ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

Salient Motion Detection in a Video Signals

Mahadev R Ingawale[1], Dr.Mrs.S.B.Patil[2]

Student, M.E (E&TC), Dr.J.J.M.C.O.E. Jaysingpur, Maharashtra, India [1]

Professor, Department of Electronics and Telecommunication, Dr.J.J.M.C.O.E. Jaysingpur, Maharashtra, India [2]

Abstract: In a video signals the detection of salient motion involves determining which motion is attended or presented by the human visual system in presence of complex background motion that are constantly changing. When the video sequence is represented as a linear dynamical system, then the salient motion detection is achieved from the output pixel. Salient motion is detected by comparing predictability to the more complex unpredictable background motion. The pixel saliency map is supported by two region based saliency map. It consists of different spatiotemporal patches in a video with the salient region in a global as well as local scene.

Keywords: Centre-surround saliency, pixel saliency map, spatiotemporal saliency, video sequence.

I. INTRODUCTION

The foreground detection is one of the major tasks whose aim is to detect changes in image sequences. Numbers of applications do not need to know everything about the movements in a video sequence, but it only require the information about changes in the scene. Background modelling and subtraction is a core component in motion analysis. The components of this model are used in an autoregressive form to predict the frame to be observed. It is always desirable to achieve very high accuracy in the detection of moving objects with the lowest possible false detection rate. The performance of background subtraction depends mainly on the background modelling technique. The foreground salient motion detection problem has approached by representing the video sequence as autoregressive (AR) models [1]. The differences in the structural appearance and the motion characteristics of the predicted and observed signal are used to detect the foreground from the background.

An autoregressive (AR) model is used to predict the dynamic background and salient motion detection that is performed by comparing the prediction with actual observation. Similarly the dynamic texture (DT) model [2] is used for the most recent frames in the video sequence to construct a predictive model that captures the important variation in the video signal. By fitting a dynamic texture model to the spatiotemporal patch having maximum pixel saliency, the dynamics of salient region is modelled. This model is further used for comparison with the other patch models in the spatiotemporal saliency map to compute the region-based saliency maps.

The continuously changing background in [3] is predicted using an autoregressive model for dynamic texture and the foreground objects are detected as outliers to this model. When there are no foreground objects in the scene then the autoregressive model evaluates the initial state vector. The foreground image is formed by minimizing the error of observed output and predicted output. So in this method, the assumption of availability of "background only" scenes is a drawback of these methods. In this method, the sequence of frames in a video signal is represented as a multi-input multi-output (MIMO) autoregressive state space model.

The salient region is modelled by analysing the spatiotemporal patch around the output pixels set within a linear dynamical system framework. A distance measure is used to compare the AR model around the most relevant spatiotemporal patch with other models. The salient region is modelled by fitting a dynamic texture model to the spatiotemporal patch and this patch having maximum pixel saliency. Finally saliency map is computed from the region-based saliency maps and the pixel saliency maps.

PROPOSED WORK

The first step of a salient motion detection algorithm is the foreground-background segmentation and subtraction step. To segment out the foreground from background there are lot of techniques available. The pixels intensity values are added with the colour illumination into the algorithm, so it is easy to improve the image subtraction techniques. The background subtraction model in [4] models each pixel as a mixture of Gaussians. Mixture of several Gaussians is preferred to model the background. The model is updated regularly and evaluated to determine which of the Gaussians are sampled from the background process. When changes in the background are too fast then the variance of the Gaussians becomes too large and non parametric approaches are more suited. The concept of observabilty from the output pixels is provides a good clue to the salient motion detection in natural videos.

II. SPATIOTEMPORAL SALIENCY

Spatiotemporal saliency and the related task of background subtraction are commonly used as a pre-processing step for object and event detection activity and gesture recognition, tracking, surveillance, and video retrieval. The most advanced techniques requires for spatiotemporal saliencies are,

1. Explicit, or approximate, compensation of camera motion
2. Foreground objects that move in a consistent direction or have faster variations in appearance than the background.
3. Explicit background models.
4. Static cameras

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

2.1 Discriminant Centre-Surround Saliency:

A novel method for detecting salient regions in both images and videos based on a discriminant centre-surround hypothesis [5], the salient region stands out from its surroundings. Background subtraction is considered to the salient motion detection, for which the discriminant centre-surround saliency observations proposes a solution. Under these observations, saliency is the result of optimal discrimination between centres and surround event at each location of the visual field. A set of visual features is collected from centre and surround windows and the locations between the features can be performed with the smallest expected error probability and declared as most salient. Background subtraction then reduces to ignoring the locations declared as non salient. A simple modification of the features and probabilistic models, discriminant saliency is applicable to various problems. A dynamic texture model is adopted due to their versatility in moving patterns and ability to reproduce natural scenes.

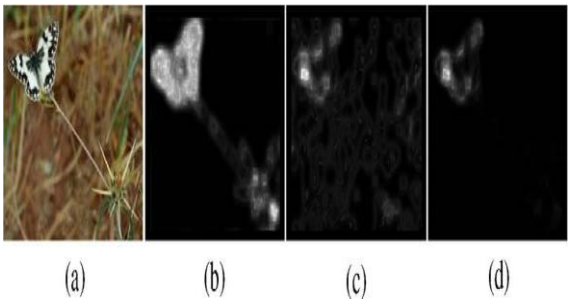

Fig.1 (a) Original image. (b) Edge saliency. (c) Colour saliency.
(d) Spatial saliency

It will have a maximum value on locations where both edge and colour saliencies agree and also vary proportionately in the regions that are in favour of edge and colour saliency. The example of our spatial saliency is shown in Fig. 1. We can see that non-salient regions are effectively suppressed [5].

The ability of human visual system to detect visual saliency is extraordinarily fast and reliable. Still a computational modelling of this behaviour is remains a challenge. The spectral residual of an image in spectral domain is extracted by analysing the log-spectrum of an input image and propose a fast method to construct the corresponding saliency map in spatial domain. The output indicates fast and robust saliency detection.

2.2 Dynamic Textured Model:

The dynamic texture is modelled by an Autoregressive Moving Average Model (ARMA). A robust Kalman filter algorithm estimate foreground objects region, also it estimate the intrinsic representation of the dynamic texture. The challenge of foreground segmentation given dynamic textured scenes is that the background is continuously changing [2]. When the background is a dynamic texture, then the work describes a framework for detection and segmentation of foreground objects in video. Dynamic textures exhibit repetitive patterns in space-time. Example dynamic textures are shown in Fig.2. Foreground objects are those that appear in front of a dynamic texture with distinctive statistics in space-time. Real world examples include ships on the sea, people riding an escalator, etc. It is assumed that foreground objects are distinctive in at least their spatial or temporal statistics.

Fig.2. Examples of dynamic textures

The foreground objects and dynamic textured background have difference in their motion patterns but they are similar in colour distributions. The non-stationary nature and clutter-like appearance of dynamic textures cause many traditional background subtraction methods to fail, since these methods assume a static or slowly changing background.

2.3 Image Descriptors

Visual descriptors or image descriptors are descriptions of the visual features of the contents in images, videos, algorithms, or applications that produce such descriptions. Visual descriptor describes the characteristics of images or videos such as the shape, colour, texture or the motion. Today's new communication technologies and use of internet in our society increases the amount of audio visual information in digital format. Therefore, it has been necessary to design some systems that allow us to describe the content of several types of multimedia information in order to search and classify them. The audio visual descriptors have a good knowledge of events and the objects found in an audio, video or in image. They allow efficient and quick searches of the audio-visual content.

Types of visual descriptors: Descriptors find out the connection between pixels contained in a digital image and observed image or a group of images. Visual descriptors are described in two groups;

1. General information descriptors: General information descriptors consist of a set of descriptors that covers different basic and elementary features such as motion, colour, shape, location and texture. This description is automatically generated by using signal processing.
2. Specific domain information descriptors: Specific domain information descriptor gives the information about events and objects in the scene and these events are not easily extractable, when the extraction is to be automatically done. Face recognition is an example of an application that tries to automatically obtain the information.

Descriptors applications:

a) Multimedia documents search engines and classifiers.
b) Personalized electronic news service.
c) Possibility of an automatic connection to a TV channel broadcasting a soccer match, for example, whenever a player approaches the goal area.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

III. OPTICAL FLOW

Object tracking generally achieved through monitoring optical flow [6]. Optical flow vectors have also been utilized to detect salient motion. Optical flow or optic flow is an apparent motion pattern of objects, edges and surfaces in a visual scene and it is caused by the relative motion between an observer (an eye or a camera) and the scene.

The optical flow concept describes the visual stimulus provided to animals moving through the world. This feature has since been elected by robot cists, which use optical flow techniques for image processing and control of navigation.

It includes time-to-contact information, motion detection, focus of expansion calculations, luminance, object segmentation; stereo disparity measurement and motion compensated encoding. Optical flow is useful beyond spatiotemporal saliency framework. It Change of image position in between two frames. Optical flow is the apparent motion of brightness patterns in the image.

Estimation of the optical flow: Sequences of ordered images allow the estimation of motion as either instantaneous image velocities or discrete image displacements. The optical flow methods are based on local Taylor series approximations of the image signal, so these methods are called differential.

The optical flow methods use partial derivatives with respect to temporal coordinates and spatial coordinates. The optical flow method calculates the motion between two image frames taken at a time and at every pixel position.

Consider a five-frame clip of a ball moving from the bottom left of a field of vision, to the top right as shown in Fig. 3. Motion estimation techniques can determine that on a two dimensional plane. The ball is moving up and to the right and vectors describing this motion can be extracted from the sequence of frames. For the video compression e.g., MPEG, the sequence is described.

In the machine vision field either ball is moving to the right or the observer is moving to the left is not knowable. Even if a ball is not static, patterned background is present in the five frames then we can confidently state that the ball was moving to the right, because the pattern has infinite distance to the observer.

The application of optical flow includes the problem of inferring the motion of the observer and objects in the scene. Also it includes the environment and the structure of objects. Since elementary state of motion and the generation of maps of the structure of our environment are critical components of animal and human vision.

Motion estimation and video compression have developed as a major aspect of optical flow research. The optical flow field is similar to a dense motion field derived from motion estimation techniques.

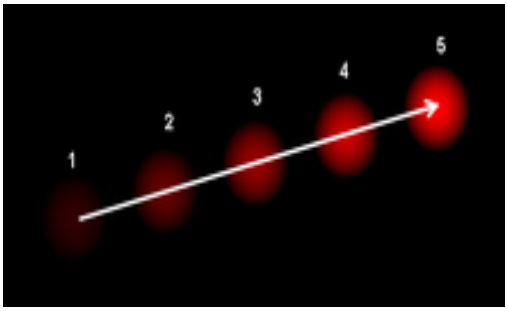

Fig. 3 The optical flow vector of a moving object in a video sequence

Optical flow is the study of the determination of the optical flow field itself as well as it is used in estimating the three-dimensional (3D) structure of the scene and nature. Also optical flow field estimate the observer relative to the scene and the three-dimensional (3D) motion of objects. Optical flow is used by robotics researchers in many applications such as image dominant plane extraction and movement detection, robot navigation, visual odometry, object detection and tracking. Optical flow information has been recognized as being useful for controlling micro air vehicles

IV. PIXEL SALIENCY MAP

Saliency map represent the saliency at every location in the visual field by a scalar quantity when all existing saliency-based approaches suffer the integrity problems for interesting object extraction. Pixel saliency map is proposed to extract objects using two saliency maps. The salient motion is considered as a more predictable for a human observer compared to the relatively unpredictable backgrounds. Pixels belonging to complex unpredictable backgrounds will have less observability to state variables compared to pixels belonging to regions of simple predictable motions. This is due to the fact that we are modelling the video sequence over a period of time and the motion components belonging to complex unpredictable regions are hard to estimate from its organic characterization.

4.1 Pixel saliency map using observability from output pixels:

The basic definition of observability and a simple rank-based method is evaluated to view whether a system is observable or not [6]. These methods help to make a binary decision on the observability of a system. They do not provide any insight into the observability from different outputs to the Eigen values of state transition matrix. Recall that observability is evaluated with respect to Eigen values of state transition matrix, since it is invariant under various equivalent system representations. The observability measure evaluated for pixels in the output vector provides a good clue for salient motion [2].

4.2 Sustained Observability

In this method some other pixels in the dynamic background can also possess some observability as can be seen in the Cyclists video sequence whose frame is shown in Fig.4 (a), Fig.4 (b) and Fig.4 (c).

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

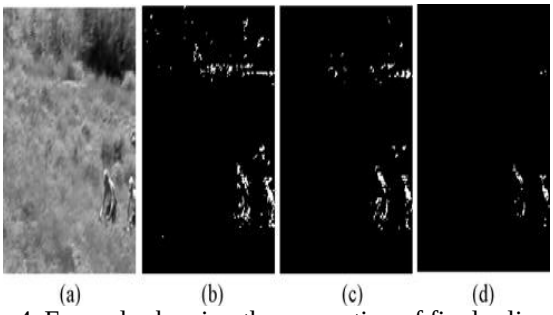

Fig.4. Example showing the generation of final saliency map with sustained observability. (a) Frame of Cyclists video. (b) and (c) Saliency maps (d) Final saliency map.

Fig. 4 shows the saliency maps for adjacent frame buffers, where in some part the dynamic background possess observability. Pixels having "sustained observability" are those that maintain high observability over successive saliency maps computed for adjacent frame buffers. The DT models evaluated for the adjacent buffers resulting in saliency maps b and c. The final saliency map is computed by multiplying the two maps and normalizing it to [0, 1]. Thus we give more saliency to those pixels that hold the observability across the adjacent models. The final saliency map correctly identifies points that show high observability consistently as shown in Fig.4 (d).

4.3 Centre-Surround Saliency

Centre-surround is implemented in the model as the difference between fine (centre) and coarse (surround) scales. The saliency map combines information from each of the feature maps into a global measure where points corresponding to one location in a feature map project to single units in the saliency map. Saliency at a given location l is determined by the degree of difference between surrounding region and its location as shown in fig. 5.

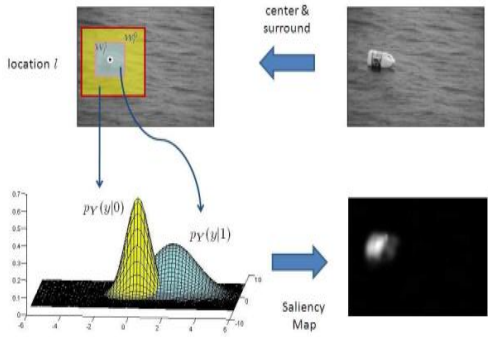

Fig.5. Centre-surround saliency

4.4 Region-based saliency map using salient patch dynamics:

The pixel saliency map computed in above section evaluates the importance of each pixel in inferring the holistic dynamics of the scene. Large homogeneous salient objects that occupy a majority of the pixels in the frame can appear to be relatively static over the frame buffer. They will be marked as less salient due to their apparent lesser involvement in the holistic dynamics of the scene. The drawback of the pixel saliency map is noticed by computing region-based saliency maps. In this process the regions correspond to spatiotemporal patches of fixed size selected from the spatiotemporal volume buffer frame. The dynamics of the most salient region hold very specific information of the region dynamics and may not always generalize well to the dynamics of the salient object. So we compare the dynamics of different regions to the dynamics of the whole of spatiotemporal volume. The dynamics of different regions are compared with the dynamics of the most salient region. This can be considered as a combination of local and global approaches in deciding region saliency.

4.5 Region-based saliency map using local dynamics:

In the region-based saliency map the first step is to understand the nature of the dynamics of the most salient region in the frame. The pixel saliency map having the maximum sustained observability and contains the most salient pixel. Spatiotemporal patch of fixed size is selected with largest average saliency around the most salient pixel. This spatiotemporal patch is deemed to fall on the salient object [2]. Also it can be modelled as a DT that is a fair approximation of the dynamics of the salient object.

4.6 Region-based saliency map using global dynamics:

The dynamics of the most salient region in nature is local and it may not always agree with the dynamics of the entire salient object. It is assumed that the salient object should have dynamics similar to the salient region but need not match its dynamics exactly. The basic idea of this assumption is the existence of minor deviations in the model when generalizing the dynamics of the object. This problem is addressed by a global approach to comparing the dynamics in the spatiotemporal buffer.

It is consider that the spatiotemporal patches of the local approach and its dynamics compare with the model of the entire frame buffer. The poles of the frame buffer model are prioritized according to its affinity toward the poles of the salient patch. So it gives more importance to the motion occurring around the most salient region. The final saliency map is computed as the product of three maps local region-based saliency map, global region-based saliency map and pixel saliency map.

V. APPLICATIONS

There are numerous technical applications in which the saliency map is typically used to identify the most important information in visual input streams and it is used to improve the performance in generating or transmitting visual data. Even an "inverse" saliency map has been used to direct attention to other regions. Saliency maps have also been integrated in a VLSI hardware model of visual selective attention.

The detection of salient motion in videos finds applications in object tracking and surveillance, video retargeting, and activity recognition. Tracking and surveillance applications are aided by the information; video retargeting applications can be used to resize video frames with minimum distortion in salient motion.

VI. CONCLUSION

The key idea is to relate the notion of predictable motion in a video and the concept of observability from the output of a linear dynamical system. We have modelled videos as a linear dynamical system to identify salient motion under complex unpredictable backgrounds. The final saliency map is computed from the pixel saliency maps and the region-based saliency maps. The method able to adapt with global and local illumination changes, weather changes, changes of the natural scene etc. The pixel has moved in a consistent direction. Objects are moving in a straight line rapidly take on salience magnitudes that are significantly larger than that of vegetation. The core contribution of our approach is the integration of a powerful set of filter operators within a linear prediction model towards the detection of events using measures that are adaptive to the complexity of the scene. This technique is able to detect such events with a minimal false alarm rate. Detection performance is a function of the complexity of the observed scene.

REFERENCES

[1] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in Proc. Int. Conf. Comput. Vision, 2003, pp. 1305–1312.

[2] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, "Dynamic textures," Int. J. Comput. Vision, vol. 51, no. 2, pp. 91–109, Feb. 2003.

[3] J. Zhong and S. Sclaroff, "Segmenting foreground objects from a dynamic textured background via a robust Kalman filter," in Proc. Int. Conf. Comput. Vision, 2003, pp. 44–50.

[4] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in Proc. IEEE Conf. Comput. Vision Patt. Recog. Jun. 1999, pp. 2246–2252.

[5] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," IEEE Trans. Patt. Anal. Mach. Intell., vol. 32, no. 1, pp. 171–177, Jan. 2010.

[6] L. E. Wixson, "Detecting salient motion by accumulating directionally consistent flow," IEEE Trans. Patt. Analy. Mach. Intell., vol. 22, no. 8, pp. 774–780, Aug. 2000.

BIOGRAPHY

Mahadev R Ingawale received B.E degree in Electronics and Telecommunication Engineering from Dr.J.J.M. C.O.E. Jaysingpur, Shivaji University, Kolhapur, Maharashtra, India in 2011 and currently pursuing Master degree in Electronics and Telecommunication Engineering.