International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 4 491 - 501

Spatiotemporal Saliency Detection: State Of Art

Sultana kadri
Research Scholar
CT Group of Institutions
Jalandhar, Punjab, India
rajankadri786@gmail.com

Pooja
Assistant Professor
CT Group of Institutions
Jalandhar, Punjab, India
poojachoudhary80@gmail.com

Manju Bala
Professor
CT Group of Institutions
Jalandhar, Punjab, India
manju.ctgroup@gmail.com

***Abstract*—**Saliency detection has become a very prominent subject for research in recent time. Many techniques has been defined for the saliency detection.In this paper number of techniques has been explained that include the saliency detection from the year 2000 to 2015, almost every technique has been included.all the methods are explained briefly including their advantages and disadvantages. Comparison between various techniques has been done. With the help of table which includes authors name,paper name,year,techniques,algorithms and challenges. A comparison between levels of acceptance rates and accuracy levels are made.

Keywords- *Spatio temporal saliency detection,temporal detection,Saliency detection etc*

*****

I. INTRODUCTION

Saliency of an image refers to spatial saliency. Itti et al[1].proposed a biological plausible model, which combines intensity, color and orientation information in order to generate a saliency map.the saliency of videos is called spatiotemporal because it requires the temporal information that is added to spatial information.the image has the values that have the local variations in both of time and space.Saliency plays an essential role in visual attention detection.the eye system can see the large amount of eye information that is effectively in a very short interval of time..Recently, there has been an great increasing interest in extending spatial saliency to spatiotemporal saliency for video sequences also. As because the motion information is very essential in dynamic scenes, motion features have been given focus in many spatiotemporal saliency methods.Itti and Pierre[1] proposed a model that computes/calculate low-level surprise at every location in video streams, where flicker and motion feature channels are combined to form the final spatiotemporal saliency map also Gao et al. proposed a discriminant centersurround hypothesis for visual saliency. Mahadevan and Vasconcelos modeled the video patches for dynamic textures to achieve the joint described representation of the spatial and temporal elemnts of saliency. Zhang et al[3] described saliency as the selfinformation with the help of statistics derived from prior experience, neglecting the current image, to obtain the visual saliency of dynamic scenes. Large amount of methods of spatiotemporal saliency combine numerous/various saliency features together to get a saliency map.In this paper a novel system is proposed for fast saliency detection. In this system, a motion estimation method which combines classical block matching and optical flow gives motion vectors. Mostly two conditions on the motion vectors are in use to select features for the saliency map. When the conditions are satisfied/verified, significant motion contrast is there in the video squence. Two elements Intensity and motion features are applied to form a saliency map. In situation when there is low motion contrast in video,colour is added and also orientation features in current frame to the saliency map of the previous frame to create a new saliency map. The motion-decision based saliency algorithm neglect the use of features to form a saliency map for each and every frame.Visual attention (saliency) analysis is an important issue in image/video applications, such as video surveillance, video object discovery, and video retargeting. Visual attention (saliency) analysis simulates the human visual system (HVS) by automatically producing saliency maps of the images/videos and detects regions of interest (ROIs) in what we perceive.Based on the employed features, visual attention (saliency) analysis approaches can be generally classified into three categories: spatial, temporal, and spatiotemporal. The spatial visual attention (saliency) analysis approaches include several sub-categories, namely, frequency-based, Harris corner detector, Itti et al. and related approaches, contrast-based, distribution-based, and others. Primate vision provides natural solutions to many machine vision problems. If it was possible to embody them in a computational theory, then machine vision would be successful. Recently, a central part of the human visual system (HVS), namely the ability to concentrate on salient regions of the visual input, has attracted several researchers both from the field of neuroscience and computer vision.This ability of the HVS states that despite the common belief that we see everything around us, only a small fraction of the surrounding visual information is processed at any time and leads to higher-level understanding of the visual scene. One of the dominant theories in the field is saliency-based visual attention.Complete vision-application systems invoke important mechanisms in order to show the computational load of various higher-level processing steps. If the attended areas show the good input, a huge amount of search can be avoided. Two major attentional mechanisms are known to control the visual-selection process: First, bottom-up attentional selection, which is a fast and often compulsory, stimulus-driven mechanism. Second, top down attentional selection that initiates from the higher cognitive levels in the brain and influences the attentional system to bias the in favour of a particular (or a combination of) feature(s). Only information about the region that is pre-attentively extracted can be used to change the preferences of the attentional system. Applications in the field include object recognition, context-based scene recognition and object detection with pre-

defined properties in noisy and cluttered video sequences.In the field of computational video analysis, image sequences are usually processed and analysed in a frame-by-frame basis to infer the short-term objects' temporal evolution. Such methods use information over a small number (typically two) of frames. Linking of the obtained results together generates long-term dynamics. The actual long-term temporal dimension of the video data is therefore disregarded by incorporating parametric motion model assumptions or smoothing constraints. Such methods are prone to noise and can lead to high computational.As detection of spatial points has created the interest of many researchers and also the spatiotemporal counterpart has created very less interest. The very well known space-time interest point detectors is the further extension of the Harris corner detector to 3Dimensional by Laptev et al..A spatio-temporal corner is representation of a region having a spatial corner whose velocity vector is changing direction. The output points are sparse also roughly correspond to initial and final points of a replacement when applied to action recognition. Doll'ar et al. found the shortcoming of spatiotemporal corners to show actions in particular domains or areas (e.g. rodent behaviour recognition also facial expressions recognition) and propose a detector which is based on the result of Gabor filters applied to both spatially and temporally. The detector gives a large and denser set of interest points also proves to be more representative of a larger and wider domain of actions or inputs. According to Lowe, sparseness is desirable or required to an level, but very less features are not sufficient and also they are problmatic in representing actions efficiently. Oikonomopoulos et al. use a very different technique and propose a spatiotemporal extension of the salient point detector of Kadir and Brady. They corelate the entropy of space and time regions to saliency also represent a framework to recognise points of interest on the basis of their characteristic scale determined by increasing and maximizing their entropy. This detector is calculated/evaluated on a dataset of aerobic actions which gives good ans promising results. Wong and Cipolla in report a more detailed evaluation/calculation of the latter and propose new detector which is based on the global information. The new detector proposed is evaluated against the state-of-the-art in various action recognition and results/outperforms the ones proposed by Laptev et al., Dollar et al.Human activity recognition and classification systems can provide useful semantic information to solve higher level tasks, for example to summarize or index videos based on their semantic content. Robust activity classification is also important for video-based surveillance systems, which should act intelligently, such as alerting an operator of a possibly dangerous situation.

Spatial saliency map

According to a very important psycho visual backing, the spatial model contains of three sequential steps: first is visibility , second is perception and finally perceptual grouping stage. The visibility portions attempts to simulate the biologically limited sensitivity of the visual system/environments: , a conversion of the RGB component into the and Cr2 is achieved.

Second, early visual features extraction is achieved by a channel decomposition (DCP) composed by splitting the 2D spatial frequency domain and in orientation. This decomposition is applied on every perceptual components leading to or resulting 17 psycho visual channels (distributed over 4 crowns) for chromatic component. Third, contrast sensitivity functions are used to assess the natural components of visibility of image, taken into observation that we are not able to assess all details/preciseness with the same accuracy. Three anisotropic CSF are usually used to weight the components.

With the help of difficult problem of the saliency maps of fusion, the following relation is proposed:

$$S(s) = \_ST\ (s) + (1 - \_)SS(s) + \_ST\ (s)SS(s)\ (3)$$

Where SS and ST are normalized spatial and temporal saliency map respectively. A and B controls the strength of the reinforcement. Figure shows for the sequence Stefan the different Computation of the a coefficient based on the spatio-temporal activity of the sequence.

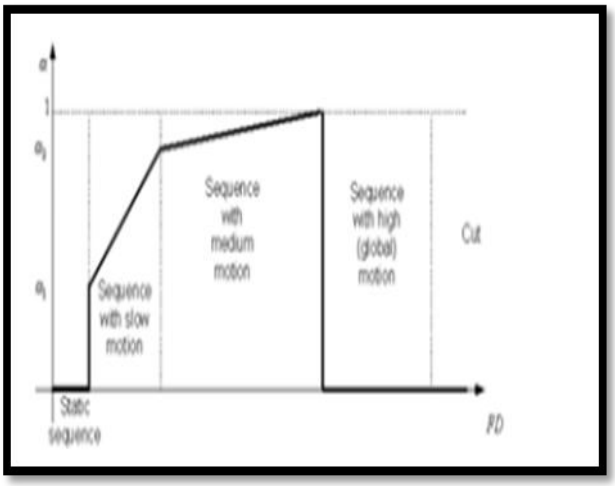

Fig-1[10]Computation of the a coefficient depending on the spatio-temporal activity of the sequence.

Temporal saliency map

The method is basically based on the fact that the visual attention is influenced by motion contrast. This type of contrast is deduced from the local and also the global motion. The new proposed technique is same in spirit to the flow/approach proposed in which the saliency of a motion region is inversely proportional to its occurring probability.

Spatial saliency detection

A biologically inspired early representation model for visual saliency detection was proposed by Koch and Ulman.This further inspired Ittietal. to propose a highly influential computational visual saliency detection method which performs local center-surrounded difference analysis of image features of color,intensity.The center-surround differencing mechanism is perform during the Difference of Gaussian approach. Saliency maps are biologically inspired approaches are blurry and often contain highly emphasized small local features in the image which might be noises as considered. Several approaches were proposed to improve Itti's model including a fuzzy growing model that estimates pixel-level dissimilarity in an image for saliency estimation.The color histogram-based computation of region-based center-surround

difference mechanism.Saliency maps produced by these methods often tend to over emphasize saliency near the edges rather than high- lighting the salient region uniformly. Most of the sebiolo- gically motivated local contrast based require multi-scale feature analysis which makes them comp- utationally in feasible or applications that need faster performance.

Temporal saliency detection

Several temporal saliency detection mode ls have been propose dover the years for detecting background regions in a visual scene,which is a complementary mechanism of saliency detection.Distribution of pixel intensities is represented by probability density function to predict the probability of background pixels in newly arrived video frames.This is used for the probabilistic models for background modeling. Since these models need exquisite tuning of several para- meters that are involved,Elgammal et al. proposed a parameter free probabilistic model for background detection.Proposed patch level texture based background modelling using pattern of histogram.This method is often affected by small scale texture dimage noises. These probabilistic model susually need a training phase in order to learn statistics of the background features.Background probability based temporal saliency detection often fails to work on videos with dynamic background or moving camera.

II. PROGRESS SINCE 2002 TO 2015

Daniel De Menthon [1] et al. (2002) Each pixel of a 3D space-time video stack is feature point whose coordinates include three color components two motion angle components and two motion position components which are mapped to a 7D. The clustering of these feature points provides color segmentation and labeling of regions over time which amounts to region tracking.For this task in this algorithm adopted a hierarchical clustering method which operates by repeatedly applying mean shift analysis over increasingly of large ranges,with weights equal to the counts of the points that contributed to the clusters.This technique has lower complexity for large the mean shift of radii than ordinary mean shift analysis because it can use binary tree structures more during range search efficiently. In addition, it provides a hierarchical segmentation of the data.Applications include video compression and compact descriptions of sequences for video indexing and retrieval applications.

Rita Cucchiara [2] et al. (2003) Background subtraction methods are widely most exploited for moving object detection in videos in many applications, such as monitoring of traffic, human motion of capture, and under video surveillance.This work proposes a general-purpose method that combines the statistical of assumptions with the object level knowledge of moving objects, apparent objects, and shadows of acquired in the processing of the previous frames. Pixels belonging to moving objects, ghosts, and shadows are processed differently an object-based selective update in order to supply . The proposed approach exploits color information for the both of background subtraction and shadow detection to improve object segmentation and background update. The approach proves that it is fast, flexible, and precise in terms of both pixel accuracy and re-activity to background changes.

O. Le Meur [3] et al. (2005) A new spatio-temporal model for bottom up visual attention simulating is proposed. It has been built from numerous properties of the Human Visual System that are important.the paper focuses both on the architecture of the model and performances of its.the spatial model of the bottom-up visual attention has already been defined that is given, the temporal dimension is described more accurately. A qualitative and quantitative comparison with human fixations collected from an eye tracking the apparatus is undertaken. From the former, consists of the sum all visual features the quality of the prediction is deemed very good where as the latter illustrates that the best predictor of the human fixation.

Yun Zhai [4] et al. (2006) Human vision system actively seeks the interesting regions in images that are used to reduce the search effort in tasks, such as object detection and recognition of object .The prominent actions in video sequences are more likely to attract our first is, which is estimated by Applying RANSAC on point correspondences in the scene.To compensate of interest-points the non-uniformity of spatial distribution, spanning areas of motion segments are in- Corporate in the time.In the spatial attention model, a fast method for the computing of the pixel-level maps of saliency has been developed using color of the histograms of images. A hierarchical spatial attention representation is established to reveal the interesting points that are used in images as well as the interesting regions. Finally, a dynamic fusion technique is applied to combine both of the temporal and spatial maps of saliency, where temporal attention is dominant over the spatial model when large motion contrast that exists.

Sebastian Nowozin [5] et al. (2007) The approaches to action classification in the videos that have used sparse spatio-temporal words encoding local appearance around movements that are interesting. Most of approaches used as a histogram representation, discarding the temporal order of the among features. where the global temporal order of motions is the important to the discriminate between the two.In this technique work is used to use a sequential representation which retains this temporal order. Further,in this paper introduce Discriminative Sub sequence Mining to find optimal discriminative sub sequence patterns. In combination with the LP Boost classifier,that amounts to simultaneously learning a classification function and performing the features of selection in the space of all possible feature sequences. The resulting classifier linearly combines a small number of decision functions The approaches to action classification in the videos that have used sparse spatio-temporal words encoding local appearance around movements that are interesting. Most of approaches used as a histogram representation, discarding the temporal order of the among features. where the global temporal order of motions is the important to discriminate between the two.In this work the propose technique is used to use a sequential representation which retains this temporal order. Further,in this introduce Dis criminative Sub sequence Mining to find optimal dis criminative subsequence patterns.

In combination with the LP Boost classifier,that amounts to simultaneously learning a classification function and

performing feature selection in the space of all possible feature sequences.

K. Rapantzikos [6] et al. (2007) The human visual system has the ability to quickly fixate on the most informative regions of the scene and thats why reducing the inherent of uncertainty visuals.Computational visual attention schemes have been approach to account for the characteristic of the HVS that is uncertainty. A video analysis framework based on a spatio temporal model is presented. A novel scheme has been generating saliency in video sequences by taking into account both the spatial extent and dynamic evolution of regions.The goal that is used for image-oriented computational model of saliency-based visual is extended to handle spatio temporal analysis of video in a volumetric framework of attention. The main claim is that attention acts as an efficient preprocessing step to obtain a representation of the visual content in the form of salient events and objects compact.The model has been implemented, and qualitative as well as quantitative examples illustrating its performance are shown.

Chenlei Guo [7] et al. (2008) Salient areas in natural scenes are generally that regarded as the candidates of attention focus in human eyes, which is the key stage in object of the detection. In computer vision, many models have been proposed to simulate the behaviour of eyes , Neuro morphic Vision Toolkit and etc., but they demand high computational cost and their results mostly rely on the parameters.A simple and fast approach based on Fourier transform called spectral residual was proposed, which used amplitude spectrum to obtain the saliency map of SR. The results are good, but the reason is questionable. In this paper,they propose it is the phase spectrum, not the amplitude spectrum, of the Fourier transform that is the key in the obtaining of location of salient areas.they provide some examples to show that the PFT can get better results comparison with SR and requires less computational complexity as well PFT can be easily extended from a two-dimensional Fourier transform to a Quaternion Fourier Transform if the value of each pixel is represented as a quaternion composed of intensity, color and feature of motion.

Konstantinos Rapantzikos [8] et al. (2009) Spatial and temporal detectors the have been used in video analysis for action recognition.Feature points are detected using a number of the measures,saliency of names,periodicity, motion activity etc.Each of these measures is usually intensity-based and provides a different trade-off between the density and informativeness.In this paper,they use saliency for feature point of detection in the videos and incorporate color and motion apart from intensity.There method uses a multi-scale of the volumetric representation of the video and involves spatiotemporal operations at the voxel level. Saliency is computed by a global minimization of process constrained by pure volumetric constraints, each of them being related to an informative visual aspect of , namely spatial proximity, scale and feature similarity. Points are selected as the extrema of taliency response and prove to balance well between density and informativeness.they provide an intuitive view of the detection of points and visual comparisons the against state-of-the-art space-time detectors.

Sophie Marat [9] et al. (2009) It presents a spatio-temporal saliency model that predicts the eye movement during video free of viewing.This model is inspired by the biology of the first steps of human visual the system.The model extracts two signals from video of the stream corresponding to the two outputs of the mains retina: parvocellular and magnocellular. Then, both signals are split into the elementary feature maps by cortical-like filters.These feature maps are used to the form of two saliency maps:static and the dynamic one. These maps are then fused into a spatio-temporal saliency map. The model is evaluated by comparing the salient areas of each of frame predicted by the spatio-temporal saliency map to the eye positions of different subjects of during a free video viewing experiment with a large database. In parallel, the static and the dynamic pathways are finds to understand that what is more or less salient and for what type of videos our model is a good or a poor predictor of eye movement.

Konstantinos Rapantzikos [10] et al. (2009) In this paper,they propose and elaborate on a saliency detection model that treats a video of sequence as a spatio temporal volume and generates the local saliency measure for each visual unit .This computation involves an optimization process that incorporating of inter- and intrafeature competition at the voxel of level. Perceptual decomposition of the input, spatio temporal of the interactions and of the integration of center-surround heterogeneous feature values are described conspicuity and an experimental framework for video classification is set up.The framework consists of a series of results in division of performance center-surround shows and let us draw conclusions on how to well the detected salient regions represent the visual input**.**

Chenlei Guo [11] et al. (2010) Salient areas in the natural scenes are generally regarded as the areas which the human eye will typically focus on the, and finding the these areas is the key step in object of detection.In computer vision many models have been proposed to simulate the behavior of eyes such as Saliency Tool Box, Neuro morphic Vision Toolkit, and others, but they demand high cost of computational and computing the useful results mostly relies on their choice of parameters. Although some of the region-based approaches were proposed to reduce the computational complexity of feature maps, these approaches were not able to work in the real time. Recently, a simple and fast approach called spectral residual was proposed,uses of the amplitude spectrum to calculate the images of map of saliency.In the work,they pointed out that it is the phase spectrum, not the amplitude of spectrum, of an image's Fourier transform that is key to calculating the location of salient areas, and proposed the phase of spectrum of Fourier transform model. In this paper,they present a quaternion is intensity, color, and motion features is composed.

Tong Yubing and Faouzi Alaya Cheikh [12] et al. (2011) A sequence of still images are called the video sequence. It contains a strong spatial–temporal correlation between the regions of consecutive the frames. The most important characteristic of videos is the across the frames perceived motion foreground objects .The motion of foreground objects dramatically changes the importance of the objects in the scene and different saliency map of the frame to be lead

representing the scene.This makes the saliency analysis of videos of still images much more than complicated. In this paper, they investigates the saliency in video sequences and that is proposed a novel spatiotemporal saliency model devoted for video surveillance applications.Both bottom-up and top down attention the mechanisms are involved in this model. Stationary saliency and motion saliency are, respectively, analyzed. First, a new method for the background of subtraction and foreground of extraction is developed based on content analysis of the scene in the domain of video the surveillance. Then, a stationary saliency model is setup based on multiple features that are computed from the foreground. Every feature is analyzed with a multi-scale Gaussian pyramid, and all the features maps are combined conspicuity using different weights. The stationary model integrates faces as a supplement feature to other low level features such as the color, intensity and orientation. Second, a motion saliency map is calculated using the statistics of the motion that are vectors field.

Yaping Zhu [13] et al. (2011) In this paper an adaptive spatiotemporal saliency algorithm for video attention Detection using motion vector decision an adaptive spatiotemporal saliency algorithm for video attention Detection using motion vector decision,of video sequences for visual system in humans.in this novel system can detect the saliency regions quickly by using classic motivated by the importance of motion information. Motion vectors are calculated by block matching and optical flow are used to determine decision of condition. When significant motion contrast occurs , the saliency area is detected by motion and intensity features.Otherwise,the motion is low, color and orientation of contrast features are added to form a more detailed saliency map. visual system in humans.in this novel system can detect the saliency regions quickly by using classic motivated by the importance of motion information.

Amir H. Shabani [14] et al. (2012) A sparse and compact representation is used for Local spatio-temporal salient features of video contents in many computer vision for human action recognition of tasks.To localize these features , existing methods perform either symmetric or asymmetric filtering of multi-resolution temporal and use a structural or a motion saliency criteria.

In a common framework for action classification discriminative, different saliency criteria Of the structured-based detectors and different temporal filters of the motion-based detectors are compared.they have two main observations.

Anna Belardinelli [15] et al. (2012) High level visual cognitive abilities analysis are modulated by attentive selective processes such as scene understanding and behavioural.The extraction of moving objects is a step of crucial in the processing of dynamic scenes.

Motion is course a powerful for grouping regions in cue and segregating objects. On a coarse level, most interesting moving objects with coherent motion are associated.

Rajkumar Kannan [23] et all (2015) The Detection of videos has many applications in computer.The novel spatiotemporal salient region detection is approached.The proposed approach estimating spatial and temporal of saliencies separately by computing. The spatial saliency is computed by estimation of color contrast cue and color distribution cue. The estimations of these cues exploit the patch level and region level image abstractions in a unified way. The aforementioned cues are fused to compute an initial spatial saliency map, which is refined to emphasize saliencies of objects uniformly for further, and to suppress saliencies of background noises. The final computation is done by spatial saliency map by integrating the refined saliency map with center prior map. The based on local and global temporal saliencies the temporal salienct is computed the estimations using patch level optical flow of abstractionsin last,to generate a spatiotemporal saliency map the spatial and temporal saliencies are integrated.

Petros Koutras [24] et all (2015) The paper developed a new spatio-temporal visual front end based on biologically inspired 3D Gabor filters, which is applied on both of luminance and the color streams are produces spatio-temporal energy maps.These volumes are fused for computing a saliency map on single and can detect the spatio-temporal phenomena that static saliency models cannot find.They provide a new movie database with eye-tracking annotation.the paper evaluated the spatio-temporal saliency model on the widely used on new database using different fusion schemes and feature sets.The proposed spatio-temporal computational framework that incorporates many ideas based on psychological evidences and yields significant improvements on spatio-temporal saliency estimation.

Lijuan Duan [25] et all (2015) Accurately modelling and predicting the visual attention behavior of human viewer scan help a video analysis algorithm search interesting regions by reducing effort of tasks.A great variety of model on attention for predicting the direction on images and videos have been proposed.When a views of human on video,motions of video and greatly affect the distribution the visual fixations.They develop models that lead to motion features that are extracted from videos and used in new video for detection method.The frames are partitioned into blocks on which saliency calculations are made to achieve the efficiency.

The motion features extracted from each of block are differences of motion vectors between adjacent frames.

III. COMPARISON TABLE

The comparison of various algorithms are as follows on the basis of authors name's,paper name ,algorithm and techniques.he different types of algorithms are used for this comparison.

TABLE1: Comparison of Various Algorithms

Year	Authors name	Paper name	Technique	Algorithm	Challenges
2002	Deniel de menthon, Remi Megret	Spatiotemporal Segmentation of Video by Hierarchical Mean Shift Analysis.	In this paper approach that is used for spatiotem poral segmenta tion of video sequences for indexing and also for retrieval application, video com- pression.	Algorithm which is used is for space and time segmentation of video seq uence,Regions are segme nted and tracked by the same mechanism;one parameter is specied by the user.	Efficient ways required to develop for using video strands,hierarchical segmentat-ion for indexing and retrieval of large video data sets and also for compression.
2003	Rita Cucchiara, Member, IEEE, Costantino Grana,Massimo Piccardi, Member, IEEE, and Andrea Prati, Member, IEEE.	Detecting Moving Objects, Ghosts, and Shadows in Video Streams.	In this paper the proposed method exploits color information for back ground subtraction and shadow detection to make better results for object segmentation and background update.	The technique has given results fast, flexible, and precise in terms of both shape accuracy and reactivity to background changes. These results are due to the integration of some form of object-level knowledge into a statistical background model.	The technique applied is very computationally cost-effective as it is not severe in compu-tational time
2005	Piotr Dollar Vincent Rabaud Garrison Cottrell Serge Belongie	Behavior Recognition via Sparse Spatio-Temporal Features	In this paper it is shown that the viability of applying behaviour recognition by characterizing behavior in terms of spatiotemporal features.	Recently introduced spatio-temporal interest point detector was represented, and a number of cuboid descriptors were analyzed and observed in this paper.	Future work involves using the spatio-temporal layout or framework of the features, extending such approaches as to the spatio-temporal domain. Using features detected at multiple scales may also improve performance.
2006	Yang Liu, Christos-Savvas Bouganis, Peter Y K. Cheung	A Spatiotemporal Saliency Framework	This paper shows an unsupervised spatiotemporal saliency layout that shows the results of combination of low-level processing and statistical inference.	The basic idea of this paper is the representation of the spatiotemporal saliency as the unpredictability in the motion of the feature.	The main problem in this paper is Other motion models that give rise to saliency under different circumstances.
2006	Yun Zhaim University of Central Florida Orlando,Mubara k Shah	Visual Attention Detection in Video Sequences Using Spatiotemporal Cues	In this paper, it is presented that spatiotemporal attention detection framework for detecting attention regions in video sequences.	The proposed framework in this paper is spatiotemporal attention framework that has been applied on 20 testing video sequences, And attended regions are detected to highlight interesting objects and motions present in the sequences with high user satisfaction.	Another part of this paper will be the combination of the presented bottom-up approach with the top-down technique which can result for both higher efficiency and accuracy in object classification tasks.
2007	K. Rapantzikos, N. Tsapatsoulis, Y. Avrithis and S. Kollias	Bottom-up spatiotemporal visual attention model for video analysis	A approach that exploits spatiotemporal information for video analysis based on the concept of saliency-based VA has been presented.	In this paper Fusion of the proposed bottom-up spatiotemporal visual attention method with prior knowledge for putting up new applications in the field involving classification	In this paper Segmentation and tracking may be the focus of our future research.

2007	Sebastian Nowozin,Gokhan Bakır and Koji Tsuda,	Discriminative Subsequence Mining for Action Classification	In this paper a new approach has been proposed a novel classifier for sequence representations, which is suitable for action classification in videos	A main goal of this work is to result for efficient pattern selection algorithms from the data mining community which is accessible to the computer vision community.	In this paper the future work will be applying defined approach to classify higher order action patterns. Due to the lack of an openly available action classification data set for this kind of high level actions is currently a problem.
2008	Chenlei Guo, Qi Ma and Liming Zhang	Spatio-temporal SaliencyDetection Using Phase Spectrum of Quaternion Fourier Transform	This paper proposed a method which is called PQFT to calculate spatiotemporal saliency maps used to detect salient objects in both natural images and videos.	In this paper, a technique is proposed which is the phase spectrum, and not the amplitude spectrum, of the Fourier transform which is the key in getting the location of salient areas.	The potential proposed work lies in the engineering fields, which can be extended to the application like object recognition, video coding and etc.
2009	Sophie Marat Tien Ho Phuoc Lionel Granjon Nathalie Guyader Denis Pellerin · Anne Guérin-Dugué	Modelling Spatiotemporal Saliency to Predict Gaze Directionfor Short Videos	This paper proposed a method which is called PQFT to calculate spatiotemporal saliency maps used to detect salient objects in both natural images and videos.	This work shows a new bottom-up saliency model inspired by the biology of the basic steps of the human visualsystem. Model may also be used to improve video compression and can be added to camera motion analysis to support selected frames for a summary of the video.	In future work it would be interesting to add more features such as color or a spatially varying sampling of the retina depending on eye position sto reinforce our model.
2009	KonstantinosRapantzikos NicolasTsapatsoulis , Yannis Avrithis Stefanos Kollias	Spatiotemporal saliency for video classification	This work presents a computational model for saliency detection that exploits the spatiotem poral structure of a video stream and produces a per voxel saliency measure based on a feature competition approach	The framework described consists of a series of experiments which shows the effect of saliency in classification performance also drawn conclusions on how well the detected salient regions represent the visual input. A comparison is done which shows the potential of the proposed method.	In future work it would be interesting to add more features such as color or a spatially varying sampling of the retina depending on eye position sto reinforce our model.
2009	KonstantinosRapantzikos NicolasTsapatsoulis , Yannis Avrithis Stefanos Kollias	Dense saliency-based spatiotemporal feature points for action recognition	In this paper, saliency is used for feature point detection in videos and incorporate color, motion apart from intensity.	In this paper a novel spatiotemporal feature point detector, which is based on a computational model of saliency. Saliency is resulted as the solution of an energy minimization problem which is initiated by a set of volu - metric feature conspi cities- derived from intensity, color and motion.	The motivation is taken by recent works, the focus can be computational efficiency issues also on the incorporate on of advanced spatio temporal descriptors as the ones proposed.
2010	Chenlei Guo, Liming Zhang,	A Novel Multi resolution Spatio temporal Saliency Detection Model and Its Applications in Image and Video Compression	In this paper, we present a quaternion representation of an image which is composed of intensity, color, and motion features	Phase spectrum of quaternion Fourier transform (PQFT) is proposed in this paper to calculate the spatiotemporal saliency map of an image by its quaternion representation.	Based on the principle of PFT, a novel multiresolution spatiotemporal saliency detection model called (PQFT).

2011	Tong Yubing, Faouzi Alaya, Cheikh Fahad Fazal, Elahi Guraya, Hubert Konik, Alain Tre´meau,	A Spatiotemporal Saliency Model for Video Surveillance	. In this paper, we investigate saliency in video sequences and propose a novel spatiotemporal saliency model devoted for video surveillance applications.	First, a new method for background subtraction and foreground extraction is developed based on content analysis of the scene in the domain of video surveillance. Then, a stationary saliency model is setup based on multiple features computed from the foreground.	In the next step, we will focus on more complicated scene where background and foreground objects are both moving. More refinedalgorithm should be necessary to get the suitable foreground objects for saliency analysis.
2011	Yaping Zhu1, Natan Jacobson, Hong Pan, and Truong Nguyen	Motion-Decision Based Spatiotemporal Saliency For Video Sequences	When motion contrast is low, color and orientation features are added to form a more detailed saliency map. Experimental results show that the proposed algorithm can detect salient objects and actions in video sequences robustly andefficiently.	The first contribution of the paper presents theextraction of motion vectors for video type classification, which will be used in saliency feature selection. The other contribution is in the adaptive fusion method, combining spatial (contrast-based) and temporal (motion-based) saliency maps.	The computational complexity is reduced by selectively choosing the dominant features to use.
2012	Anna Belardinelli, Andrea Carbone, Werner X. Schneide	Classification of multiscale spatiotemporal energy features for video segmentation and dynamic objects prioritisation.	In this study we focused on the natural movies set. Video sequences are provided with correspond - ing eye-tracking raw data of 54 subjects.	the system delivers reasonable results for both the segmentation and the prioritisation of meaningful objects in a scene. It should be noted that the whole process relies on the same features computed in the beginning and that these are only motion features.	A further improvement would consist in the discrimination between object and self-motion.
2012	Amir H. Shaban, David A. Clausi, John S Zelek,	Evaluation of Local Spatio-temporal Salient Feature Detectors for Human Action Recognition	They have two main observations.The motion-based detectors localize features which are more effective than those of structured-baseddetectors.	In a common discriminative framework for action classification, we compared different salient structured-based and motion-based feature detectors.	They recommend the use of asymmetric motion filtering for effective salient feature detection, sparse video content representation, and consequently, action classification.
2013	Kang-Ting Hu,Jin-Jang Leou,Han-Hui Hsiao	Spatiotemporal saliency detection and salient region determinationfor H.264 videos	In this study, a spatiotemporal saliency detection and salient region determination approach for H.264 videos is proposed. After Gaussian filtering in Lab color space, the phase spectrum of Fourier transform is used to generate the spatial saliency map of each video frame.	Finally a modified salient region determination scheme is used to determine salient regions (SRs) of each video frame. Based on the experimental results obtained in this study, the performance of the proposed approach is better than those of two comparison approaches.	Then, the spatial and temporal saliency maps of each video frame are combined to obtain the spatiotemporal saliency map using adaptive fusion.
2013	Shen Hao, Li Shuxiao, Zhu Chengfei, Chang Hongxing, Zhang Jinglan	Moving object detection in aerial video based on spatiotemporal saliency	In this paper, the problem of moving object detection in aerial video is addressed. While motion cues have been extensively exploited in the literature, how to use spatial information is still an open problem.	In this paper, we utilize spatiotemporal saliency in moving object detection. Temporal and spatial saliency is extracted in a hierarchical manner, and both pixel saliency and region saliency are extracted to give a full illustration for spatial distribution.	However, as the detection algorithms estimate objectlocations in every frame independently, false alarms are unavoidable. We will deal with this by combining tracking information in our future study.

2014	Abdalrahman Eweiwi, Muhammad Shahzad Cheema, Christian Bauckhage	Action recognition in still images by learning spatial interest regions from videos.	They propose a novel method for extracting spatial interest regions where we apply non-negative matrix factorization to optical flow fields extracted from videos.	Experimental evaluation shows that approach is able to extract interest regions that are highly correlated to those body parts most relevant for different actions.generative model achieves high accuracy in action classification.	compared to conventional part based approaches, our approach does not assume an underlying elastic model of body but provides priors even for cluttered scenes orimages of partly occluded human bodies.
2014	Kevis Maninis, Petros Koutras And Petros Maragos	Advances On Action Recognition In Videos Using An Interest Point Detector Based On Multiband Spatio-Temporal Energies	This paper proposes a new visual framework for action recognition in videos, that consists of an energy detector coupled with a carefully designed multiband energy based filterbank.	They proposed a new video energy tracking method that relies on detection of multiband spatiotemporal modulation components.	As future work, we would like to focus on improvement of action localization and extention of our experimental comparisons to more databases
2014	Guruprasad Somasundaram, Anoop Cherian, Vassilios Morellas, Nikolaos Papanikolopoulos	Action recognition using global spatiotemporal features derived from sparse representations	In this paper, we propose a novel global spatio-temporal self-similarity measure to score saliency using the ideas of dictionary learning and sparse coding.	In contrast to existing methods that use local spatio-temporal feature detectors along with descriptors, dictionary learning helps consider the saliency in a global setting in computationall efficient way.	These are currently investigating such optimization methods to improve our current classification performance especially for smaller sets of actions such as in real world applications.
2015	Petros Koutras, Petros Maragos,	A Perceptually-based Spatio-Temporal Computational Framework for Visual Saliency Estimation	The purpose of this paper is to demonstrate a perceptually-based spatio-temporal computational framework for visual saliency estimation.	We have developed a new spatio-temporal visual frontend based on biologically inspired 3D Gabor filters, which is applied on both the luminance and the color streams and produces spatio-temporal energy maps.	As future work, we focus on the reduction of the frontend's complexity and integration of our bottom-up frontend with the movies' high level semantic information.
2015	LijuanDuan, TaoXi, SongCui, HonggangQi, AlanC.Bovik	A spatiotemporal weighted dissimilarity-based method for video saliency detection	The proposed a new video saliency detection model for detecting salient regions on video, which combines two spatial features with one motion feature	These demonstrated the effectiveness of our model on four kinds of video datasets and found that it delivers highly competitive performance.	A novel feature fusion technique was applied to combine the pro-posed temporal and spatial features
2015	Nasim Souly, Mubarak Shah	Visual Saliency Detection Using Group Lasso Regularization in Videos of Natural Scenes	The propose to use group lasso regularizationto find the sparse representation of a video, which benefits from grouping information provided by super-voxels and extracted features from the cuboids	They show that without using data labeling, and learning techniques requiring eye movement data, we are able to determine salient regions of videos accurately.	In summary, we present an entirely unsupervised bottom-up method that detects the regions of videos to which people'seyes are drawn.
2015	Hansang Kim, Youngbae Kim, Jae Young Sim, and Chang-Su Kim	Spatiotemporal Saliency Detection for Video Sequences Based on Random Walk With Restart	The proposed algorithm obtains the spatialtransition probability matrix using the features of intensity, color, and compactness. Finally, the proposed algorithm computes the stationary distribution of the random walker as the overall saliency.	A novel saliency detection algorithm for video sequences based on the random walk with restart (RWR) is proposed in this paper.	Two or more objects results are not accurate.

Iv. CONCLUSION

In this paper with help of table various methods for spatiotemporal saliency detection has been explained.The best technique that comes out of all the techniques is robust saliency detection for video sequences based on random walk with restart.this is because the technique explained in this paper results in good color compaction also is a robust based noval algorithm.this technique become best as it uses trasition probability matrix.All the techniques that has been explained for spatiotemporal saliency detection includes temporal restarting distribution and noise removal. his is a comparison of feature, score, match and decision level fusions performed on biometric modalities.The accuracy percentage range is between 85%-99% range, which is very accuracy .It is explained above in detail how fusion at an earlier stage is the best for multi modal systems and this survey concretely puts forward this point.The amount of information goes on decreasing as one proceeds from sensor level to decision level.

V. REFRENCES

[1] D. De Menthon. Spatio-temporal segmentation of video by hierarchical mean shift analysis.In Statistical Methods in Video Processing Workshop, 2002.

[2] R. Cucchiara, C. Grana, Massimo Piccardi, and A. Prati, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams,"IEEE Trans.PAMI *vol.25-10(2003) 1337-1342*.

[3] O. Le Meur,D.Thoreau,P.Le Callet And D. Barba" A Spatio-Temporal Model Of The Selective Human Visual Attention".In: Proceedings Of Ieee International Conference On Image Processing , III-1188-1191(2005).

[4] Yun Zhai, Mubarak Shah, "Visual Attention Detection in Video Sequences Using Spatiotemporal Cues", ACM MM 2006, Santa Barbara, CA, USA.

[5] S Nowozin, G Bakir, K Tsuda,"Discriminative subsequence mining for action classification",Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 1-8.

[6] K Rapantzikos, N Tsapatsoulis, Y Avrithis,S Kollias,"Bottom-up spatiotemporal visual attention model for video analysis",Image Processing, IET 1 (2), 237-24.

[7] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.

[8] J. He, M. Li, H.-J. Zhang, H. Tong, and C. Zhang, "Manifold-ranking based image retrieval," in *Proc. ACM Int. Conf. Multimedia*, 2004,pp. 9–16.

[9] T. Lu, Z. Yuan, Y. Huang, D. Wu, and H. Yu, "Video retargeting with nonlinear spatial-temporal saliency fusion," in *Proc. IEEE ICIP*,Sep. 2010, pp. 1801–1804.

[10] R. P. N. Rao, G. J. Zelinsky, M. M. Hayhoe, and D. H. Ballard, "Eye movements in iconic visual search," *Vis. Res.*, vol. 42, no. 11,pp. 1447–1463, Nov. 2002.

[11] V. Navalpakkam and L. Itti, "Top–down attention selection is fine grained," *J. Vis.*, vol. 6, no. 11, pp. 1180–1193, Oct. 2006.

[12] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach.Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[13] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. 11th ACM Int. Conf. Multimedia*,Nov. 2003, pp. 374–381.

[14] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequencytuned salient region detection," in *Proc. IEEE CVPR*, Jun. 2009,pp. 1597–1604.

[15] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Haung, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. IEEE CVPR*,Jun. 2011, pp. 409–416.

[16] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp.1915–1926, Oct. 2012.

[17] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE CVPR*, Jun. 2013, pp. 1155–1162.

[18] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 545–552.

[19] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *IEEE Trans. Image Process.*, vol. 19,no. 12, pp. 3232–3242, Dec. 2010.

[20] J.-S. Kim, J.-Y. Sim, and C.-S. Kim, "Multiscale saliency detection using random walk with restart," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 198–210, Jun. 2013.

[21] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE CVPR*, Jun. 2013, pp. 3166–3173.

[22] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE CVPR*, Jun. 2007, pp. 1–8.

[23] B. Schauerte and R. Stiefelhagen, "Quaternion-based spectral saliency detection for eye fixation prediction," in *Proc. ECCV*, 2012,pp. 116–129.

[24] J. Li and W. Gao, *Visual Saliency Computation: A Machine Learning Perspective*. New York, NY, USA: Springer, 2014.

[25] C. Jia, F. Hou, and L. Duan, "Visual saliency based on local and globalfeatures in the spatial domain," *Int. J. Comput. Sci.*, vol. 10, no. 3,pp. 713–719, 2013.

[26] O. Boiman and M. Irani, "Detecting irregularities in images and in video," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 17–31, 2007.

[27] X. Hou and L. Zhang, "Dynamic visual attention: Searching for coding length increments," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 681–688.

[28] L. Zhang, M. H. Tong, and G. W. Cottrell, "Sunday: Saliency using natural statistics for dynamic analysis of scenes," in *Proc. 31st Annu. Cognit. Sci. Conf.*, 2009, pp. 2944–2949.

[29] Y. Xue, X. Guo, and X. Cao, "Motion saliency detection using lowrank and sparse decomposition," in *Proc. IEEE ICASSP*, Mar. 2012, pp. 1485–1488.

[30] H. J. Seo and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," *J. Vis.*, vol. 9, no. 12, pp. 1–27, Nov. 2009.

[31] Y. Li, Y. Zhou, L. Xu, X. Yang, and J. Yang, "Incremental sparse saliency detection," in *Proc. 16th IEEE ICIP*, Nov. 2009, pp. 3093–3096.

[32] Y. Li, Y. Zhou, J. Yan, Z. Niu, and J. Yang, "Visual saliency based on conditional entropy," in *Proc. Asian Conf. Comput. Vis.*, 2009, pp. 246–257.

[33] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. ACM Int. Conf. Multimedia*, 2006, pp. 815–824.

[34] S. Marat, T. H. Phuoc, L. Granjon, N. Guyader, D. Pellerin, and A. Guerin-Dugue, "Modelling spatio-temporal saliency to predict gaze direction for short videos," *Int. J. Comput. Vis.*, vol. 82, no. 3, pp. 231–243, 2009.

[35] J. Peng and Q. Xiaolin, "Keyframe-based video summary using visual attention clues," *IEEE Trans. Multimedia*, vol. 17, no. 2, pp. 64–73, Apr./Jun. 2010.

[36] X. Xiao, C. Xu, and Y. Rui, "Video based 3D reconstruction using spatio-temporal attention analysis," in *Proc. IEEE ICME*, Jul. 2010, pp. 1091–1096.

[37] W. Kim, C. Jung, and C. Kim, "Spatiotemporal saliency detection and its applications in static and dynamic scenes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 446–456, Apr. 2011.

[38] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 3, pp. 334–352, Aug. 2004.

[39] J. B. Kim and H. J. Kim, "Efficient region-based motion segmentation for a video monitoring system," *Pattern Recognit. Lett.*, vol. 24,nos. 1–3, pp. 113–128, 2003.

[40] H. Li and K. N. Ngan, "Saliency model-based face segmentation and tracking in head-and-shoulder video sequences", *J. Vis. Commun. Image Representation*, vol. 19, no. 5, pp. 320-333, 2008

[41] Y.-T. Chen and C.-S. Chen, "Fast human detection using a novel boosted cascading structure with meta stages", *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1452-1464, 2008

[42] L. Itti , C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254-1259, 1998

[43] W.-H. Chen , C.-W. Wang and J.-L. Wu, "Video adaptation for small display based on content recomposition", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, pp. 43-58, 2007

[44] M. Rubinstein , A. Shamir and S. Avidan, "Improved seam carving for video retargeting", *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1-8, 2008

[45] J.-S. Kim , J.-H. Kim and C.-S. Kim, "Adaptive image and video retargeting technique based on Fourier analysis", *Proc. IEEE Comput. Vision Pattern Recognition*, pp. 1730-1737, 2009

[46] Y. F. Ma and H. J. Zhang, "Contrast-based image attention analysis by using fuzzy growing", *Proc. ACM Int. Conf. Multimedia*, pp. 374-381, 2003

[47] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention", *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304-1318, 2004

[48] D.-Y. Chen , H.-R. Tyan , D.-Y. Hsiao , S.-W. Shih and H.-Y. M. Liao, "Dynamic visual saliency modeling based on spatiotemporal analysis", *Proc. IEEE ICME*, pp. 1085-1088, 2008

[49] W. Kim , C. Jung and C. Kim, "Saliency detection: A self-ordinal resemblance approach", *Proc. IEEE ICME*, pp. 1260-1265, 2010
[50] V. Gopalakrishnan , Y. Hu and D. Rajan, "Salient region detection by modeling distributions of color and orientation", *IEEE Trans. Multimedia*, vol. 11, pp. 892-905, 200
[51] C. Kim and B. Vasudev, "Spatiotemporal sequence matching for efficient video copy detection",*IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, pp. 127-132, 2005
[52] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression", *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185-198, 2010
[53] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 171-177, 2010
[54] D. Walther and C. Koch, "Modeling attention to salient proto-objects", *Neural Netw.*, vol. 19, no. 9, pp. 1395-1407, 2006
[55] Y. Fu , J. Cheng , Z. Li and H. Lu, "Saliency cuts: An automatic approach to object segmentation",*Proc. IEEE ICPR*, pp. 1-4, 2008.