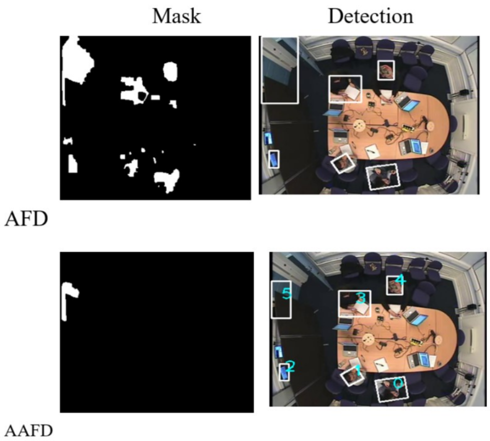


Figure 5: *Adaptive accumulated frame differencing.Frame 300 shows that blob 0, blob 1, blob 3, and blob 4 were detected correctly and accepted as good blobs, while a second detection was applied on blob 5 using a small window to segment this object in fast motion.*

3.3.2. Evaluation of the combination of AAFD and corner detection. The performance of our detection and tracking algorithm has, to date, been evaluated visually (qualitatively). As depicted in figures 6, 7 and 8, the qualitative result highlights a good performance of the proposed algorithm in different scenarios.

Figure 6 illustrates the detection result using our algorithm which shows a slow movement scenario as all people are sitting around the table making very small motions. The first column provides the original frame, the second column shows the result of region of movement extraction using AAFD and the last column presents the final detected corner points.

Figure 7 demonstrates the case where there is varying amounts of motion. For example, in Frame 5873 and Frame 6456 one person is standing at the white board while others are sitting. The person starts moving to the whiteboard in Frame 6968. Frame 6893 shows a person returning to his seat and another one starting to move to the white board.

Figure 8 shows an example of detection when people are near to each other.

The proposed method gives robust results to detect and track people in a meeting room with no prior knowledge and whether their motion is slow or fast. However, the proposed method does not consider overlapping between people (where one person's movement is occluded by another's).

4. Conclusion and ongoing work

An approach for the detection and tracking of meeting participants by combining adaptive accumulated frame differencing (AAFD) and corner features is presented in this paper. We firstly extract region of movements for each person adaptively using AAFD, where a large temporal window size is used to find slow moving objects and a small temporal window size is used to detect fast moving objects based on analysis of the object shape features. Secondly, Shi-Tomasi corner detection is applied. Then, the score for each corner based on its accumulated frame differencing data is calculated and, finally, the corners with high scores are returned and the minimum area rectangle fitting them is found.

We evaluate performance visually, which demonstrates that the proposed method can robustly detect and track people in meeting videos with no prior knowledge and whether their motion is slow or fast. Overlapping regions of interest (when people are very close together or walk past each other) remain a challenge to this approach. We are currently investigating methods to cope with overlapping; one promising approach is to use the vector dot product of (i) a person's normalised tracked velocity and (ii) the normalised velocity indicated by optical flow. This allows us to reject extracted corner points if the direction of movement suggested by optical flow disagrees with the known direction of movement of the person over a number of previous frames.

In addition to visual evaluation, performance can be evaluated quantitatively, which is more time consuming and is part of our current efforts. First of all, a valid ground truth needs to be established, which is the exact output that the algorithm should produce. In our algorithm, the ground truth is the determination of real-world object positions. This process is usually done by human beings, yet it is relatively difficult to achieve exact results from the same data by different people or even the same person at different times. It is also a time consuming process [MDP12] [HKL*01]. Labelling the ground truth of people's positions and implementing quantitative measurement is in progress.

For future work, we aim to use the detection result to implement a timeline of people's movements in meeting videos and use this information to build a robust background model.

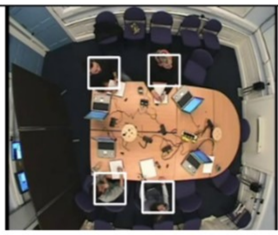
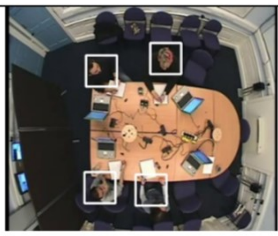
Frame No	Original frame	Region of Movement	Detection results
3900			
4029			

Figure 6: *Detection results for slow movement (i.e. people are sitting). Left to right, frame number, original frame, region of movement, detection results.*

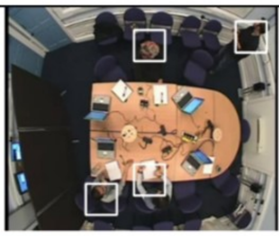
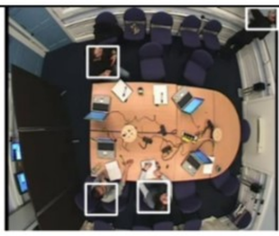
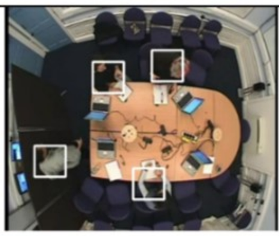
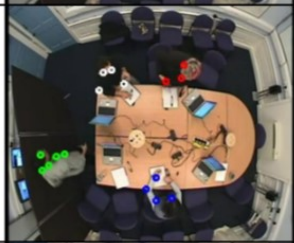
Frame No	Original frame	Region of Movement	Detection result
5873			
6456			
6968			
6893			

Figure 7: *Detection results for different motions of people (i.e. low motion when people are sitting and significant motion when they are moving in a meeting room).*

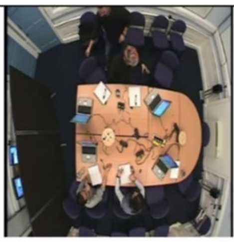
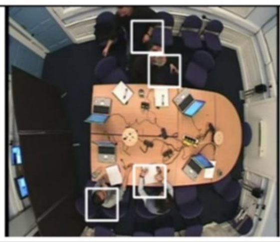
Frame No	Original frame	Region of Movement	Detection result
6166			

Figure 8: *Sample image to show detection result when meeting participants are near each other.*

References

[AAA16] ALSMADI M., ABDULRAHIM K., ABDULSALAM R.: Cumulative frame differencing for urban vehicle detection. In *First International Workshop on Pattern Recognition* (2016), vol. 10011, International Society for Optics and Photonics, p. 100110G. 1, 2

[AR17] ALGETHAMI N., REDFERN S.: Detecting and tracking meeting participants using motion heat maps. In *Irish Machine Vision and Image Processing Conference Proceedings 2017*, McDonald J., Markham C., Winstanley A., (Eds.). Irish Pattern Recognition &Classification Society, Maynooth university, 2017, pp. 99–106. 2

[BK17] BALAJI S., KARTHIKEYAN S.: A survey on moving object tracking using image processing. In *Intelligent Systems and Control (ISCO), 2017 11th International Conference on* (2017), IEEE, pp. 469–474. 1

[CU12] CHINCHKHEDE D., UKE N.: Image segmentation in video sequences using modified background subtraction. *International Journal of Computer Science & Information Technology 4*, 1 (2012), 93. 2

[Far03] FARNEBÄCK G.: Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis* (2003), Springer, pp. 363–370. 1

[FC13] FU Q., CELENK M.: Video object segmentation using spatio-temporal information and marked-watershed operation. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)* (2013), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), p. 1. 2

[GWB*17] GUO J., WANG J., BAI R., ZHANG Y., LI Y.: A new moving object detection method based on frame-difference and background subtraction. In *IOP Conference Series: Materials Science and Engineering* (2017), vol. 242, IOP Publishing, p. 012115. 2

[HGL*15] HAN X., GAO Y., LU Z., ZHANG Z., NIU D.: Research on moving object detection algorithm based on improved three frame difference method and optical flow. In *Instrumentation and Measurement, Computer, Communication and Control (IMCCC), 2015 Fifth International Conference on* (2015), IEEE, pp. 580–584. 1

[HKL*01] HU J., KASHI R., LOPRESTI D., NAGY G., WILFONG G.: Why table ground-truthing is hard. In *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on* (2001), IEEE, pp. 129–133. 5

[HS88] HARRIS C., STEPHENS M.: A combined corner and edge detector. In *Alvey vision conference* (1988), vol. 15, Citeseer, pp. 10–5244. 2

[LD07] LENG B., DAI Q.: Video object segmentation based on accumulative frame difference. In *Broadband Network & Digital Media Lab of Dept.* (2007), Automation in Tsinghua University, 100084, Beijing, P.R.China. 2

[MCK*05] MCCOWAN I., CARLETTA J., KRAAIJ W., ASHBY S., BOURBAN S., FLYNN M., GUILLEMOT M., HAIN T., KADLEC J., KARAISKOS V., ET AL.: The ami meeting corpus. In *Proceedings of the 5th International Conference on Methods and Techniques in Behavioral Research* (2005), vol. 88, p. 100. 4

[MDP12] MARTÍNEZMARTÍN E., DEL POBIL A. P.: *Robust motion detection in real-life scenarios*. Springer Science & Business Media, 2012. 2, 5

[MS18] MAHALINGAM T., SUBRAMONIAM M.: A robust single and multiple moving object detection, tracking and classification. *Applied Computing and Informatics* (2018). 2

[Od18] OPENCV-DOCUMENTATION: Goodfeaturestotrack open cv method at: Opencv3.0.0, documentation, org.opencv.imgproc package, imgproc class., 2018. (accessed April 8, 2018). URL: `http://docs.opencv.org/java/3.0.0/`. 4

[SG99] STAUFFER C., GRIMSON W. E. L.: Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.* (1999), vol. 2, IEEE, pp. 246–252. 2

[SP18] SUPREETH H., PATIL C. M.: Efficient multiple moving object detection and tracking using combined background subtraction and clustering. *Signal, Image and Video Processing* (2018), 1–9. 2

[ST94] SHI J., TOMASI C.: Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on* (1994), IEEE, pp. 593–600. 2, 4

[VVV15] VASUHI S., VIJAYAKUMAR M., VAIDEHI V.: Real time multiple human tracking using kalman filter. In *Signal Processing, Communication and Networking (ICSCN), 2015 3rd International Conference on* (2015), IEEE, pp. 1–6. 2

[YGJ13] YE Q., GU R., JI Y.: Human detection based on motion object extraction and head–shoulder feature. *Optik-International Journal for Light and Electron Optics 124*, 19 (2013), 3880–3885. 2

[YN10] YU Z., NAKAMURA Y.: Smart meeting systems: A survey of state-of-the-art and open issues. *ACM Computing Surveys (CSUR) 42*, 2 (2010), 8. 1

[Ziv04] ZIVKOVIC Z.: Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (2004), vol. 2, IEEE, pp. 28–31. 2

[ZL01] ZHANG D., LU G.: Segmentation of moving objects in image sequence: A review. *Circuits, Systems and Signal Processing 20*, 2 (2001), 143–183. 2

[ZZZ*15] ZHANG T., ZHOU G., ZHANG C., LI G., CHEN J., LIU K.: A novel temporal-spatial variable scale algorithm for detecting multiple moving objects. *IEEE Transactions on Aerospace and Electronic Systems 51*, 1 (2015), 627–641. 2