ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

Deep Learning Features Based Video Saliency Detection Model for Effective Compression

K.Pushpalatha[1], P.Manimegalai[2], G.Kanaga[3], R.Vedhapriyavadhana[4]
[1,2,3] UG Student, Department of ECE, Francis Xavier Engineering College, Tirunelveli.
[4]Associate Professor, Department of ECE, Francis Xavier Engineering College, Tirunelveli.

Abstract: Saliency detection is widely wont to extract regions of interest in images for various image pro-cessing applications. Recently, many saliency detection models are proposed for video in uncompressed (pixel) domain. However, video over Internet is usually stored in compressed domains, like MPEG2, H.264, and MPEG4 Visual. Deep learning features are recently addressed by the researcher thanks to its distinct involvement in image also as video feature extraction. During this paper, we propose a com-pletely unique video saliency detection model supported feature contrast in compressed domain. Four sorts of features including luminance, color, texture, and motion are extracted from the discrete cosine transform coefficients and deep learning features in video bit stream. The static saliency map of unpre-dicted frames (I frames) is calculated on the basis of luminance, color, and texture features, while the motion saliency map of predicted frames (P and B frames) is computed by deep learning feature. A re-placement fusion method is meant to mix the static saliency and motion saliency maps to urge the ulti-mate saliency map for every video frame. Thanks to the directly derived features in compressed domain, the proposed model can predict the salient regions efficiently for video frames. Results on the database show superior performance of the proposed video saliency detection model in compressed domain.

Keywords: Deep learning, Saliency Detection, image Processing

I. INTRODUCTION

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a way wider range of algorithms to be applied to the input file and may avoid problems like the build-up of noise and signal digital image processing could also be modeled within the sort of Multidimensional Systems.

Many of the techniques of digital image processing, or digital picture processing because it often was called, were developed within the 1960s at the Jet Propulsion Laboratory, Massachusetts Institute of Technology, Bell Laboratories, University of Maryland, and a couple of other research facilities, with application to satellite imagery, wire-photo standards conversion, medical imaging, videophone, character recognition, and photograph enhancement. The value of processing was fairly high, however, with the computing equipment of that era. That changed within the 1970s, when digital image processing proliferated as cheaper computers and dedicated hardware became available. Images then might be processed in real time, for few dedicated problems like television standards conversion. As general-purpose computers became faster, they begin to take over the role of dedicated hardware for about the foremost specialized and computer-intensive operations. With the fast computers and signal processors available within the 2000s, digital image processing has become the most common sort of image processing and usually, is employed because it is not only the most versatile method, but also the most cost effective.

II. PROPOSED SYSTEM

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

Saliency detection has recently attracted an excellent amount of research interest. The reason behind this growing popularity lies in the effective use of these models in various vision tasks, such as image segmentation, object detection, video summarization and compression, to name a few. Saliency models can be broadly classified into two categories: human eye fixation prediction or salient object d g to the sort of input, they will be further c tic and dynamic saliency models. While st still images as input, dynamic models w quences. In this paper, we focus on d ive regions in dynamic scenes. C al networks (CNNs) are successfully u fundamental areas of computer vision, i tection, semantic segmentation and still s

Saliency Map

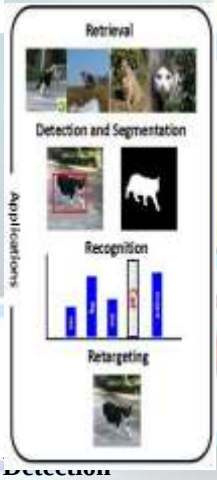
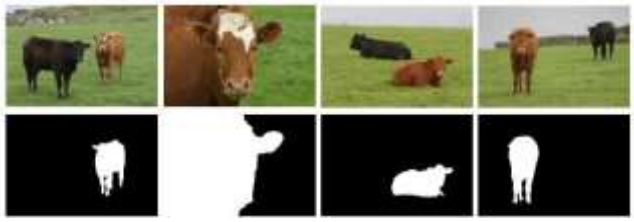


Fig. Salient

Saliency Detection:

Saliency detection aims to detecting the salient regions automatically, which has been applied in image/video segmentation, image/video retrieval, image retargeting, video coding, quality assessment, action recognition, and video summarization. The last decade has witnessed the remarkable progress of image saliency detection, and a plenty of methods have been proposed based on some priors or techniques, such as uniqueness prior, background prior, compactness prior, sparse coding, random walks, and deep learning.



Co-saliency Detection

Co-saliency detection aims at detecting the common and salient regions from an image group containing multiple related images, while the categories, intrinsic attributes, and locations are entirely unknown. Therefore, the inter-image correspondence among multiple image-plays a useful role in representing the common attribute.

Fig. Co-saliency Detection

Video saliency detection aims at continuously locating the motion-related salient object from the given video sequences by considering the spatial and temporal information jointly. we divide the video saliency detection methods into two categories, i.e., low-level cues based method and learning based method.

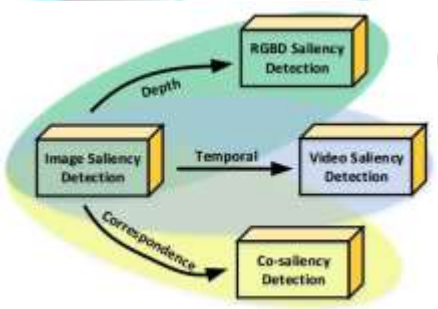


Fig. Relationship and Comprehensive Information

The image saliency detection model is the basis for other three models. With the acquisition technology development, more comprehensive information is available, such as the depth cue for RGBD data, the inter-image constraint for image group, and the temporal relationship for video data.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

Motivations and Contributions.

Depth map evaluation. A good depth map can be benefit for the saliency detection no matter how it produced. According to the observation of depth distribution, a confidence measure for depth map is proposed to reduce the influence of poor depth map on saliency detection; Extension of compactness. A novel model for compactness integrating color and depth information is put forward to compute the compactness saliency; Multiple cues fusion. A foreground seeds' selection mechanism based on depth refined is presented. The saliency is measured against this between the target regions with seed regions, which integrate color, depth, and texture cues.

Inter Saliency Detection

Acquiring the corresponding relationship among multiple images is the key point of co-saliency detection model. Within the proposed model, the matching methods on two levels are designed to represent the correspondence among multiple images.

The primary one is the superpixel-level similarity matching scheme, which focuses on determining the matching superpixel set for the present superpixel based on three constraints from other images. The second is that the image-level similarity measurement, which provides a World Wide relationship on the entire image scale.

With the corresponding relationship, the inter saliency of a excellent pixel is defined because the weighted sum of the intra saliency of corresponding super pixels in other images.

Optimization and Propagation

Within the proposed method, the optimization of saliency map is casted as a "label propagation" problem, where the uncertain labels are propagated by using two sorts of certain seeds, i.e. background and salient seeds. The proposed CLP method is employed to optimize the intra and inter saliency maps during a cross way, which suggested the propagative seeds are crosswise interacted. The cross seeding strategy optimizes the intra and inter saliency maps jointly, and improves the robustness.

The first problem of applying CNNs to video saliency that the lack of sufficiently large, densely labeled video training data. As far as we know, the successes of CNNs in computer vision are largely attributed to the availability of large-scale annotated images (e.g., Image Net). However, existing video datasets are too small to provide adequate training data for CNNs. In Table 1, we list the statistics of the Image Net dataset and widely adopted video object segmentation datasets, including FBMS, SegTrackV2, VSB100 and DAVIS.

It is often observed that, the prevailing video datasets rarely match existing image datasets like Image Net, in either quality or quantity. Besides, considering the high correlation between the frames from same video clip, existing video datasets are far unable to meet the satisfy the requirements of coaching CNNs for pixel- level video applications, like video salient object detection. On the opposite hand, for the instant, creating such a large-scale video dataset is typically infeasible, because annotating videos is complex and time-consuming. To the present end, we propose a video data augmentation approach to synthetically generating labeled video training data, which explicitly leverages existing large- scale image segmentation datasets. The simulated video data are easily accessible and rapidly generated, on the brink of realistic video sequences and present various motion patterns, deformations, companied with automatically generated annotations and optical flow. The experimental results via these automatically generated videos clearly demonstrate the practicability of our strategy.

Our video data synthesis approach clears the underlying challenge for learning CNNs for several applications in video processing, where dynamic saliency detection is of no exception. Another challenge for detecting saliency in dynamic scenarios derives from the natural demand of this task. As suggested by human beholding research when computing dynamic saliency maps, video saliency models got to consider both the spatial and therefore the temporal characteristics of the scene. We propose a deep video saliency model for producing spatiotemporal saliency via fully exploring both the static and dynamic saliency information. The proposed model adopts fully convolutional networks (FCNs) for pixel-wise saliency prediction. Related to existing rich image saliency data, the static saliency is deeply exploited and explicitly encoded within the deep learning process via transferring and fine-tuning recent

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

success in image classification. For learning dynamic saliency cues, the proposed deep video saliency model learns from an outsized number of labelled videos, including both human-generated and natural video data, during a supervised learning mode. The static saliency is integrated into dynamic saliency detection process, thus for directly producing final spatiotemporal saliency estimation.

Another important contribution of this work is that our deep video saliency model is much more computationally efficient compared with existing video saliency models. Salient object detection may be a key step in many image analysis tasks because it not only identifies relevant parts of a visual scene but can also reduce computational complexity by filtering out irrelevant segments of the scene. In recent years, some notable video saliency models are proposed and show usefulness in many computer vision applications, like video segmentation and video re-timing. However, time efficiency becomes the common major bottleneck for the applicability of existing video saliency algorithms; most computation time has been spent for optical flow computation. Additionally, from the attitude of learning deep networks in dynamic scenes, many schemes take optical flow as input, causing high computational expenses.

In this work, we propose a both effective and efficient video saliency model, which frees itself from the computationally expensive optical flow estimation. One among the key insights of this paper is that, unlike high-level video applications such as action detection, video saliency can derive from short-term analysis of video frames. Thus we directly capture temporal saliency via learning deep networks from frame pairs, rather than using long-term video information, like optical flows from multiple adjacent video frames.

We comprehensively evaluate our method on the FBMS dataset, where the proposed video saliency model produces more accurate saliency maps than state-of-the-arts. Meanwhile, it achieves a frame rate of 2fps (including all steps) on a GPU. Thus it is a practical video saliency detection model in terms of both speed and accuracy. We also report results on the newly released DAVIS database and observe performance improvements over current competitors.

Our methods are computationally efficient, much faster than traditional video saliency models and other deep networks in dynamic scenes.

Saliency detection has been extensively studied in computer vision, and saliency models generally can be categorized into visual attention prediction or salient object detection. The previous methods attempt to predict scene locations where a person's observer may fixate. Salient object detection aims at uniformly highlighting the salient regions, which has been shown benefit to a good range of computer vision applications. More detailed reviews of the saliency models are often found. Saliency models are often further divided into static and dynamic ones consistent with their input. During this work, we aim at detecting saliency object regions in videos.

Image saliency detection has been extensively studied for many years and most of the methods are driven by the well- known bottom-up strategy. Early bottom-up models are mainly based on detecting contrast, assuming salient regions in the visual field would first pop out from their surroundings and computing feature-based contrast followed by various mathematical principles. Meanwhile, another mechanisms are proposed to adopt some prior knowledge, like background prior, or global information, to detect salient objects in still images. More recently, deep learning techniques are introduced to image saliency detection. These methods typically use CNNs to look at an outside number of region proposals, from which the salient objects are selected. Currently, more and more methods tend to find out in an end-to-end manner and directly generate pixel-wise saliency maps via fully convolutional networks(FCNs).

Compared with saliency detection in still images, detecting saliency in videos may be a far more challenging problem thanks to the complication within the detection and utilization of temporal and motion information. So far, only a limited number of algorithms are proposed for spatiotemporal saliency detection. Early models are often viewed as simple extensions of exiting static saliency models with extra temporal dimension. Some more recent and notable approaches to the present task are proposed, showing inspired performance and good potentials in many computer vision applications. However, the applicability of those approaches is severely limited by their high-computational costs. The main computational bottleneck

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

comes from optical flow estimation, which contributes much to the promising results.

In recent years, the border of saliency detection has been reach capturing common saliency among related images/videos inferring the salient event with video sequences or scene understanding. However, there are significant differences between above methods and traditional saliency detection, especially considering their goals and core difficulties.

III.DEEP LEARNING MODELS

We mainly specialize in famous, deep learning models for computer vision applications in dynamic scenes, including action recognition, object segmentation, object tracking attention prediction and semantic segmentation and explore their architectures and training schemes. This may help to clarify how our approach differs from previous efforts and can help to spotlight the important benefits in terms of effectiveness and efficiency.

Many approaches directly feed single video frames into neural networks trained on image data and adopt various techniques for post-processing the results with temporal or motion information. Unfortunately, these neural networks hand over learning the temporal information which is often very important in video processing applications.

A famous architecture for training CNNs for action recognition in videos is proposed during which incorporates two- stream convolutional networks for learning complementary information on appearance and motion. Other works adopt this architecture for dynamic attention prediction and video object segmentation. However, these methods train their models on multi-frame dense optical flow, which causes heavy computational burden.

In the areas of human pose estimation and video object processing, online learning strategy is introduced for improving performance. Before processing an input video, these approaches generate various training samples for fine-tuning the neural networks learned from image data, thus enabling the models to be optimized towards the thing of interest within the test video sequence. Obviously, these models are quite time-consuming and therefore the fine-tuned models are only specialized for specific classes of objects.

We show the chances of learning to detect generic salient objects in dynamic scenes by training on videos and pictures via a completely offline manner. A completely unique technique for synthesizing video data via leveraging large amounts of image training data. The CNNs model are often efficiently and completely trained on rich video sequences and pictures, thus successfully learning both static and dynamic saliency features. Meanwhile, it directly learns inner relationship between frames, getting obviate of time-consuming motion computation. Thus, our algorithm is significantly faster than traditional video saliency methods and therefore the deep learning architectures that demand optical flow as input.

IV.DEEP NETWORKS FOR VIDEO SALIENCY DETECTION

During this work, we describe a procedure for constructing and learning deep video saliency networks employing a novel synthetic video data generation approach. Our approach generates a outsized amount of video data (150K paired frames) from existing image datasets and associates these annotated video sequences with existing video data to find out deep video saliency networks. We first introduce the proposed CNNs based video saliency model.

Architecture Overview

We start with an summary of our deep video saliency model before going into details below. At a high level, we feed frames of a video into a neural network, and therefore the network successively outputs saliency maps where brighter pixels indicate higher saliency values. The network is trained with video sequences and images and learns spatiotemporal saliency in general dynamic scenes. Fig. 1 shows the architecture of proposed deep video saliency model. Inspired by classical human beholding research, which suggests both static and dynamic saliency cues contribute to video saliency, we design our model with two modules, simultaneously considering both the spatial and temporal characteristics of the scene.

The first module is for capturing static saliency, taking single frame image as input. It adopts fully convolutional networks (FCNs) for generating pixel-wise saliency estimate and utilizes previous excellent pre-trained models on large-scale image datasets. Boosted from rich image saliency benchmarks, this module is efficiently trained for capturing diverse static

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

saliency information of interesting objects. The second module takes frame pairs and static saliency from the first module as input, and generates final dynamic saliency results. This network is trained from both synthetic and real labelled video data.

Deep Networks for Static Saliency

A static saliency network takes a single frame image as input and produce a saliency map with the same size of the input. We model this process with a fully convolutional network (FCN). The bottom of this network is a stack of convolutional layers. Convolutional layer is defined on shared parameters (weight vector and bias) architecture and has translation invariance characteristics. The input and output of each convolutional layer are a set of arrays, called feature maps, with size h w c, where h, w and c are height, width and the feature or channel dimensionality, respectively. For the first convolutional layer, the input is the color image , with pixel size h and w, and three channels. At the output, each feature map indicates a particular feature representation extracted at all locations on the input, which is obtained via convolving the input feature map with a trainable linear filter (or kernel) and adding a trainable bias parameter. If we denote the input feature map as X, whose convolution filters are determined by the kernel weights W and bias b, each convolutional layer, point-wise nonlinearity (e.g., ReLU) is applied for improving feature representation capability. Additionally, convolutional layers are often followed by some form of non-linear down-sampling (e.g., max pooling). This leads to robust feature representation which tolerates small variations within the location of input feature map.

Due to the stride of convolutional and feature pooling layers, the output feature maps are coarse and reduced-resolution. However, for saliency detection, we are more curious about pixel-wise saliency prediction.

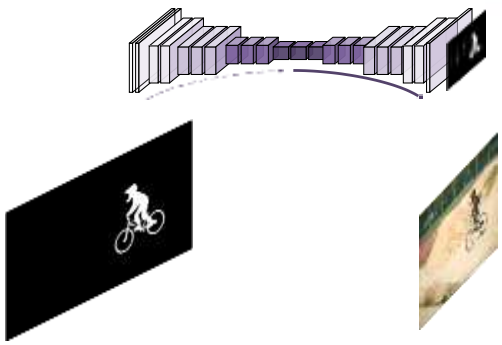


Fig. 2. Illustration of our network for static saliency detection.

The network takes single frame image (for example, 224 224) as input, adopting multi-layer convolution networks that transforms the input image to multidimensional feature representation, then applying a stack of deconvolution networks for upsampling the feature extracted from the convolution networks. Finally, a fully convolution network with 11 kernel and sigmoid activity function is used to output of a probability map in the same size as input, in which larger values mean higher saliency values.

Deep Networks for Dynamic Saliency

Now we describe our spatiotemporal saliency network. As depicted in Fig. 3, the network has a similar structure as our static saliency network, which is based on FCN and includes multi-layer convolution and deconvolution nets. The dynamic network learns dynamic saliency information jointly with the static saliency results, thus directly generating spatiotemporal saliency estimates.

The training set consists of a collection of synthetic and real video data, which efficiently utilizes existing large-scale well-annotated image data (described in Sec. IV). More specifically, we feed successive pair of frames (I_t, I_{t+1}) and the ground truth G_t of frame I_t in the training set into this network for capturing dynamic saliency. Meanwhile, since saliency in dynamic scenes is boosted by both static and dynamic saliency information, the network incorporates the saliency estimate P_t generated by static saliency network as saliency priors indicative of potential salient regions. Thus our dynamic saliency network directly generates final spatiotemporal saliency estimates for frame I_t ,which is achieved via exploring dynamic saliency cues and leveraging static saliency prior from the static saliency network.

We train the proposed architecture in an end-to-end manner. It is commonplace to initialize systems for several of vision tasks with a prefix of a network trained for image classification. This has shown to substantially reduce training time and improve accuracy. During training, our convolutional layers are initialized with the weights in the first five convolutional exploring dynamic saliency cues and leveraging static saliency prior from the static saliency network.

We concatenate frame pair (I_t, I_{t+1}) and static saliency size of h w 7. Then we feed **I** into our FCN based P_t

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

within the channel direction, thus generating a tensor **I** with dynamic saliency network, which has similar architecture of static saliency network. Only the primary convolution layer is modified accordingly: blocks of VGG Net, which was originally trained over 1.3 million images of the Image Net dataset .

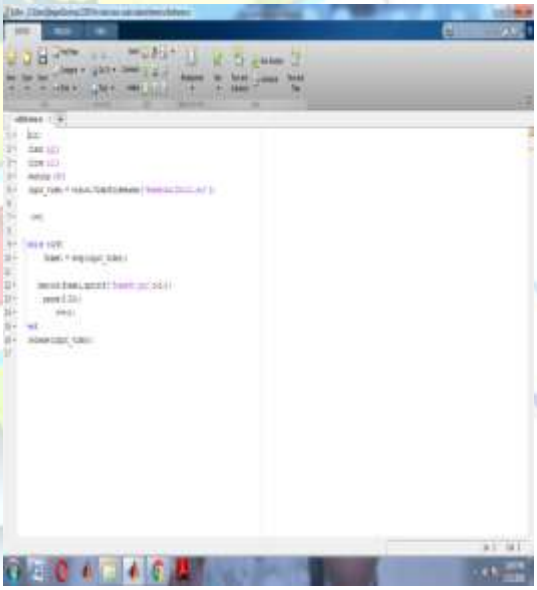
Successive frame pairs (I_t, I_{t+1}) from real video data or synthesized from existing image datasets (described in Sec. IV), and static saliency information inferred from our static saliency network, are concatenated and fed into the dynamic network, which features a similar FCN architecture with the static network. The dynamic network captures dynamic saliency, and considers static saliency simultaneously, thus directly generating spatiotemporal saliency estimation .where g_i G and p_i P ; α refers to ratio of salient pixels in ground truth G. where Ws represent corresponding convolution kernels; b is bias parameter. During training, stochastic gradient descent (SGD) is used to attenuate the weighted cross-entropy loss described before. After training, given a frame image pair and static saliency prior, the deep dynamic saliency model is in a position to output final spatiotemporal saliency estimate. For testing, we first detect the static saliency map P_t for frame I_t via our static saliency network. Then frame image pair (I_t, I_{t+1}) and therefore the static saliency map P_t are fed into the dynamic saliency network for generating the ultimate spatiotemporal saliency for frame I_t. After obtaining the video saliency estimate for frame I_t, we keep iterating this process for subsequent frame I_{k+1} until reaching the top of the video sequence. More implementation details are often found. Qualitative and quantitative study of the effectiveness of our dynamic saliency model is described.

Compared with the favored two-stream network structure used. We merge the output of the static network into the dynamic saliency model, which directly produces spatiotemporal saliency results. This architecture brings two advantages. Firstly, the fusion of dynamic and static saliency is explicitly inserted into the dynamic saliency network, instead of training two-stream networks for spatial and temporal features and specially designing a fusion network for spatial and temporal feature integration. Secondly, the proposed model directly infers the temporal information from two adjacent frame rather than previous methods using optical flow images, thus our model gaining higher computation efficiency.

where I is that the input image; $F_S(\ )$ denotes the output feature S; D_S denotes the de convolution layers that up sample the map generated by the convolutional layers with total stride of input by an element of S to make sure an equivalent spatial size of the output Y and therefore the input image I. The de convolution operation is achieved via reversing the forward and backward passes of corresponding convolution layer. All the parameter s of convolution and de convolution layers are learnable. with a 11 kernel is adopted for mapping the feature map Finally, on the top of the network, a convolutional layer Y into a particular saliency prediction map P through a sigmoid activation unit. We use the sigmoid layer in order that each entry in the output has a realvalueintherangeof0and 1. Thanks to the utilization of FCN, the network is allowed to work on input images of arbitrary sizes and preserves spatial information.

V. RESULT AND DISCUSSION

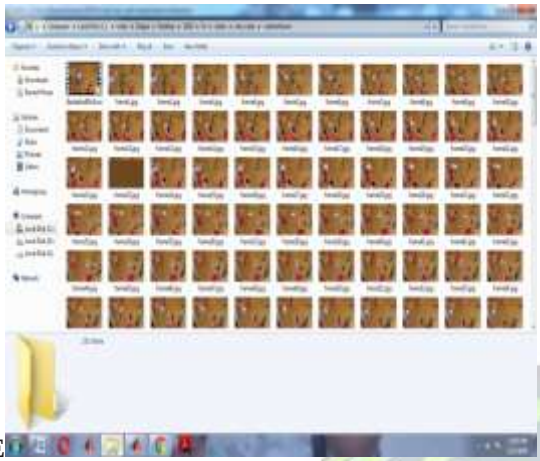
VIDEO TO FRAME CODING



The salient objects within the video should be salient in each individual frame

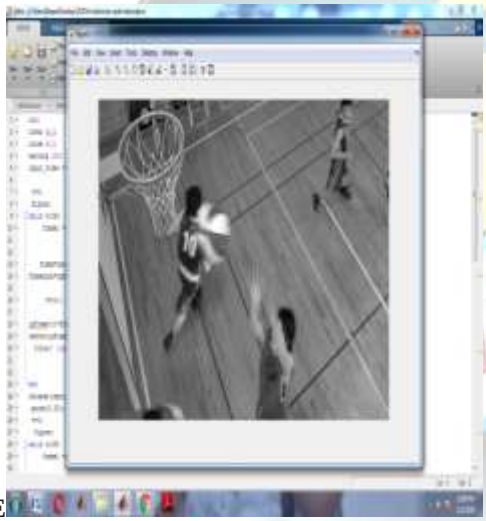
ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020



FRAME

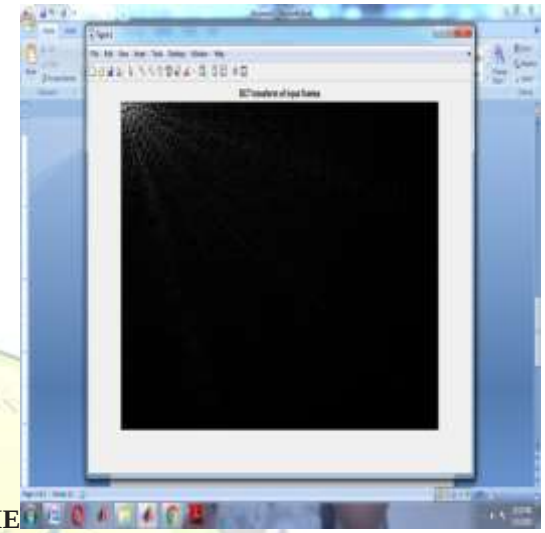
The salient objects in the video should be salient in each individual frame



INPUT FRAME

A single frame processed for detecting the salient object is shown in the above figure.

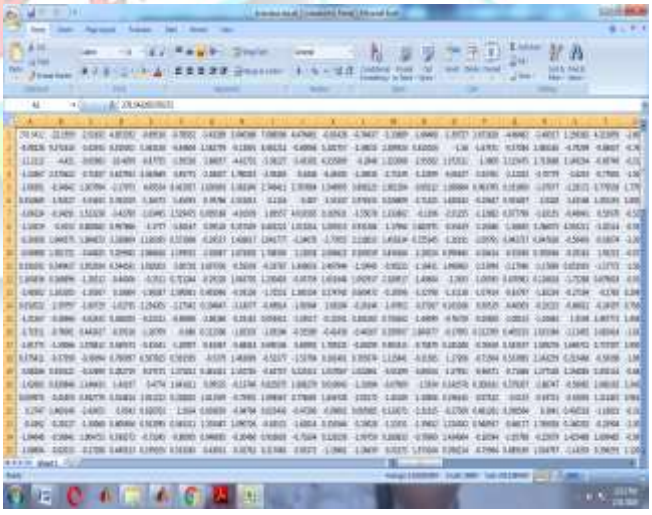
DCT TRANSFORM INPUT



FRAME

- It helps separate the image into parts (or spectral sub-bands) of differing importance (with reference to the image's visual quality).
- It is almost like the discrete Fourier transform: it transforms a signal or image from the spatial domain to the frequency domain.

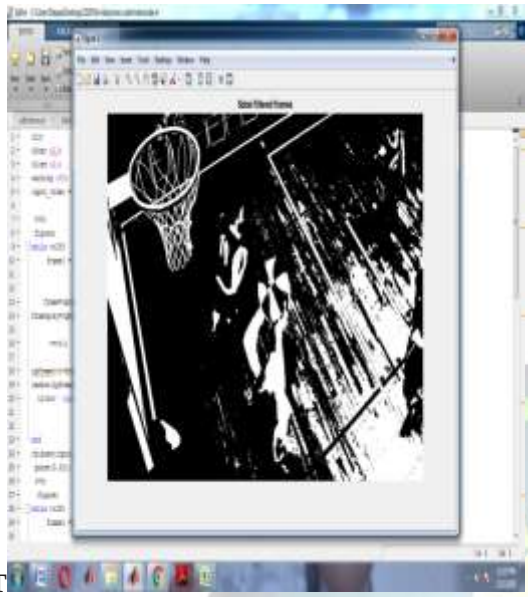
DCT OUTPUT VALUE



SOBEL FILTER

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

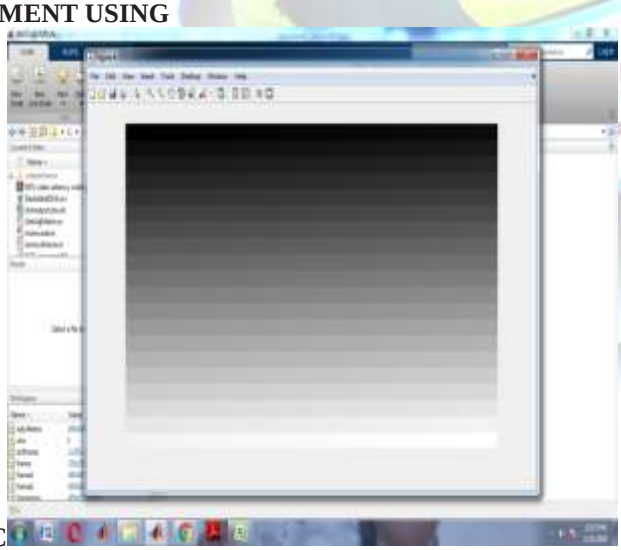


OUTPUT

Sobel Filter

The Sobel operator, sometimes called the Sobel–Feldman operator or Sobel filter, is employed in image processing and computer vision, particularly within edge detection algorithms where it creates a picture emphasising edges**.**

SEGMENT USING



SLIC

- It is the state of the art algorithm to segment super pixels which doesn't require much computational power.
- In brief, the algorithm clusters pixels within the combined five-dimensional color and image plane space to efficiently generate compact, nearly uniform super pixels.

CNN Classification

- This type of architecture is dominant to acknowledge objects from an image or video.
- It is a class of deep learning neural networks. CNNs represent a huge breakthrough in image recognition. They're most ordinarily wont to analyze visual imagery and are frequently working behind the scenes in classification.

VI. CONCLUSION

In this paper, the co-saliency detection model which introduces the inter-image correspondence constraint to discover the common salient object in an image group is used. Our comprehensive analysis demonstrated that the proposed method outperforms the state-of-the-art saliency, co-saliency, and video saliency models. In the future, we plan to incorporate our models into a deep learning framework.

REFERENCES

[1] H. Fu, D. Xu, and S. Lin, "Object-based multiple foreground segmen-tation in RGBD video," IEEE Trans. Image Process., vol. 26, no. 3, pp. 1418–1427, Mar. 2017.

[2] X. Cao, C. Zhang, H. Fu, X. Guo, and Q. Tian, "Saliency-aware nonparametric foreground annotation based on weakly labeled data," IEEE Trans. Neural Netw. Learn. Syst., vol. 27, no. 6, pp. 1253–1265,Jun. 2016.

[3] W. Wang, J. Shen, Y. Yu, and K.-L. Ma, "Stereoscopic thumbnail cre-ation via efficient stereo saliency detection," IEEE Trans. Vis. Comput.Graphics, vol. 23, no. 8, pp. 2014–2027, Aug. 2017.

[4] W. Wang, J. Shen, and H. Ling, "A deep network solution for attention and aesthetics aware photo cropping," IEEE Trans. Pattern Anal. Mach.Intell., to be published.

[5] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," IEEE Trans. Image Process., vol. 25, no. 12, pp. 5664–5677, Dec. 2016.

[6] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, and P. Han, "Hierarchical features driven residual learning for depth map super-resolution," IEEE Trans.Image Process., vol. 28, no. 5, pp. 2545–2557, May 2019.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

[7] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "Learning sparse representa-tion for objective image retargeting quality assessment," IEEE Trans.Cybern., vol. 48, no. 4, pp. 1276–1289, Apr. 2018.

[8] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image qual-ity assessment," IEEE Trans. Multimedia, vol. 20, no. 8, pp. 2035–2048, Aug. 2018.

[9] L. Zhou, Z. Yang, Q. Yuan, Z. Zhou, and D. Hu, "Salient region detection via integrating diffusion-based compactness and local contrast," IEEETrans. Image Process., vol. 24, no. 11, pp. 3308–3320, Nov. 2015.

[10] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in Proc. CVPR, Jun. 2014, pp. 2814–2821.
pp.X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detec-tion via dense and sparse reconstruction," in Proc. ICCV, Dec. 2013, 2976–2983.

[11] N. Li, B. Sun, and J. Yu, "A weighted sparse coding framework for saliency detection," in Proc. CVPR, Jun. 2015, pp. 5216–5223.

[12] Y. Yuan, C. Li, J. Kim, W. Cai, and D. D. Feng, "Dense and sparse labeling with multidimensional features for saliency detection," IEEETrans. Circuits Syst. Video Technol., vol. 28, no. 5, pp. 1130–1143,May 2018.