

METHODS AND ALGORITHMS FOR DETECTING OBJECTS IN VIDEO FILES

Cuong Nguyen The^{1,2*}, *Dmitry Shashev*¹

¹National Research Tomsk State University, 634050, Tomsk, Russia

²Vietnam Maritime University, 484 Lach Tray, Le Chan, Hai Phong, Viet Nam

Abstract. Video files are files that store motion pictures and sounds like in real life. In today's world, the need for automated processing of information in video files is increasing. Automated processing of information has a wide range of application including office/home surveillance cameras, traffic control, sports applications, remote object detection, and others. In particular, detection and tracking of object movement in video file plays an important role. This article describes the methods of detecting objects in video files. Today, this problem in the field of computer vision is being studied worldwide.

1 Introduction about video files

Video files are files that store motion pictures and sounds of real-life. Video is a form of animation that gathers many images appearing continuously in sequence. Usually one second of video has 24 (or maybe more) still images and 24 or more photos, the human eyes (with retinal image retention mechanism of 1/24 sec.) will feel the motion image as continuous, smooth and beautiful.

2 The application for searching objects in video files

Detecting and tracking moving objects in video files has wide range of applications in real life. If a person just watches a video file, human eyes will not be able to give specific details about the time and the process of changing location of objects, especially those with a fast change process (such as detection of human face, smile detection when taking pictures, cars on the road, the path of a moving tennis ball and others), objects with complicated orbits (such as soccer players running on the pitch), or objects with a process of change slowly compared to the background (objects between the sea and sky, daylight changes, and tides). In addition, the object detection in the video file at the moment will help a lot in real life, for example, to determine if there was a goal or not (in football), if a tennis balls is in/out of court (in tennis), or which athlete has finished first (in speed races).

* Corresponding author: cuongntit@vimaru.edu.vn

3 Overview about searching objects

As we all know, video is a collection of discrete images that are constantly displayed to create motion effects. Thus, object detection in video files is also based on the idea of detecting objects in an image file. We need to perform the image-partitioning step to determine where the area of objects is, and where the area of the background is. Image segmentation is an important step in image processing, and this phase analyses the image into components of the same nature based on the boundaries or interconnected regions. An image is a detail, an object in panorama. An image area describes the surface properties of an image; this area is surrounded by a boundary and points with a relatively uniform grey level. Based on the physical properties of the image area, we define a number of partitioning techniques. The main partitioning methods include:

- Classification or threshold-based method.
- Structure-based method.
- Boundary-based method [1].

Object detection in video files should be based on the object's history of motion over discrete images over time. In function $I(x, y, t)$, 'I' is an image, 'x', 'y' are the coordinates of the object's location in two-dimensional space, and 't' is the time of the object at that coordinate. An approach based on the history of motion pictures is a view based on the method of pattern detection over time. This is a simple but effective way of displaying motions; it is used in a number of studies related to motion recognition, motion analysis, and other related applications.

An overview of real-time video detection, classification, and mark up of motion shows that this system can implement the following operations: (1) distinguish between transient or stationary objects and static background objects in the scene; (2) detect and distinguish objects, which move and disappear; (3) classify objects that have been discovered in different groups such as people, groups, vehicles, and others; and (4) mark the objects and generate information about the paths in the video image.

Video processing-based applications have common characteristics. Commonly used technologies for detecting motion objects are background subtraction, static methods, time difference, optical flow, and others.

3.1 Background subtraction method

Background subtraction is a part of motion-moving technology in static scenes. It tries to detect motion regions by subtracting pixels from the current image to a background image that was created by the average background image for a period of one initialization cycle. There are several approaches, which can be presented as follows.

Let $I_n(x, y)$ be a representation of the greyscale intensity value at a pixel location (x, y) with a value of $[0 \div 255]$ in the n^{th} case of the video sequence I . Let $B_n(x)$ be a corresponding background magnitude value for a pixel at time (x, y) estimated from the video image I_0 to I_{n-1} . A pixel in position (x, y) in the current image belongs to the dominant component (highlighted) if

$$|I_i(x,y) - B_i(x,y)| > T_n(x, y) \quad (1)$$

is satisfied, where $T_n(x, y)$ is a predefined threshold.

The B_t background image is updated by using the Infinite Impulse Response (IIR) filter as follows:

$$B_{t+1} = \alpha I_t + (1 - \alpha) B_t \tag{2}$$

This background is updated for all types of pixels. In the background subtraction methods, the source background image is updated only for pixels in the background.

3.2 Static methods

The W4 method uses a static background pattern, where each pixel is represented again with the smallest value (M) and its maximum value (N) in terms of intensity and maximum intensity difference (D) between any consecutive frames observed during the training initialization cycle, in which the scene does not contain moving objects. One point in the current image is classified as a prominent feature if it satisfies:

$$|M(x, y) - I_t(x, y)| > D(x, y) \quad \text{or} \quad |N(x, y) - I_t(x, y)| > D(x, y) \tag{3}$$

Every pixel is modelled by a blend of the Gaussian methods and is updated online by the input image data. Gaussian distributions will evaluate whether a point belongs to the dominant point processing or to the underlying processing.

3.3 The difference over time

Difference over time tries to detect motion regions using the difference between pixels in consecutive frames (two or three) in a sequence of videos. This method is highly adaptable to the scene; however, it often causes errors in detecting moving objects. The single-color area of the left-hand side causes the time difference algorithm to extract all pixels of the human motion.

A theory of image identification and representation was proposed by Bobick and Davis [2]. In particular, this theory analyses motion-based identification along with description of the spatial element and describes how the object moves. It represents formation of the Motion Energy Image (MEI) or the Binary Motion Region (BMR), and we can see that it represents a position of the transfer available in a sequence of images. MEI represents the shape of movement and the spatial division of motion. Then the MHI (motion history image) is made. The intensity of each pixel in the MHI represents an interval of the number of times that can be encoded by each frame, and this MHI method will hook up the time proportion of human gestures. If to put MEI and MHI together, we will see that MHI and MEI can be viewed as two component versions of the sample over time. In a vector image, each component of each pixel is a number of functions that move at that pixel location. Specific specimens that are grafted will have competition with the storage patterns of the known motion view. Unification of MEI and MHI models is constructed using the MHI method with the historical motion function $H_T(x, y, t)$ calculated from the difference in image and optical flow.

At the simplest level, each object is identified and represented by a boundary box as shown in Figure 1. As the object moves, the boundary box moves around. The difference in image is done by the entire image in the area of the bounding box. Combination with existing patterns will give the type of object to be identified.



Fig. 1. Illustration of the object in the bounding box [3].

At a more granular level, each object is identified by an image area that is distinct from the background (the boundary will have a shape of the object). Movement of the object is monitored by subtracting the difference in optical current. Illustration is made by converting to binary image as shown in Figure 2.

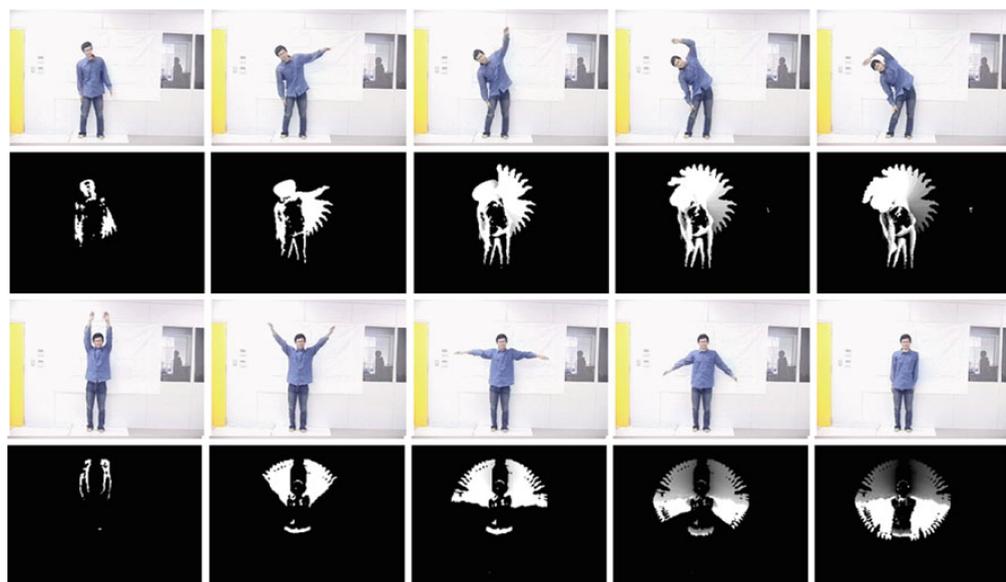


Fig. 2 Development of the MHI images for two different actions. The MHI images produced are shown under the actions sequentially [4].

4 Issues of detecting and tracking an object in video files

Detecting and tracking objects in video files encounters seven following issues: (1) Brightness change. The brightness of an image can change in space and time; some parts of an object may be brighter and brightness of objects in the current image may become darker on the next day. (2) Ability to expand or shrink an object. An object can move close to or move out of the camera's frame. (3) Rotations of an object. An object can rotate relative to the camera in all three axes of the three-dimensional space. (4) Screening of objects. An object can be overlapped by other objects. (5) Video has distorted objects or noise. Noise in the image can occur because of natural and artificially created reasons. The natural work can be attributed to the incorrect operation of the photo (video) camera and all the natural conditions such as snowfall or rain, under which the image (video) is obtained. Artificially created reasons include all human actions aimed to change the image quality such as changing colour space, compression of the image with loss of quality, blurring of the image, and others. (6) Change of the geometric shape of an object. The observed object can be deformed because it can turn to

the camera with a part, which differs from the form observed before. An object can consist of small inter-connected objects that can move in space along different trajectories. For example, the desired object is a person whose hands can move in different directions. (7) Change of the colour histogram of an object. An object can have different colour histograms on different sides of its body [5].

Scalability and rotations refer to affine transformations [6, 7]. According to [8], projective transformations include affine transformations and take into account the distortions associated with the slope of the camera.

5 Methods and algorithms for searching objects

The main methods and algorithms for searching objects include the following:

- Dense optical flow. These algorithms help to estimate the motion vector of every pixel in a video frame.
- Sparse optical flow. These algorithms, like the Kanade-Lucas-Tomashi (KLT) feature tracker, track the location of a few feature points in an image.
- Kalman Filtering. This is a very popular signal-processing algorithm used to predict location of a moving object based on prior motion information. One of the early applications of this algorithm was missile guidance! Moreover, the on-board computer that guided the descent of the Apollo 11 lunar module to the moon had a Kalman filter.
- Meanshift and Camshift. These are algorithms for locating the maxima of a density function. They are also used for tracking.
- Single object trackers. In this class of trackers, the first frame is marked using a rectangle to indicate location of the object we want to track. The object is then tracked in subsequent frames using the tracking algorithm. In most real life applications, these trackers are used in conjunction with an object detector.
- Multiple object track finding algorithms. In cases, when we have a fast object detector, it makes sense to detect multiple objects in each frame and then run a track finding algorithm that identifies, which rectangle in one frame corresponds to a rectangle in the next frame [9]. Such algorithm is shown in Figure 3.



Fig 3. Tracing multiple objects [9].

6 Conclusion

Currently, anti-jamming algorithms (and methods) are not suitable for actual finding objects in the video files. The method proposed is based on finding the main characteristics of the object.

In comparison with other methods, the algorithms proposed remain unchanged for projective transformations, are resistant to noise and brightness change, and are characterized by less computational costs. The searching method requires immutability to object transformations and minimally achievable computational complexity in a real-time mode [10–12].

The further research is oriented to find an algorithm for detecting fast-moving objects in video files in order to use it to for detecting a potentially dangerous person. These algorithms can be used for detecting dangerous people who carry suspicious weapons or other things that must to be checked. Such situations include, for example, a person sitting but suddenly standing up and running at high speed, or an object, which suddenly moves to another object at high speed. The authors will study the main characteristic features and colour charts in the process of determining an object (as a person) in order to detect object deformation in video files (because rotation of objects is divided into different parts: head - body - leg).

Acknowledgements

The research presented in this paper was supported by the Russian Foundation for Basic Research (grant No 16-29-04388/17). The authors are grateful to Tatiana B. Rummyantseva from Tomsk State University for her assistance in preparing the article.

References

- [1] S. M. Bryan, *Lecture 15: Segmentation (Edge Based, Hough Transform)* (Brigham Young University, 2000)
- [2] A. Bobick A, J. Davis, An appearance-based representation of action, International Conference on Pattern Recognition (1996)
- [3] Motion detection in OpenCV, URL: https://www.youtube.com/watch?v=z1Cvn3_4yGo (2013)
- [4] Md. Atiqur Rahman Ahad, J. K. Tan, H. Kim, S. Ishikawa, Machine Vision and Applications **23**, 255 (2012) doi :10.1007/s00138-010-0298-4
- [5] A. V. Pastushkov, Izvestiya Vysshikh Uchebnykh Zavedenii. Physics **56**, 334 (2013)
- [6] A. A. Potapov, *Newest methods of image processing* (FIZMATLIT, Moscow, 2008)
- [7] D. Shashev, S. Shidlovskiy, V. Syriamkin, A. Yurchenko. International scientific conference on radiation-thermal effects and processes in inorganic materials **81**, (2015)
- [8] I.S. Gruzman, *Digital imaging in information systems* (Publishing house of the NSTU, Novosibirsk, 2002)
- [9] Satya Mallick, Object Tracking using OpenCV (C++/Python), URL: <https://www.learnopencv.com/object-tracking-using-opencv-cpp-python/>
- [10] Diego Cavalca, Object Detection with Tensorflow API, URL: https://www.youtube.com/watch?v=_zZe27JYi8Y
- [11] A. V. Pastushkov, Dissertation for the degree of candidate of technical sciences (2017)
- [12] S.V. Shidlovskii, J. Comp. Sys. Sci. Int **45**, 282 (2006)
- [13] D.V. Shashev, S.V. Shidlovskiy, Opt. Inst. D. Proc. **51**, 227 (2015)