Visual- cognizance to divulge scarcely moving objects based on local motion stabilization

Sheela B. P
sheelabp@gmail.com
Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka

Ashwini
ashwini.is18.rymec@gmail.com
Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka

Sushma Patil
sushmapatif743@gmail.com
Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka

Dr. Sreepathi .B
sreepathib@gmail.com
Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka

ABSTRACT

Motion is one of the most important sign to separate foreground objects from background in a video. Using a stationary camera, it is usually assumed that the background is static while the foreground objects are moving most of the time. In practical, the foreground objects may show scarcely motions, such as impulsive objects and sleeping persons. Meanwhile, the background may contain frequent local motions, such as waving trees or grass. Such complexities may prevent the existing background subtraction algorithms from correctly identifying the foreground objects. We propose a new approach that can detect the foreground objects with frequent or scarcely motions. Specifically, we use a visual attention mechanism to deduce a complete background from a subset of frames and then propagate it to the other frames for accurate background subtraction. Furthermore, we develop a feature-matching based local motion stabilization algorithm to identify frequent local motions in the background for reducing false positives in the detected foreground. The proposed approach is fully unsupervised without using any supervised learning for object detection and tracking. Foreground detection technologies have emerged as an important research area with increasing popularity of computer vision and camera devises. This study proposes a method that can integrate arbitrary detection technologies to detect foreground in real time.

Keywords: Object detection, Scarcely moving objects, Visual, Local Motion Stabilization.

1. INTRODUCTION

Nowadays moving object detection has become a very prime area for research due to its use in various computer vision applications. Beside from the vital benefit of being able to differentiate video streams into moving and background content, detecting moving objects provides a purpose of attention for recognition, classification and activity scrutiny making these later steps more effective. This research paper presents the thorough survey of background subtraction methods for object detection with a brief information about other methods for object detection. In many videos, it is usually needed to separate foreground object of interest, which can be persons, vehicles, animals and so on. Based on the extracted foreground objects, high-level tasks, such as detecting objects and recognizing activities from videos, can be addressed more effectively. Assuming that the camera is stationary, motion plays a key role in video-based foreground/background separation: foreground objects are usually moving while the background is relatively static. Many approaches, such as optical flow and background subtraction, have been developed to detect the motions of foreground objects, based on which the foreground and the background can be separated. Optical flow requires that the foreground objects move all the time. However, in practical, foreground objects may show scarcely motions, e.g., deserted objects, removed objects, and persons stopping for a while and then walking away. As an example shown in Fig.1, we first run the NIPE-based and the GMM-based background subtraction method that output a background subtraction map respectively. Second, these background subtraction maps are combined via a foreground AND operator as attention mechanism to finally get a more robust and accurate background subtraction map. This is a natural fusion because the NIPE-based method using block feature and the GMM-based method using pixel feature have complementary strengths, in which the NIPE-based method is insensitive to the noises and the small movement of dynamic scenes and the GMM-based method contains rich information. Third, we use a connected components algorithm to give the green bounding boxes of the labeled foreground pixels. Finally, at each extracted green bounding box, we compute local trimap matting via a heuristic seeds selection scheme, in which the labeled foreground pixels in the green bounding box are used as the foreground seeds and the pixels outside the red bounding boxes and inside the yellow bounding boxes are used as the background seeds. The yellow bounding boxes are obtained by doubling the size of the green bounding boxes, and
the red bounding boxes are obtained by extending the size of the green bounding boxes in each dimension a quarter. This matting task is guided by top-down knowledge.

To address the local frequent motions in the background, we further develop a feature-matching based local motion stabilization algorithm that can reduce the foreground false positives in background subtraction. There are three major contributions are a visual-attention analysis based algorithm is developed to evaluate if an RoD shows the background in a frame; 2) a forward/backward background propagation algorithm is developed to construct complete background images; 3) a feature-matching based local motion stabilization algorithm is proposed to suppress frequent local motions in the background and reduce false positives in foreground detection.

Fig. 1. Illustration of the heuristic seeds selection scheme for moving objects matting and segmentation

Our overall framework of foreground detection is illustrated in Fig. 2. The proposed method has been evaluated extensively on a large amount of data that contain objects with Scarcely motions: 18 long videos (580,041 frames in total) from DARPA Mind’s Eye project Y2 dataset containing significant illumination changes, cluttered background, and motions in the scene, and 6 videos (18,650 frames in total) from the category of “Intermittent Object Motion” in the Change Detection dataset. Experiment results have

Demonstrated that the proposed method performs several state-of the-art motion detection methods, especially with the Scarcely moving foreground objects.

2. RELATED WORK

Background subtraction may be the simplest approach for foreground detection. The basic idea is to obtain a background image that does not contain any object of interest. A video will be compared with the background image for the foreground object detection. The most critical and challenging task in background subtraction is background modeling, i.e., obtaining a clean background image is the important task, which generally includes background initialization and updating the difference between the foreground and background objects. Assuming that foreground objects have different color or intensity distribution from that of the background, the majority of background modeling approaches learn a background distribution at each pixel location, which is then used to classify each pixel in a video frame as background or foreground. The background distribution at each pixel can be modelled parametrically, such as a Gaussian Mixture Model, or non-parametrically, such as kernel density estimation. More recently, statistical background modeling has been extended to estimate the background distribution in a spatial or spatiotemporal neighborhood. Sheikh and Shah challenged the idea of modeling background distribution at each pixel, and employed the correlation between spatially proximal pixels. Narayana et proposed a kernel estimate at each pixel using data samples extracted from its spatial neighborhood in previous frames. Moshe et directly modeled the statistics from 3D spatiotemporal video patches to capture both the static and dynamic information of the scene. Hofmann et proposed a pixel-based adaptive segmentator, which used a history of N background values to construct the background model and a random update rule. Hernandez-Lopez et proposed to regularize the likelihood of each pixel belonging to background or foreground based on a Quadratic Markov Measure Field model. However, this method assumes that the first frame of the video does not contain the foreground and thus, cannot handle the case that the foreground objects are present at the beginning of the video. Shimada proposed a bidirectional background modeling approach based on case-base reasoning, where a background model was retrieved from an online constructed background database. In the authors proposed to fuse the motion detection based on spatiotemporal tensor formulation and the foreground and background modeling scheme based on split Gaussian models. Wang and Dudek modeled background for each pixel by using a number of background values, followed by a classification process based on matching the background model templates with the current pixel values. Besides modeling the statistics, foreground or background separation can be performed through low-rank subspace separation. Cui proposed a model using both low rank and group sparsity constraints, which represented two observations, i.e. “background motion caused by orthographic cameras lies in a low rank subspace” and “pixels belonging to one trajectory tend to group together”, respectively. He introduced an online background modeling algorithm, named Grassmannian Robust Adaptive Subspace Tracking Algorithm, for
low-rank subspace separation of background and foreground from randomly subsampled data. Lin proposed to pursue low-rank subspace in spatiotemporal domain. However, the low-rank constraint tends to treat the objects with scarcely motions as the background. In addition to color or intensity information, local texture information has also been employed in background modeling. Liao employed the local texture information by using a scale-invariant local ternary pattern, which is modified from Local Binary Patterns. Han integrated the histogram of SILTP features and color information in a blockwise background model. Liu extended the SILTP to spherical-center-symmetric SILTP (SCS-SILTP) by integrating spatiotemporal statistics for background modeling with a pan-tilt-zoom (PTZ) camera. Kim proposed to use SIFT features to generate adaptive multi-homography matrices, which are then used to compensate for the global camera motion to detect the moving objects under the moving camera. Yao and Odobez combined the local textures represented by LBPs and color features. However, there is a common assumption in the existing background modeling algorithms that the background is more frequently visible than the foreground. As a result, they are more likely to treat an object with scarcely motions as part of the background. In this paper, we employ a visual attention analysis based mechanism to explicitly deal with the foreground objects with scarcely motions. There is another set of literatures focusing on detecting abandoned/removed objects. Propose two background models to handle abandoned objects. The long-term background model is updated slowly by using a large learning rate, while the short-term background model is updated fast. Thus the abandoned objects can be detected through comparing background subtraction results using the long-term and short-term background models. However, the long-term background model will cause the “ghosting” artifacts. Since the long-term background model is still updating, the abandoned objects will be treated as background finally. Model the background by using three Gaussian mixtures to represent the background and changes in different temporal scales, which also suffers the “ghosting” artifacts. Maddalena and Petrosino explicitly detect the stopped objects from the moving ones by counting the consecutive occurrences (i.e., detected as a foreground) of an object from a sequence of frames. However, this model cannot detect the removed objects since it employs the first frame to initialize the background. The saliency detection has recently raised a great amount of research interest and has been shown to be beneficial in many applications. We would like to emphasize that the proposed approach is totally different from these approaches. Regions with high visual saliency are identified on each frame as foreground without considering any motion cue. Spatio temporal segments with high visual saliency are identified from a video as foreground. While this method considers motion cue in evaluating the visual saliency, it will fail to detect an infrequently-moving object once it stays static and generates no motions. We do not directly use visual saliency to separate foreground and background. Instead, we identify RoDs and compare the saliency values of anRoD in different frames to help construct the complete background images. Together with a step of background propagation, our method can better detect infrequently moving objects. In addition, directly using saliency in each frame to distinguish the foreground and the background may work poorly when the background is highly textured – highly textured regions are usually considered to be salient in most visual attention models. The proposed method compares the relative saliency of a region across frames to distinguish the foreground and the background, and can better identify the highly textured background, as shown later in the experiments.

3. OVERVIEW

To accurately extract moving objects in dynamic scenes, we combine a background subtraction and an alpha matting technique via a heuristic seeds selection scheme. The flowchart of our method is shown in Fig.3 in which the complementary pixel color information and statistics of neighborhoods are first exploited to generating candidate BGS maps. Then, an automatic matting technique driven by a heuristic seeds selection scheme is utilized to extract a high-quality alpha matte and moving objects from the dynamic scenes.

4. EXISTING SYSTEM AND PROPOSED SYSTEM

The abandoned objects can be detected through comparing background subtraction results using the long-term and short-term background models. However, the long-term background model will cause the “ghosting” artefacts. Since the long-term background model is still updating, the abandoned objects will be treated as background finally.

A common assumption in the existing background modeling algorithms that the background is more frequently visible than the foreground. As a result, they are more likely to treat an object with Scarcely motions as part of the background. In this paper, we employ a visual attention analysis based mechanism to explicitly deal with the foreground objects with motions. That foreground objects can be detected through background subtraction. The proposed background-modeling method involves the estimate of between temporally nearby frames. In the following, we first introduce the operation of estimation and then introduce the proposed background modeling method.
The proposed method for background modeling and background subtraction, starting from an input long streaming video. For a long streaming video, there may be intermittent, abrupt background changes, such as those caused by sudden illumination change or camera shake. In this paper, we first divide the long video into a set of super-clips so that each super-clip does not contain abrupt background change. In this way, we can perform background modeling for each super-clip independently.

The pixel-wise background subtraction as presented above is sensitive to frequent local motions in the scene (background), such as trees and/or grass waving in the breeze. As a result, the waving trees and/or grass will be misdirected as foreground objects. To suppress the effect of local motions in background subtraction, we propose a local motion stabilization method based on feature-matching.

5. BACKGROUND SUBTRACTION AND LOCAL MOTION STABILIZATION

A. Background Subtraction Once the background image is constructed for each video clip Ci, background subtraction can be conducted by subtracting every frame in the video clip from the background image. We use the same algorithm for calculating the RoDs (see Section III-A) for background subtraction. The only difference is that we input a frame and the background image, instead of two frames, to calculate the RoDs, which are taken as the detected foreground objects.

B. Local Motion Stabilization based on Feature Matching The pixelwise background subtraction as presented above is sensitive to frequent local motions in the scene (background), such as trees and/or grass waving in the breeze. As a result, the waving trees and/or grass will be misdetected as foreground objects. To suppress the effect of local motions in background subtraction, we propose a local motion stabilization method based on feature-matching. In this case, the detected RoDs from background subtraction (i.e., subtracting a frame f to the background image b) may come from the foreground objects or the background local motions. We examine each RoD R in f and identify it to be part of the foreground or the background. Our basic idea is that, if R is part of the background in f, then f(R), the region R in f, and b(R), the region R in b, should share a lot of appearance features, such as SIFT features, although there is background local motion between f and b. The SIFT features are invariant to image scale and rotation and robust to changes in illumination, noise, and minor changes in viewpoint. Since SIFT features are invariant to image scale and rotation, robust to changes in illumination and noise, and have the highest matching accuracy compared to other local features, we detect and match the SIFT features between f(R) and b(R) and define the background likelihood of R in f as

$$
\Omega(f(R)) = \frac{N_{\text{matched}}}{\max(N_f, N_b)},
$$

Where N matched denotes the number of matched SIFT pairs between f(R) and b(R); Nf and Nb denote the number of detected SIFT feature points on f(R) and b(R), respectively. In our experiments, if \( \Omega(f(R)) \) is larger than a predefined threshold \( \tau \), R is considered to be part of the background in f and we remove it from the foreground detection result. Recent work on background subtraction has shown developments on two major fronts. In one, there has been increasing sophistication of probabilistic models, from mixtures of Gaussians at each pixel, to kernel density estimates at each pixel, and more recently to joint domain-range density estimates that incorporate spatial information. Another line of work has shown the benefits of increasingly complex feature representations, including the use of texture information, local binary patterns, and recently scale-invariant local ternary patterns. In this work, we use joint domain-range based estimates for background and foreground scores and show that dynamically choosing kernel variances at each individual pixel can significantly improve results. We give a heuristic method for selectively applying the adaptive kernel calculations which are nearly as accurate as the full procedure but runs much faster. We combine these modeling improvements with recently developed complex features and show significant improvements on a standard backgrounding benchmark.

![Fig 4: Example of Background subtraction](image)
6. ADVANTAGES AND DISADVANTAGES

Advantages

- A different “threshold” is selected for each pixel
- These pixel wise “thresholds” are adapting by time
- Objects are allowed become a part of the background without destroying the existing background model
- Provides fast recovery

Disadvantages

- Cannot deal with sudden, drastic lighting change
- Initializing the median filtering is important
- There are relatively many parameters, and they should be selected intelligently

7. CONCLUSION

In this work, we proposed a novel method to detect moving foreground objects, which is especially capable of detecting objects with Scarcelymotions. Specifically, we improve the background subtraction method by integrating a visual attention mechanism to distinguish the foreground and background. The identified background regions can be propagated back-and-forth along the whole super-clip. Furthermore, we also proposed a SIFT-matching based local motion stabilization algorithm to deal with the frequent local motions in the scene. Extensive experimental validations on two challenging datasets have demonstrated that the proposed method outperforms the state-of-the-art background subtraction methods in comparison. As shown in the experimental results, the performance improvement is more impressive for detecting objects with Scarcelymotions. In this work, a simple video decomposition strategy has been used to divide the long video into super-clips and works well under the assumption that the camera keeps static in the most of the time. In order to handle complicated camera motions, in the future, we plan to try more sophisticated video decomposition methods, such as, to generate super-clips. From the scene, the offline mode will be triggered and the background model will be updated using the bi-direction background propagation. We also plan to use the proposed model in surveillance applications, especially when the events of interest involve infrequent moving objects, e.g., abandoned object detection and fall detection.

8. REFERENCES