Journal of Engineering Technology (ISSN: 0747-9964)
Volume 6, Issue 1, Jan, 2017, PP.146-163

ORIGINAL ARTICLE

Visual Attention Modelling and Applications in Image and Video Processing

Uzair Ishtiaq[1] and Tanzila Saba[2]

[1]Computer Science Department CIIT Vehari Campus Pakistan
[2]College of Computer and Information Sciences Prince Sultan University Riyadh, 11586 Saudi Arabia

Abstract: Humans attention is normally drawn towards visually salient or distinct stimuli. Actually, identifying any or all of interesting targets will be computationally complex. Visual Saliency helps to identify the target in bottom-up or top-down fashions. If the target is stimulus-driven then it would be in a bottom-up manner and if it is user-driven then it would be in a top-down manner. During the past few decades, the researchers have developed such methods and models that could be applied to different frames of a video or an image and find out its visual distinctiveness. There are a number of applications that are exhibited by visual attention modeling as well such as image and video quality assessment, video summarization that includes video skimming and key frame extraction based on visual saliency. Accordingly, this paper presents a methodology of visual attention modelling and their applications in image and video processing. Additionally, current models in this field are compared on benchmark dataset and future directions are suggested.

Keywords: Visual saliency; Video summarization; Video Attention; Features mining.

1. Introduction

According to William James, Principles of Psychology , 1890, attention is a well-known concept to all of us. It refers to what human mind takes possessions of any one thing out of several others to be observed simultaneously, in a clear and transparent manner. At the same moment, humans are not paying attention to others but concentrating only one point of interest. Visual attention is a theoretical notion which corresponds to the attitude depicted by a person to keep concentration over a particular target. This is non-linear in nature. It means that naturally a human being cannot concentrate on all portions of his conspicuity area of a frame. Actually, a human could attend to one thing at a time and the attention helps us to make a decision to move eyes in a specific direction. Human perception is dependent on three basic factors, which are attention, eye movement and memory. The spatial region all over the center of gaze within which the target can be sensed in the first glimpse or fixation is known as the conspicuity area where the gaze is the center of a stable intentional look, and fixation is, sustaining the gaze in a constant direction [1-5].

Previously, the psychophysical measurement procedures were slow and complex but the researchers have presented such psychophysical procedures that assess complete conspicuity area while encompassing full awareness of the target while the scene was holding its location. With such method, the target can easily be resolved from its surroundings [6-10]. The distinctive property of an image or any frame of a video that how much distinct it is from its background is known as visual saliency. According to human observer studies, it has come to know that saliency is one of the greatest properties of interest, gaze allocation and attention of a person when he is freely viewing and observing any static or any dynamic image [11-16]. Image saliency is the practical concept of visual

ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Journal of Engineering Technology (ISSN: 0747-9964)
Volume 6, Issue 1, Jan, 2017, PP.146-163
ORIGINAL ARTICLE

attention. Image saliency is helpful to find how much an image or an object is distinct from rest of its background. However, saliency can be measured computationally [17-22].

Fig. 1: Visual Saliency of different images

Figure 1 is showing the Visual Saliency of different images, the salient parts of the images are highlighted when these images are applied different saliency techniques [23-28].

Visual saliency is done in the following two ways:

- Bottom-up Visual Saliency
- Top-Down Visual Saliency

1.1 Bottom-Up Visual Saliency:

Bottom-up visual saliency is a stimulus-driven signal, which means that the saliency is depended over the stimulus [29,30]. The factors affecting bottom-up image saliency include color, luminance, orientation contrast, etc.

- For example: Consider there is a green field. However, if there is any red object in that field, it will surely capture attraction towards itself, which is in a bottom-up manner [31-35].

Fig. 2: Original Image

Fig. 3: Bottom-up visual saliency effects on Fig 2

Figure 2 is showing some models of houses where all the houses are blue except one which is red in color. Figure 3 is showing the resultant visually salient image of the houses when a bottom-up visual saliency technique is applied on it. The red house can easily be figured out [36].

AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Journal of Engineering Technology (ISSN: 0747-9964)
Volume 6, Issue 1, Jan, 2017, PP.146-163
ORIGINAL ARTICLE

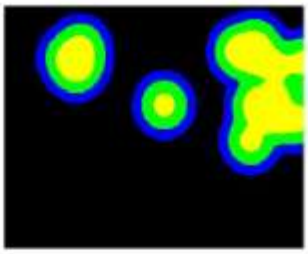
1.2 Top-Down Visual Saliency:

Top-Down visual saliency is a user-driven signal, which refers that the saliency is depended on the user. The factors affecting top-down image saliency include: how and where the image tend to appear in the scene [37-40].

For example: Consider a toy basket and someone is looking into it for the cricket ball then it will be noticed how and where all of our toys located, still continuously searching for the cricket ball, which is in a top-down manner [41-44].

Targets: paintings

Fig. 4



Subject Consistency

Fig. 5

Figure 4 presents the actual image in which the target is defined to be the paintings [45]. In figure 5 a top-down image saliency technique is applied and the resultant image is obtained in which the paintings are highlighted.

2. Models for Visual Saliency

To calculate visual saliency maps, several algorithms have been proposed during the past decade. With the use of these algorithms, a given input image could be converted into its consequent scalar-valued map and these maps are normally compared with human observer data for its validity. Researchers have presented a number of computational saliency models, including, Attention-based on Information Maximization (AIM) [2], Saliency Using Natural Image Statistics (SUN) [3], Spectral Whitening (SW) [4], Edge Distance Saliency (EDS) [5], Saliency Toolbox 2.2 [6,7], Extended Saliency (ESaliency) [8], Multiscale Contrast Conspicuity (MCC) [1], Rarity Based Saliency [9], Graph-Based Visual Saliency (GBVS) [10], Phase Spectrum of Quaternion Fourier Transform (PQFT) [11], Frequency-Tuned Saliency (FTS) [12] and Adaptive Whitening Saliency [13].

There can be a lot of applications for these visual attention models, like, image and video quality assessment [14], video summarization [15], progressive image transmission [16], image segmentation [17], object recognition [18] and so forth.

3. Related Work

In this section, first, the human visual conspicuity and what is conspicuity measurement procedure for humans is explained [46-50]. Additionally, several, computational saliency models are analyzed.

3.1 Human Visual Conspicuity

The area all around the center of gaze and attention from which the target could be identified in a single sight is known as visual conspicuity. If the target is implanted in the background of the image, then the conspicuity area would be small, whereas, it would be large if the target is clear and

Journal of Engineering Technology (ISSN: 0747-9964)
Volume 6, Issue 1, Jan, 2017, PP.146-163

not embedded in its background. The conspicuity measurement procedure is given in the following steps [1]:

(i) The observer fully observes the target to achieve full knowledge of its locality.
(ii) He marks a point that is at a large angular space from the point where the target as well as it is located in front of the parallel position to the target. This is the initial fixation.
(iii) Then the observer changes fixations repeatedly and thus gradually gets closer to the target until it is identified in its spatial region.
(iv) The angular distance between the initial fixation and target's center is recorded.
(v) This process is repeated for three times.

Conspicuity estimates have the following two types.

(i) Detection Conspicuity:

It imitates the concept of bottom-up saliency in which the previous target fixation is not that much important for the detection of the target. It finds out how obviously the target area differs from its background area [51-55].

(ii) Identification Conspicuity:

Top-down saliency concept is depicted in identification conspicuity where the observer uses target features to identify the target which he is interested in. It is used to find out how obviously the image details show that whether it is the target or it can be extracted as the target [56-63].

3.2 Computational Saliency Models

Following are some of the computational saliency models that are used to find the distinctive property of an image or a video [64-70].

(a) AIM: Attention-based on Information Maximization

Bruce and Tsotsos [2] present a bottom-up attention model in such a way that they wanted to maximize the information from the sampled scene. The purpose of AIM is to convert the image feature plane into visual saliency map using Shannon's self-information measure. The core idea of AIM was that how much saliency an image feature is providing associated with its local surroundings [71-75]. Bruce [19] presented an image operator which is used to calculate the gaze in random natural scenes and landscapes when a human is freely viewing them, based on local information and statistics. In this technique, a scene is sampled in such a way that maximum information is obtained from the scene under consideration [76-80].

(b) SUN: Saliency Using Natural Image Statistics

The SUN, Saliency Using Natural Image Statistics [3] is used by local image features to calculate their bottom-up saliency. The core idea is to detect the important targets i.e. the important features of the image. In SUN, both the saliency techniques, i.e. bottom-up and top-down saliencies are used. The bottom-up saliency is used in free viewing when the target is not specified whereas top-down saliency is used when the target is specified [81-85].

(c) RBS: Rarity-Based Saliency

(i) Global Rarity: It takes the intensity of the local image and calculates its mean and variance.
(ii) Local Rarity: It takes the intensity of the local image and calculates its contrast [9,86,87].

(d) EDS: Edge Distance Saliency

In EDS [5], first, the original image is decomposed into edge transformed images which are then thresholded at different gray-scale levels, thus binary edge images are produced. Now these binary edge images are then combined to produce the saliency map. Saliency and the pixel values in this map are not directly related but inversely related to each other, i.e. the pixel that is near the edge has lower value and the pixel that is away from the edge has comparatively higher value [88].

(e) Saliency Toolbox 2.2

Saliency Toolbox 2.2 [6,7] uses contrast, color, and luminance maps and computes its saliency map. The core idea behind Saliency Toolbox 2.2 is that local contrast is important for the target.

(f) GBVS: Graph-Based Visual Saliency

In Graph-Based Visual Saliency [10], dissimilarity feature maps are allocated distance-weights to calculate saliency. In this approach, the image under consideration is converted into a specific representation that resembles a pyramid with three levels.

- At the first level, intensity map is computed.
- Then the color map and then the orientation distribution map is calculated.
- Then a fully linked graph representation is created for these three feature maps.

After the creation of these above-mentioned map representations, weights are allocated among different nodes that are not directly proportional rather they are inversely proportional to the similarity that was found in the feature values and their spatial distance among different nodes [89].

(g) SW: Spectral Whitening

In SW (Spectral Whitening) [4], a map is constructed that only focuses on salient features and ignores non-informative background information that is of no use, this is an example of top-down saliency. This is similar to the common human eye phenomenon to focus on the informative features of a scene and ignore the other non-informative features.

(h) PQFT: Phase Spectrum of Quaternion Fourier Transform

Information of the interesting details is provided by the phase spectrum while considering any image. In this model of image saliency, phase spectrum of an image's quaternion Fourier transform is used to compute the saliency map [11]. This saliency map is generated by representing each pixel by a quaternion (a group of four) that comprises of color, intensity and motion features.

(i) FTS: Frequency Tuned Saliency

In FTS (Frequency Tuned Saliency) [12], the target is fixated by the local image feature contrast. Contrast feature is used for color and luminance, which gives the concept of bottom-up saliency.

(j) SDSR: Saliency Detection by Self-Resemblance

In SDSR [20], the target is focused by local feature contrast. The contrast matrix of the pixel under consideration is compared with the contrast matrices of the nearby pixels and their measured similarity is used to calculate the saliency map. SDSR is an example of bottom-up saliency map.

(k) Esaliency: Extended Saliency

In Esaliency (Extended Saliency) [8], the image that is given as an input is fragmented and then the saliency is calculated as the global dissimilarity of those segments. According to this technique:

AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Journal of Engineering Technology (ISSN: 0747-9964)
Volume 6, Issue 1, Jan, 2017, PP.146-163
ORIGINAL ARTICLE

(i) Natural landscapes are made up of slam components having similar feature properties.
(ii) If there are two or more segments that seem similar, they all can either be targets or non-targets.
(iii) Total targets in a scene are a few.

(l) MCC: Multiscale Contrast Conspicuity

For human fixation, intensity contrast is an important property. In MCC (Multi-scale Contrast Conspicuity) [1] saliency is computed based on intensity contrast of the target. Target contrast is the relation between the intensity of the supposed target and the surrounding area of the target including the target itself.

Therefore, for this first, these two values will be calculated. Surrounding area's contrast is calculated by progressively growing the width of target surround area. When the target and the total area which includes the target and surround area are equal, the target contrast is equal to 1 as there is no surrounding area at all. And if the target and the total area which includes the target and surround area are not equal, that means that the value of contrast will be between greater than 0 and less than 1.

Target contrast will increase if the surrounding area is minimized, conversely, it will decrease if surround area is maximized. MCC (Multi-scale Contrast Conspicuity) is an example of bottom-up saliency.

4. Benchmark comparison

This section compares different models employed for the visual attention measurement in state of the art using benchmark dataset.

4.1 The Models

In SUN [3], the low-level features of an image are taken into account to detect the important feature of a given input image, low-level features of an image include, color, contrast, luminance, are used to find the bottom-up image saliency. SW [4] is used to construct a saliency map which only concentrates on the significant and informative elements of an image while ignoring the non-informative parts. In EDS [5], the actual image under consideration is first decomposed into edge transformed smaller images. These smaller images are given thresholds with respect to their grayscale levels which forms binary edge images. Then these binary edge images are fused with one another and ultimately a saliency map is formed. Pixel values of the saliency may are inversely related, which means that those pixels which are close to the edge will be having lower values and the pixels which will be away from the edge will be having higher values. In MCC [1], the saliency is detected on the basis of target contrast at different scales. First, the target intensity is calculated and then again intensity is calculated by gradually increasing the target surround area. Target intensity decreases when the surround area increase and vice versa.

4.2 The Benchmark Dataset

The data set used by [1,3,4,5] is same, i.e., TNO Human Factors Search_2 image dataset. These four models operate on the single spatial domain as well as they are parameter-free. The images available in the dataset TNO Human Factors Search_2 are very high resolution (6144 X 4096) pixels. These models sub-sampled this dataset images by a factor of 4 and the resulting images were of (1536 X 1024) pixels. Table 1 exhibits the Spearman's Rank Order Correlation of detection saliency exhibited by SUN [3], SW [4], EDS [5] and MCC [1]. Table 2 is showing the Spearman's Rank Order

ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Journal of Engineering Technology (ISSN: 0747-9964)
Volume 6, Issue 1, Jan, 2017, PP.146-163
ORIGINAL ARTICLE

Correlation of identification saliency exhibited by SUN [3], SW [4], EDS [5] and MCC [1]. Finally, Table 3 demonstrates mean search time of SUN [3], SW [4], EDS [5] and MCC [1].

Table 1: Models and their Detected Saliency

Models	**Detection Saliency**
SUN [3]	0.438
SW [4]	0.217
EDS [5]	0.443
MCC [1]	0.653

Table 2: Spearman's Rank Order Correlation

Models	**Identification Saliency**
SUN [3]	0.702
SW [4]	0.522
EDS [5]	0.608
MCC [1]	0.843

Table 3: Mean search time

Models	**Mean Search Time**
SUN [3]	0.735
SW [4]	0.398
EDS [5]	0.550
MCC [1]	0.735

Correlation of all of these saliency models (except EDS) over the target area are having maximum saliency values in accordance with these three aspects, which include, detection saliency, identification saliency as well as mean search time. When compared with human estimates, MCC Metric has the largest estimate.

In EDS, the mean output values are large which means that the target saliency is low and the maximal target values represent the internal structure of the target. If the maximal salience is small then the target's internal structure is prominent and if it is large, then the target's internal structure is not articulate but it's boundaries are clear. According to the results, the correlation between mean EDS values and human observer data is strong compared to the maximal EDS values. Therefore, the targets which are having less internal structure can be extracted as salient targets, whereas, those targets which are having more internal details are not that much salient since the former is having clear boundaries.

The results are showing that the highest correlations are between identification saliency and maximal salience of MCC Metric (which is 0.843). MCC is an efficient metric and is used to calculate the bottom-up saliency since it uses a simple conspicuity contrast approach. The results also depict that the highest correlation between the calculated maximum saliency and the mean search time is again MCC Metric (which is 0.735).

There are several aspects to imitate human visual system, like 1) distinctness of local image considered statistically (as found in SUN), when 2) contrast is taken into account (as found in SW) or when 3) edginess is under consideration (as found in EDS). Thus, bottom-up saliency can be expressed for these saliency models to find out what local feature differences are there, which may include the differences in color, texture, shape, size, luminance, etc.

5. **Applications**

Visual attention modeling has a number of applications, which include video summarization [15], image and video quality assessment [14], progressive image transmission [16], image segmentation [17], object recognition [18].

5.1 Video Summarization

There is a massive data placed on the internet which includes images and videos that are needing to be retrieved in a well-organized manner. With the technological advancement, the quantity of producing new and new video data is also mounting quickly. There must be some ways and means to browse this video data efficiently. Earlier techniques used to select the key frames randomly or on the interval basis, that is, selecting the frames after a particular interval of time. This problem can be resolved in many ways; one solution can be in the form of providing summaries of the video so that the users may browse through the video quickly. Except browsing, the users can easily reach the portion of the video of their interest.

Video Summarization is the technique in which a video is summarized into a smaller sized and only important and visually salient parts of the video are highlighted. There are two ways to perform video summarization:

5.2 **Video Skimming**

Video summarization is an old concept and a lot of work has been done by researchers in this field. Ma and Zhang [21] performed video skimming with their motion attention-based model. Then Ma et al. [22] extended this work and produced a collection of visual, linguist and auditory features in an open framework. Chernyak and Stark [23] presented a top-down approach as they took visual attention depending upon the observer. Jiang and Qin [24] estimated the visual attention with the use of Visual Attention Index (VAI) then they grouped the extracted frames by K-means algorithm into different clusters and at the end those frames with the highest values of Visual Attention Index (VAI) were selected. Peng and Xiaolin [25] used the color histogram to cluster the frames and after that, they chose such a frame from each cluster which was the most salient. Lai and Yi [26] used a time controlled algorithm to group related frames together which gives the concept of clustering and then that frame was selected as the key frame from each cluster that was visually most prominent and distinct based on these features like, color, texture, and motion.

In video skimming, the original video is cut short into a much smaller video which is shorter in duration compared to the real video. Video skimming produces skims which are more meaningful and enjoyable in contrast to the key frames. Whereas in key frames extraction based technique, some visually salient or distinct frames are extracted from the real video. Keyframes let the user experience

Journal of Engineering Technology (ISSN: 0747-9964)
Volume 6, Issue 1, Jan, 2017, PP.146-163

the whole significant and important parts of the video in just one view exclusive of even viewing a single small clip. Most of the video summarization techniques use low-level index features whereas better approach could have been employing high-level semantic contents which include objects, proceedings, and actions in the video. Some researchers extracted those frames from the videos which were visually distinct from their background with the use of some visual attention modeling techniques [27]. One way to achieve this is extracting the key frames non-linearly to find out such visual attention signs which are static and the others might be dynamic and once they are found then merge these signs. Temporal gradients are used for dynamic attention modeling and that image saliency discovery which is based on image signature is used for static modeling. In image signature based saliency discovery technique, the foreground of the image is approximated. The underlying hypothesis is that the foreground of the image is more informative and prominent as compared to the background of the image.

There are two methods for attention detection, which are,

- dynamic attention detection and
- static attention detection.

According to dynamic attention detection, human's visual attention can normally be attracted by motion contrast [24]

It is a natural fact that the brain and Human Visual System coordinate with each other to identify the visually salient portions from the images and videos. Human beings are able of focusing on particular areas of the images and videos by keenly observing them which is a neurobiological process known as human attention. Some precise mechanism for human attention needs to be explored, as no one has yet explored it. Human attention is expressed by two mechanisms, i.e., the first one is bottom-up attention mechanism and the other one is known as top-down attention mechanism. The top-down attention is derived from high-level features like, objects, actions, and events which attract the observer's gaze. Bottom-up attention, in contrast, is derived from low-level features which include color, texture, motion contrast, etc.

Journal of Engineering Technology (ISSN: 0747-9964)
Volume 6, Issue 1, Jan, 2017, PP.146-163

ORIGINAL ARTICLE

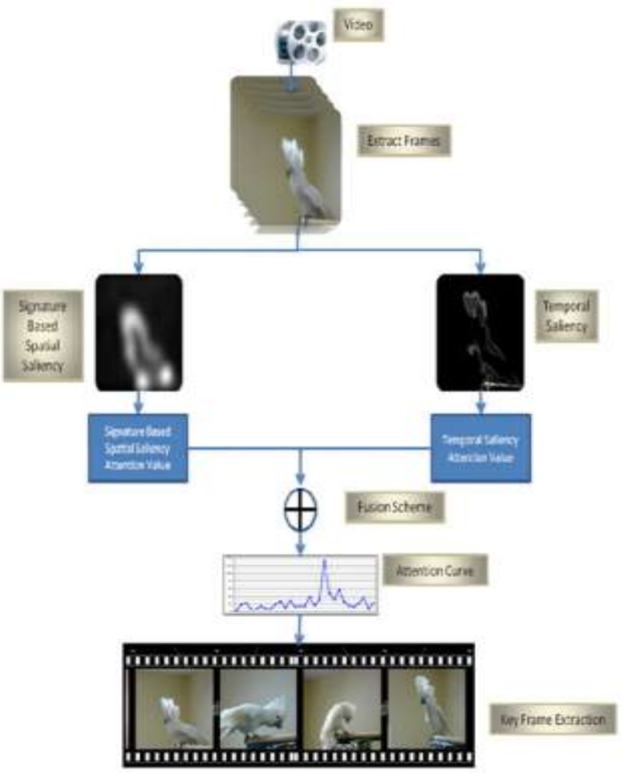


Fig. 6: N.Ejaz et al. Architecture for key frame extraction [27]

In the above architecture, presented by Ejaz et al. [27], first of all, frames are figured out from the video. Then first for dynamic frames, time-based temporal saliency is used, resulting in temporal saliency attention values. Similarly, for static frames, image signature technique is used; resulting in signature based spatial saliency attention values. Then these two found values are fused together and an attention curve is obtained, thus key frames are extracted which are visually salient.

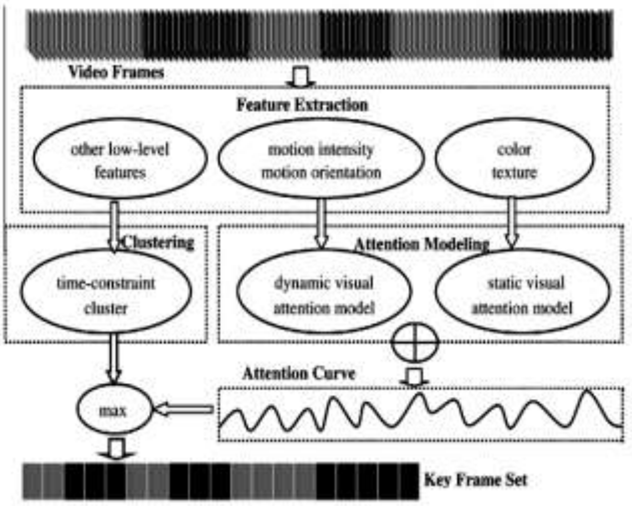


Figure 7: Lai and Yi et al. [26] Architecture for key frame detection

Journal of Engineering Technology (ISSN: 0747-9964)
Volume 6, Issue 1, Jan, 2017, PP.146-163

This architecture for key frames extraction was presented by Lai and Yang [26] and is very much similar to Ejaz et al. [27]. In this architecture, first of all, frames are extracted from the video based on their features, as mentioned here like color, texture, motion intensity, motion orientation, and other low-level features. Then they have used color and texture features to create the static visual attention model, whereas motion intensity and motion orientation are used for dynamic visual attention modeling. These two attention models (dynamic and static) are then merged together through some fusion technique to achieve an attention curve. On other hand, frames with similar contents are clustered to create time-constraint clusters [26]. Now the attention curve and the time constraint cluster are used in a function named max which is used to get the final set of visually distinct key frames.

Fig. 8: Extracted key frames by different techniques [27]

Figure 8 is showing the extracted key frames by different techniques for a video. This image is taken from [27] just to list what type of key frames are extracted using different techniques.

5.3 Image and Video Quality Assessment

In those applications in which images and videos are processed, their quality evaluation and assessment is one of the most significant concerns. There are a lot of indices that are used to cater human visual sensitivity which includes the followings [14]:

- Structural Similarity (SSIM),
- Visual Information Fidelity (VIF), etc.,

For Structural Similarity (SSIM), when humans are observing a scene, in a video or an image, luminance in the scene is normally observed. Luminance is basically the product of illumination that falls over the objects and their reflectance. So, the structural information of the object absolutely does not depend on the intensity of luminance and that of contrast. To extract that information, the structure is separated from the influence of the intensity of luminance and then the SSIM Index is calculated [14]. Structural Similarity (SSIM) Index is used for quality assessment. Applying Structural Similarity (SSIM) Index locally rather than globally is a better approach. It has a number of

Journal of Engineering Technology (ISSN: 0747-9964)
Volume 6, Issue 1, Jan, 2017, PP.146-163
ORIGINAL ARTICLE

benefits. First, the statistic features, such as the luminance and the contrast, of natural images are not spatially stationary and they keep on changing. Second, image distortions may also be spatially different. Third, Human Visual System (HVS) is capable of foveation in which the image resolution varies all over the image as there are normally more than one fixation points, humans perceive only such particular area in an image which is having the highest resolution and is catered by the retina (also known as the fovea). Additional, information about image degradation having local quality measurements which could be helpful in assessing different applications.

For Visual Information Fidelity (VIF), the correctness of the image is measured. Most of the researchers have concentrated on measuring the signal fidelity for the assessment of visual quality. A number of models can be used to determine image and video quality assessment, such as [2,11]. A lot of research has been done to simulate the Human Visual System (HVS) by the researchers [28]. The underlying principle of visual quality assessment is that an object present in the salient region of the image could be more disturbing than it to be in the other non-informative parts. This can be found by recording the eye movements.

Fig. 9 Eye Tracking Apparatus [28]

6. Conclusion

Visual saliency is a non-linear concept and a distinguishing property of an image through which a salient part of the image is highlighted. Bottom-up image saliency is the stimulus-driven and top-down image saliency is the user-driven saliency. Visual attention models are used to figure out the distinctness in the images. These models are used in many real world environments for video summarization, key feature-based visual attention, image and video quality assessment etc.

References

[1] Toet, A. (2011). Computationally versus Psychophysical Bottom-Up Image Saliency: A Comparative Evaluation Study, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33(11), pp. 2131-2146.

[2] Bruce, N.D.B and Tsotsos, J.K. (2009). Saliency Based on Information Maximization, Advances in Neural Information Processing Systems, Y. Weiss, B. Scholkopf, and J. Platt, eds., vol.18(6), pp.155-162.

[3] Kanan, C. et al., (2009) SUN: Top-Down Saliency Using Natural Statistics, Visual Cognition, vol.17(6/7), pp.979-1003.

[4] Bian, P. and Zhang, L. (2009). Biological Plausibility of Spectral Domain Approach for Spatio-temporal Visual Saliency, Advances in Neuro-Information Processing, pp.251-258.

[5] Saba, T. (2016) Pixel intensity based cumulative features for moving object tracking (MOT) in darkness, 3D Research vol. 7 (10), pp.1-6, doi. 10.1007/s13319-016-0089-4.

[6] Walther, D. (2006). Interactions of Visual Attention and Object Recognition: Computational Modeling, Algorithms and Psychophysics , PhD thesis, California Institute of Technology, pp. 79-83.

[7] Walther, D. and Koch, C. (2006) Modeling Attention to Salient Proto-Objects, Neural Networks, vol.19(9), pp.1395-1407.

[8] Avraham, T. and Lindenbaum, M. (2010). Esaliency (Extended Saliency): Meaningful Attention Using Stochastic Image Modeling, IEEE Trans. Pattern Analysis and Machine Intelligence, vol.32(4), pp.693-708.

[9] Mancasetal., M. (2006). A Rarity-Based Visual Attention Map Application to Texture Description, Proc. IEEE Int'l Conf. Image Processing, pp. 445-448.

[10] Joudaki, S., Mohamad, D. Saba, T. Rehman, A. Al-Rodhaan, M. Al-Dhelaan, A. (2014) Vision-Based Sign Language Classification: A Directional Review, IETE Technical Review, vol.31(5), pp. 383-391, doi.10.1080/02564602.2014.961576

[11] Saba, T. Rehman, A. Elarbi-Boudihir, M. (2014). Methods and strategies on off-line cursive touched characters segmentation: a directional review, Artificial Intelligence Review, vol. 42(4), pp. 1047-1066 , doi.10.1007/s10462-011-9271-5

[12] Achanta, R. et al.,(2009) Frequency-Tuned Salient Region Detection, Proc. IEEE International Conference Computer Vision and Pattern Recognition.

[13] Garcia-Diaz, A. et al., (2009) Saliency Based on Decorrelation and Distinctiveness of Local Responses, Proc. 13th Int'l Conf. Computer Analysis of Images and Patterns, pp. 261-268, 2009.

[14] Q. Ma, L. Zhang, and B. Wang,(2010). New Strategy for Image and Video Quality Assessment, J. Electronic Imaging, vol. 19, pp. 1-14.

[15] S. Marat, M. Guironnet, and D. Pellerin, (2007) Video Summarization Using a Visual Attention Model, Proc. 15th European Signal Processing Conf., pp. 1784-1788.

[16] Chun, C. and Aun, C. (2014) Construction of Business Process Reengineering for Professional Management Journal of Business and Technovation, vol. 3(3), pp. 149-164.

[17] B. C. Ko and J.Y. Nam,(2006) Object-of-Interest Image Segmentation Based on Human Attention and Semantic Region Clustering, J. Optical Soc. of Am. A, vol. 23(10), pp. 2462-2470.

[18] Walther, D. et al., (2005). Selective Visual Attention Enables Learning and Recognition of Multiple Objects in Cluttered Scenes, Computer Vision and Image Understanding, vol. 100(1/2), pp. 41-63.

[19] N.D.B. Bruce, (2005). Features That Draw Visual Attention: An Information Theoretic Perspective, Neuro computing, vol. 65/66 pp. 125-133.

[20] H. J. Seo and P. Milanfar, (2009). Static and Space-Time Visual Saliency Detection by Self-Resemblance, Journal of Vision, vol. 9(12), pp.1-27.

[21] Y. F. Ma, H. J. Zhang, (2002). A model of motion attention for video skimming , in Proc. of International Conference on Image Processing, vol. 1, pp. 129-130.

[22] Y. F. Ma, X. S. Hua, L. Lu, H. J. Zhang, (2005). A generic framework of user attention model and its application in video summarization, IEEE Transactions on Multimedia, vol. 7(5), pp. 907–919.

[23] D. A. Chernyak, L. W. Stark, (2001) Top-down guided eye movements , IEEE Transaction On Systems, Man and Cybernetics, vol. 31(4) pp. 514–522.
[24] Jiang, P., Qin, X.L.(2010) Key frame based video summarization using visual attention clue , IEEE Multimedia, vol. 17(2), pp.64–73.
[25] Peng, J. Xiaolin, Q. (2010) Key frame based video summary using visual attention clues , IEEE Transactions on Multimedia, vol. 17(2), pp. 64–73.
[26] Lai, L.J., Yi, Y. (2012). Key frame extraction based on visual attention model , Journal of Visual Communication and Image Representation, vol. 23(1), pp. 114–125.
[27] Norouzi, A. Rahim, MSM, Altameem, A. Saba, T. Rada, A.E. Rehman, A. Uddin, M. (2014) Medical image segmentation methods, algorithms, and applications IETE Technical Review, vol.31(3), doi.10.1080/02564602.2014.906861.
[28] lsayed, H.A.G., Alharbi, A.N. AlNamlah, H. Saba, T. (2015) Role of Agile Methodology in Project Management and Leading Management Tools, Journal of Business and Technovation, vol.3(3),pp. 188-193
[29] Loe, Q. and Liu, P. (2015). Factors Effecting Enterprise Resource Planning, Journal of Business and Technovation, vol. 3(2), pp. 64-71
[30] Khodamoradi, M. Bozorgmanesh, M. and Ghorbani, E.Usage of Data and Correspondence Advancements (ICT) In Instruction (2014), Journal of Business and Technovation, vol.2(2), pp.131-135.
[31] http://www.vision.ime.usp.br/~rafaellg/2012/07/07/simple-bottom-up-saliency-model-in-python-opencv/
[32] Rehman, A. and Saba, T. (2014). Evaluation of artificial intelligent techniques to secure information in enterprises, Artificial Intelligence Review, vol. 42(4), pp. 1029-1044, doi. 10.1007/s10462-012-9372-9.
[33] Saba, T. and Altameem, A. (2013) Analysis of vision based systems to detect real time goal events in soccer videos, Applied Artificial Intelligence, vol. 27(7), pp. 656-667, doi. 10.1080/08839514.2013.787779
[34] Rehman, A. and Saba, T. (2012). Off-line cursive script recognition: current advances, comparisons and remaining problems, Artificial Intelligence Review, vol. 37(4), pp.261-268. doi. 10.1007/s10462-011-9229-7.
[35] Fadhil, MS. Alkawaz, MH., Rehman, A., Saba, T. (2016) Writers identification based on multiple windows features mining, 3D Research, vol. 7 (1), pp. 1-6, doi.10.1007/s13319-016-0087-6
[36] Rehman, A. and Saba, T. (2011). Document skew estimation and correction: analysis of techniques, common problems and possible solutions, Applied Artificial Intelligence, vol. 25(9), pp. 769-787. doi. 10.1080/08839514.2011.607009.
[37] Rehman, A. and Saba, T. (2011). Document skew estimation and correction: analysis of techniques, common problems and possible solutions Applied Artificial Intelligence, vol. 25(9), pp. 769-787. doi. 10.1080/08839514.2011.607009
[38] Saba, T., Rehman, A., and Sulong, G. (2011) Improved statistical features for cursive character recognition International Journal of Innovative Computing, Information and Control (IJICIC) vol. 7(9), pp. 5211-5224

[39] Muhsin; Z.F. Rehman, A.; Altameem, A.; Saba, A.; Uddin, M. (2014). Improved quadtree image segmentation approach to region information. the imaging science journal, vol. 62(1), pp. 56-62, doi. http://dx.doi.org/10.1179/1743131X13Y.0000000063.
[40] Neamah, K. Mohamad, D. Saba, T. Rehman, A. (2014). Discriminative features mining for offline handwritten signature verification, 3D Research vol. 5(3), doi. 10.1007/s13319-013-0002-3
[41] Mundher, M. Muhamad, D. Rehman, A. Saba, T. Kausar, F. (2014) Digital watermarking for images security using discrete slantlet transform, Applied Mathematics and Information Sciences, vol 8(6), pp. 2823-2830, doi.10.12785/amis/080618.
[42] Rehman, A. Mohammad, D. Sulong, G. Saba, T.(2009). Simple and effective techniques for core-region detection and slant correction in offline script recognition Proceedings of IEEE International Conference on Signal and Image Processing Applications, pp. 15-20.
[43] Rehman, A. Kurniawan, F. Saba, T. (2011) An automatic approach for line detection and removal without smash-up characters, The Imaging Science Journal, vol. 59(3), pp. 177-182, doi. 10.1179/136821910X12863758415649
[44] Saba, T. Rehman, A. Sulong, G. (2011) Cursive script segmentation with neural confidence, International Journal of Innovative Computing and Information Control (IJICIC), vol. 7(7), pp. 1-10.
[45] Rehman, A. Alqahtani, S. Altameem, A. Saba, T. (2014) Virtual machine security challenges: case studies, International Journal of Machine Learning and Cybernetics vol. 5(5), pp. 729-742, doi. 10.1007/s13042-013-0166-4.
[46] Joudaki, S. Mohamad, D. Saba, T. Rehman, A. Al-Rodhaan, M. Al-Dhelaan, A. (2014) Vision-based sign language classification: a directional Review, IETE Technical Review, vol.31 (5), 383-391, doi. 10.1080/02564602.2014.961576
[47] Saba, T. Rehman, A. Altameem, A. Uddin, M. (2014) Annotated comparisons of proposed preprocessing techniques for script recognition, Neural Computing and Applications, vol. 25(6), pp. 1337-1347 , doi. 10.1007/s00521-014-1618-9
[48] JWJ Lung, MSH Salam, A Rehman, MSM Rahim, T Saba (2014) Fuzzy phoneme classification using multi-speaker vocal tract length normalization, IETE Technical Review, vol. 31 (2), pp. 128-136, doi. 10.1080/02564602.2014.892669
[49] Saba, T., Rehman, A., Sulong, G. (2010). An intelligent approach to image denoising, Journal of Theoretical and Applied Information Technology, vol. 17 (2), 32-36.
[50] Saba, T., Almazyad, A.S. Rehman, A. (2016) Online versus offline Arabic script classification, Neural Computing and Applications,vol.27(7), pp 1797–1804, doi. 10.1007/s00521-015-2001-1.
[51] Soleimanizadeh, S., Mohamad, D., Saba, T., Rehman, A. (2015) Recognition of partially occluded objects based on the three different color spaces (RGB, YCbCr, HSV) 3D Research, vol. 6 (22), 1-10, doi. 10.1007/s13319-015-0052-9.
[52] Saba, T. Rehman, A. Al-Zahrani, S. (2013) Character segmentation in overlapped script using benchmark database, pp. 140-143, ISBN: 978-1-61804-233-0
[53] Saba, T. and Rehman, A. (2012). Effects of artificially intelligent tools on pattern recognition, International Journal of Machine Learning and Cybernetics, vol. 4, pp. 155-162. doi. 10.1007/s13042-012-0082-z.

[54] Z.S.Younus, D. Mohamad, T. Saba, M. H.Alkawaz, A. Rehman, M.Al-Rodhaan, A. Al-Dhelaan (2015) Content-based image retrieval using PSO and k-means clustering algorithm, Arabian Journal of Geosciences, vol. 8(8) , pp. 6211-6224, doi. 10.1007/s12517-014-1584-7

[55] Al-Ameen, Z. Sulong, G. Rehman, A., Al-Dhelaan, A., Saba, T. Al-Rodhaan, M. (2015) An innovative technique for contrast enhancement of computed tomography images using normalized gamma-corrected contrast-limited adaptive histogram equalization, EURASIP Journal on Advances in Signal Processing, vol. 32, doi:10.1186/s13634-015-0214-1

[56] Saba,T., Rehman, A., Al-Dhelaan, A., Al-Rodhaan, M. (2014) Evaluation of current documents image denoising techniques: a comparative study, Applied Artificial Intelligence, vol.28 (9), pp. 879-887, doi. 10.1080/08839514.2014.954344

[57] Nodehi, A. Sulong, G. Al-Rodhaan, M. Al-Dhelaan, A., Rehman, A. Saba, T. (2014) Intelligent fuzzy approach for fast fractal image compression, EURASIP Journal on Advances in Signal Processing, doi. 10.1186/1687-6180-2014-112.

[58] Ahmad, AM., Sulong, G., Rehman, A., Alkawaz,MH., Saba, T. (2014) Data Hiding Based on Improved Exploiting Modification Direction Method and Huffman Coding, Journal of Intelligent Systems, vol. 23 (4), pp. 451-459, doi. 10.1515/jisys-2014-0007

[59] Harouni, M., Rahim,MSM., Al-Rodhaan, M., Saba, T., Rehman, A., Al-Dhelaan, A. (2014) Online Persian/Arabic script classification without contextual information, The Imaging Science Journal, vol. 62(8), pp. 437-448, doi. 10.1179/1743131X14Y.0000000083

[60] Jadooki, S. Mohamad,D., Saba, T., Almazyad, A.S. Rehman, A. (2016) Fused features mining for depth-based hand gesture recognition to classify blind human communication, Neural Computing and Applications, pp. 1-10, doi. 10.1007/s00521-016-2244-5.

[61] Norouzi, A. Rahim, MSM, Altameem, A. Saba, T. Rada, A.E. Rehman, A. & Uddin, M. (2014) Medical image segmentation methods, algorithms, and applications IETE Technical Review, vol.31(3), doi.10.1080/02564602.2014.906861.

[62] Rehman, A. and Saba, T. (2014). Neural network for document image pre-processing, Artificial Intelligence Review, vol. 42(2), pp 253-273, doi. 10.1007/s10462-012-9337-z.

[63] Elarbi-Boudihir, M. Rehman, A. and Saba, T. (2011). Video motion perception using optimized Gabor filter, International Journal of Physical Sciences, vol. 6(12), pp. 2799-2806.

[64] Saba, T. and Altameem, A. (2013) Analysis of vision based systems to detect real time goal events in soccer videos, Applied Artificial Intelligence, vol.27(7), pp. 656-667, doi. 10.1080/08839514.2013.787779.

[65] Haron, H. Rehman, A. Wulandhari, L.A. Saba, T. (2011). Improved vertex chain code based mapping algorithm for curve length estimation, Journal of Computer Science, Vol. 7(5), pp. 736-743, doi. 10.3844/jcssp.2011.736.743.

[66] Selamat, A. Phetchanchai, C. Saba, T. Rehman, A. (2010). Index financial time series based on zigzag-perceptually important points, Journal of Computer Science, vol. 6(12), 1389-1395, doi. 10.3844/jcssp.2010.1389.1395

[67] Haron, H. Rehman, A. Adi, DIS, Lim, S.P. and Saba, T.(2012). Parameterization method on B-Spline curve, Mathematical problems in engineering, vol. 2012, doi. 10.1155/2012/640472.

[68] Al-Turkistani, H. and Saba, T. (2015) Collective Intelligence for Digital Marketing, Journal of Business and Technovation, vol.3(3), pp: 194-203

[69] Ming, J.,Ping, J. and Chiun, R. (2014) Provably Secure Password-based Threeparty Key Exchange Protocol with Computation Efficiency, Journal of Business and Technovation, vol.2(2), pp-117-126

[70] Arafat, S. and Saba, T. (2016) Social Media Marketing Vs Social Media Commerce: A Study on Social Media's Effectiveness on Increasing Businesses' Sales, Journal of Business and Technovation, vol. 4(3), pp. 134-147.

[71] Abbasi, H. (2015) Role of Management Accounting Information System in Organizations, Journal of Business and Technovation, vol.3(3), pp. 165-171.

[72] Amin, M.S. and Noori, A.(2016). Mechanism For Farm Mechanization and Careful Planning Using Geographic Information System (GIS), Journal of Business and Technovation, vol.4(1), pp. 23-28.

[73] Kem, Y. (2016). Use of Information Technology in Urban Planning and Design: An Urban Architecture Study, Journal of Business and Technovation, vol. 4(1), pp. 1-9

[74] Jun,S. Mii, S. and Kao, C. New CMT-SCTP With Increased Speed of Data Transmission Using Fuzzy Logic, Journal of Business and Technovation , vol. 4(2), 2016, pp. 66-74

[75] Arafat, S. and Saba, T. (2016) Social Media Marketing Vs Social Media Commerce: A Study on Social Media's Effectiveness on Increasing Businesses' Sales, Journal of Business and Technovation, vol. 4(3), pp. 134-147

[76] Patwa, N. Kavuri, S.P. Grandhi, B. Hongrak, K. (2016) Study on Effectiveness of Location-Based Advertising on Food Service Industry in Sydney, Journal of Business and Technovation, vol. 4(3), pp. 112-124.

[77] Niranjan, I., Sahu, V.K. Saravanan, A.S. (2016) A Study on Performance Management of Credit Rating Agencies, Journal of Business and Technovation, vol. 4(3), pp. 96-111.

[78] Aljalidi, N. Alshedokhi, M. and Saba, T. (2016) The Impact of TripAdvisor Ratings for Tourisim Advertisement in Saudi Arabia, Journal of Business and Technovation, vol. 4(2), pp. 90-95.

[79] Sara, J. and Piyumi, A. (2016). The Study of Organizational Anti-citizenship Behaviors on the Performance of Organizations (Case Study: Sri-Lanka Airlines Company), Journal of Business and Technovation, vol. 4(1), 2016, pp. 10-22

[80] Ahmad, H. and Rasheed, M. (2015) Surveying The Influence of Customer Relationship Management on Gaining Competitive Advantage, Journal of Business and Technovation, vol. 3(2), pp. 87-94.

[81] Syeed, H. and Ajaz, N. (2015) Surveying The Influence of Market Orientation on Competitive Strength Development, Journal of Business and Technovation, vol.3(2), pp. 95-102

[82] Mehrmanesh, H., Eyni, H., and Sargheyn, A. (2015) Customer Relationship Management: A Case Study of Iran Mellat Bank, Journal of Business and Technovation, vol.3(2), pp. 103-108.

[83] Rasool, M. and Khan, K. (2015) Determining the Concept of Ethics, Professional Ethics and Management Ethics in the View of Islam, Journal of Business and Technovation, vol. 3(1), 2015, pp.26-46

[84] Zendehdel, M. and Paim, L.H (2014) Online Sales and Purchase of Products: Security and Privacy Issues, Journal of Business and Technovation, vol. 2(2), 2014, pp.142-146.

[85] Saba, T. and Rehman, A. (2012). Machine Learning and Script Recognition, Lambert Academic publisher, pp.35-40.

[86] Saba T, Al-Zahrani S, Rehman A. (2012) Expert system for offline clinical guidelines and treatment Life Sci Journal 2012;vol. 9(4):2639-2658.

[87] Saba, T. Rehman, A. Al-Dhelaand, A. and Al-Rodhaand, M. (2014). Evaluation of current documents image denoising techniques: a comparative study, vol. 28(9), pp. 879-887, doi. 10.1080/08839514.2014.954344

[88] Selamat, A. Phetchanchai, C. Saba, T. Rehman, A. (2010). Index financial time series based on zigzag-perceptually important points, Journal of Computer Science, vol. 6(12), 1389-1395, doi. 10.3844/jcssp.2010.1389.1395

[89] Saba, T. and Alqahtani, F. (2013) Semantic analysis based forms information retrieval and classification, 3 D Research, vol. 4(4), doi. 10.1007/3DRes.03(2013)4