WHAT DO YOU MEAN BY "SALIENCE"?

V. Javier Traver
Dep. Computer Languages and Systems, Institute of New Imaging Technologies
Universitat Jaume I, Castellón, Spain

Keywords: Visual salience, Visual attention, Interest-point detection, Image segmentation.

Abstract: In the field of computer vision, the term "salience" is being used with different or ambiguous meanings in a variety of contexts. This abuse of terminology contributes to create some confusion or misunderstanding among practitioners in computer vision, a situation which is particularly inconvenient to less experienced researchers in the field. The contribution of this paper is twofold. On the one hand, by providing a categorization of some different usages of the concept of salience, its possible meanings will be clarified. On the other hand, by providing a common framework to understand and conceptualize those different meanings, the commonalities emerge and these analogies might serve to relate the ideas and techniques across usages.

1 INTRODUCTION

In everyday language, something is said to be salient if it stands out, attracts attention, or is of notable significance[1]. But, what about the use of the term "salience" in literature of computer vision or image processing? Do all the usages in these fields refer to the same meaning of salience? Actually, it can be found that different authors use the term to denote something similar, but not always the same, or they may even be referring to quite different concepts, but this distinction is, at best, implicit, and often rather unclear. This situation is certainly not desirable, since readership, particularly newcomers to the field, may be confused by this lack of consistency.

This position paper is aimed at gaining some insight and perspective, so that confusion is reduced or removed among researchers, *by* those employing the terminology *for* those reading and learning about it. It is important to note that the paper is *not* intended to be a review work about the salience problem. Firstly, some representative works that use the term salience are mentioned and classified into a few categories (§ 2). Then, a tentative framework is proposed as an aid in thinking about current and potential research on salience-based problems (§ 3). Some remarks are finally provided (§ 4).

[1]Merriam Webster's Online Dictionary, http://www.merriam-webster.com.

2 CATEGORIZATION OF SALIENCE

Among the works about salience, we identified four broad meanings for it: perceptual, uniqueness, likelihood, and semantics-based salience, which are presented in the following paragraphs. This categorization is not claimed to be unique or exhaustive, but rather, illustrative of the significantly different uses of the term and that should not be mistaken.

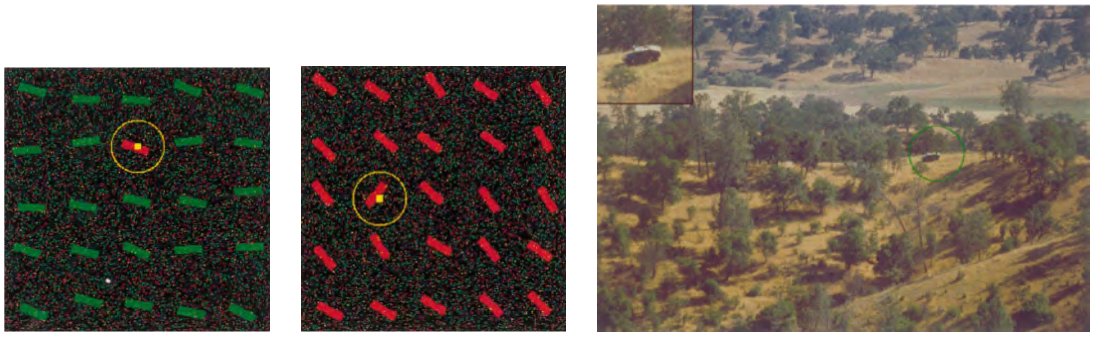
2.1 Perceptual Salience

The perceptual meaning of salience is probably the most widely known and, therefore, what most researchers understand by the term. Computation of visual salience tend to get a lot of inspiration in biological visual system, and how animals rely on salient cues to act in their environment.

The work by Itti and Koch (Itti et al., 1998) is a good representative example of this category. Fig. 1 illustrates some results of their original model. Significant research has been developed on visual attention by Itti and collaborators (iLab, 2000). Another more recent example is the work by (Meur et al., 2006), which models additional features of the bottom-up visual attention of the human visual system.

Visual salience is just one element of the very broad discipline of visual attention. There are many issues in the literature of visual attention, such as

Javier Traver V..
WHAT DO YOU MEAN BY "SALIENCE"?.
DOI: 10.5220/0003356803400345
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2011), pages 340-345
ISBN: 978-989-8425-47-8



(a) (b) (c)

Figure 1: The model by (Itti and Koch, 2000) allows to predict the visual prominence of the bar with (a) different color and (b) different orientation Its performance in a natural scene is shown in (c), where a car within a forest is signaled as salient. [Source: Reprinted from Vision Research, Vol. 40, numbers 10–12, L. Itti and C. Koch, A saliency-based search mechanism for overt and covert shifts of visual attention, Pages 1489–1506 , Copyright (2000), with permission from Elsevier.] (Note: this and Figures 3 and 4 in this paper should be viewed in color).

the bottom-up *vs.* top-down attention, or overt *vs.* covert attention (Sun and Fisher, 2003). These aspects, while important, are not actually relevant to the purpose of this paper and, therefore, are not considered here. However, some discussion related to object-based and space-based visual attention is included below (§ 2.4, § 3).

2.2 Uniqueness

Another use of salience is as a synonym of "information of great importance or interest", where the semantics of importance or interest depends on the target application, but this is only rarely or vaguely explained. Usually, in these problems, the sought points or regions should be highly discriminative and repeatable when viewing or photometric conditions change. Detecting these image locations is therefore helpful in image matching and object recognition. Under this meaning, the salience of a point, a feature or a region may refer to the extent to which they facilitate the matching or recognition tasks.

Works of *key* points or *interest* point detectors (Schmid et al., 2000; Mikolajczyk and Schmid, 2004), as well as transformation-invariant features (Lowe, 2004) would be included in this category. While most of these examples do not explicitly use the term salience, others do (Kadir and Brady, 2001), which adds a lot of confusion for two reasons. On the one hand, from the point of view of key points, they are using *different* terminology to mean the *same* thing. On the other hand, from the point of view of salience (in the *perceptual* meaning), they use the *same* terminology to mean *different* things. As a result, some readers may infer the perceptual meaning when the authors meant the uniqueness meaning, and other readers interested in the uniqueness meaning may skip works whose title or abstract may wrongly imply the perceptual meaning.

Examples of regions or points detected within images using these kind of approaches are shown in Fig. 2. These detectors can be very powerful for matching or recognizing objects or scenes even under affine or perspective distortions. However, they may (and usually do) detect *many* points within a same image, without much or any concern on their local or global distribution. As a result, they do not straightforwardly lend themselves to detect salience in the perceptual sense.

Another set of approaches that would lie in this category are those based on traditional basic features, such as edges, but in these cases they are aimed at finding the "salient" ones, mostly assuming that "there are significant edges that characterize the objects in images" (Han and Guo, 2003).

2.3 Probability

A third usage of salience is simply a means to denote how likely a given pixel or image region is of containing or representing some information which is relevant to the task at hand. Works with this "probability" meaning would include those of object categorization (Moosmann et al., 2006) or those aimed at detecting irregular, unlikely or suspicious contents (Boiman and Irani, 2007).

With respect to the first example (Moosmann et al., 2006), object categories can be inferred by appropriate learning algorithms, some of whose results are illustrated in Fig. 3.

On the other hand, the definition of salience by Itti et al. (who use the perceptual meaning, § 2.1) is criti-

⟨ Please, refer to Fig. 14(b) in http://dx.doi.org/10.1023/B:VISI.0000027790.02288.f2 ⟩

(a) scale and affine interest point detector (Mikolajczyk and Schmid, 2004)

⟨ Please, refer to Fig. 5 (bottom) in http://dx.doi.org/10.1023/A:1012460413855 ⟩

(b) multi-scale saliency (Kadir and Brady, 2001)

Figure 2: Some results of two different key-points detectors. [We regret that some original figures from other papers could not be included in this paper since permission for reproducing them were not available at the time to submit the camera-ready paper. As a replacement, we provide a clickable link for an easy access to an electronic version of the paper to see corresponding figures].

⟨ Please, refer to Fig. 3 in http://eprints.pascal-network.org/archive/00002435/01/MLJ06.pdf ⟩

Figure 3: Salience maps understood as likelihood of object presence (Moosmann et al., 2006).

cized in (Boiman and Irani, 2007) since, according to these authors, distinctiveness within a neighborhood is not always the proper measure for salience, because many local saliences within an image do not yield perceptual conspicuity. Therefore, they propose to take a more global perspective:

> "An image region will be detected as salient if it cannot be explained by anything similar in other portions of the image. Similarly, given a single video sequence (with no prior knowledge of what is a normal behavior), we can detect salient behaviors as behaviors which cannot be supported by any other dynamic phenomena occurring at the same time in the video." (Boiman and Irani, 2007)

However, their algorithm does not seem either to be necessarily consistent with human perception, not only because several (many) regions can be detected as salient, but also because it is difficult to imagine that humans can be effective in quickly spotting the irregularities resulting by their algorithm. For instance, by looking at the image of the cards on the cluttered background on the left of Fig. 4a, no image region stands out to immediately capture the visual attention of a human observer. The card which is different to the rest four cards and to the background may become salient *only after a while*, in particular if the human subject wants (or is asked) to find out what is different. And the performance of the algorithm is really good in this sense, as it is in contexts such as detecting some defects in fruits (Fig. 4b), or identifying abnormal behaviors in video sequences (Fig. 4c). Therefore, the main attractiveness of this approach may lie in its power and generality, not in being faithful to the human visual system (HVS). And given that the latter (faithfulness to the HVS) seems not to be the authors' goal, their criticism to the perceptual meaning of salience as used in (Itti and Koch, 2000) seems inappropriate, mostly because of the confusing effects on the readership: in our view, they are wrongly merging two different meanings of salience (the probability and the perceptual meanings) into one.

2.4 Semantics

Finally, a few other authors may refer to salient regions to mean regions that may correspond to semantically meaningful real-world objects, found after some segmentation or clustering process (Pauwels and Frederix, 1999).

This last meaning may be somehow related to object-based visual attention (Sun et al., 2008) which, in contrast to spatial-based visual attention, hypothesizes that it is the real-world objects, rather than arbitrary spatial locations, that draws our visual attention. However, results, as shown in Fig. 5, do not directly identify which of the detected regions are salient, or how salient each of them is.

2.5 Relationship between Meanings

Some of the different meanings that salience has been given in the computer vision literature have been illustrated above. Despite the differences in those meanings, it can be seen that some of them may overlap slightly and, more importantly, the different perspectives may enrich one another, and some definitions of salience may be inspiring for solving other problems, perhaps by suggesting new fresh approaches to solve old problems.

For instance, the uniqueness/distinctiveness approaches could be reworked to yield perceptually

⟨ Please, refer to Fig. 7(a) and Fig. 7(c) in http://dx.doi.org/10.1007/s11263-006-0009-9 ⟩

(a) Detecting a "significant" region roughly corresponding to a different card

⟨ Please, refer to Fig. 10(b) in http://dx.doi.org/10.1007/s11263-006-0009-9 ⟩

(b) Finding defects in fruits

⟨ Please, refer to Fig. 8(b) in http://dx.doi.org/10.1007/s11263-006-0009-9 ⟩

(c) Inferring abnormal dynamics in video contents

Figure 4: Some results and applications of detecting salience as understood by (Boiman and Irani, 2007).

Figure 5: Salience by segmentation (Pauwels and Frederix, 1999) may yield real-world meaningful regions [Source: Reprinted from Computer Vision and Image Unerstanding, Vol. 75, numbers 1–2, E. J. Pauwels and G. Frederix, Nonparametric Clustering for Image Segmentation and Grouping, Pages 73–85 , Copyright (1999), with permission from Elsevier].

salient images. Or, the other way around, perceptual-based salience models might be revised for them to function as key point/region detectors and descriptors. As another example, the semantics-based segmentation may inspire an algorithm that assigns saliency to the detected regions on the basis of their features (shape, size, compactness, etc.), possibly by including these features in a general framework such as that by (Itti et al., 1998). Yet another possibility is to combine ideas to get the best of all of them. An example of this idea would be the work by (Ko and Nam, 2006), who exploit the complementary role of salient regions (Itti's model) and interest points (corner-like) for object-of-interest segmentation.

3 A COMMON FRAMEWORK

To help discover similarities and differences among the several approaches or meanings of salience, as well as to guide the finding of new research problems, it would be nice to have a common conceptual framework that can be as complete and precise as possible. An attempt for such formalization, which relies on the dimensions that arise regarding the *what*, *where*, *when*, *why* and *how* questions, is presented here:

- **What** is being detected as salient?
- **Where** the detected entity "lives in" that makes it salient?
- **When** does the detected entity becomes salient?
- **Why** has the entity be signaled as salient?
- **How** has been possible to find the salient entities?

Using this framework, a categorization of existing and potential problems, approaches, or applications can be done. For instance, the traditional perceptual meaning of salience would basically be given by answering "pixels" to the *what* question and "single image" to the *where* question. On the other hand, object-based and space-based visual attention could be differentiated in the *what* question, since object-based visual attention tries to find *proto*-objects (Sun and Fisher, 2003; Walther and Koch, 2006) rather than isolated meaningless spatial regions.

The *where* question is related to the context, which is a key concept when defining salience, since it can be studied *within* an image, *between* images, etc. The *where* question is also related to what the "neighborhood" of the salient entity is like that makes it salient.

For instance, in the perceptual meaning of salience, the context is the part of the rest of the image that has features different to those of the salient image region. But in other meanings of salience, such as the probability-based, the context does not necessarily play such a role: bikes can be found next to each other even if they do not visually protrude from the background.

The *when* question is relevant when the temporal variable is included in the problem domain. This is the case, for instance, of video sequences. However, if a sequence is considered a 3D volume, then the *when* would become the *where*. Therefore, depending on its particular statement, a given problem may require different questions or different answers to these questions.

The *why* question has to do with the interpretation of the results. As far as we know, this issue has not been addressed up to now or, at least, has not received significant research attention. While human observers might be able to report why something has been detected as salient, it would be desirable to work towards the automation of the explanation process. Studies on how the task being undertaken might influence the attention (Navalpakkam and Itti, 2005) are partially related to this issue. However, generally speaking, the *why* question is an open research issue.

The *how* question provides the opportunity to distinguish salience-based solutions depending on the models, approaches, methodologies or even technologies being used. For instance, a biologically-motivated visual attention model might be used in one case, and a mathematically-based sound technique in another.

Besides being useful as a taxonomy tool, this framework can be helpful in finding new problems, in stating existing problems in new ways, etc., even for problems outside the *visual* domain. For example, feature selection is a traditional problem in Pattern Recognition, but it can also be seen as a salience problem, since discriminative features form patterns within a same class which are different from those belonging to other classes. Thus, answers to the *where* question could be, in these cases, the feature space for feature selection, or the set of classes in classification.

4 CONCLUSIONS

The main goal of this position paper was to make readers aware that the term salience is used with different meanings in the computer vision literature. This lack of agreement in the terminology may affect practitioners, specially new ones, who may find difficult or confusing some usages of the salience concept or, even worse, be confused *without* being aware they are. To explore the implications of this situation, possible meanings of salience have been grouped into four categories, and illustrated with a few representative examples drawn from the literature.

However, besides the differences across these several usages, some commonalities may also be identified. Because both, differences and similarities, can be useful to find new problems, or new approaches to solve old problems, a simple formal framework has been suggested as a tool to (i) help use a consistent vocabulary; (ii) avoid ambiguities in meanings; and (iii) get inspiration to reuse concepts and ideas, even in the case of problems not related to visual attention but that could be conceptualized similarly.

It is our hope that this paper has helped to clarify ideas on salience, motivated authors to use the term unambiguously, and suggested scientists new research avenues by exploring and exploiting similarities and differences across the several usages of salience.

ACKNOWLEDGEMENTS

The authors acknowledge the funding from the Spanish research programme Consolider Ingenio-2010 CSD2007-00018, and from Fundació Caixa-Castelló Bancaixa under project P1·1A2010-11.

REFERENCES

Boiman, O. and Irani, M. (2007). Detecting irregularities in images and in video. *Intl. J. of Computer Vision*, 74(1):17–31.

Han, J. W. and Guo, L. (2003). A shape-based image retrieval method using salient edges. *Signal Processing: Image Communication*, 18(2):141–156.

iLab (2000). iLab, University of Sourthern California. http://ilab.usc.edu.

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10–12):1489–1506.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE T-PAMI*, 20(11):1254–1259.

Kadir, T. and Brady, M. (2001). Saliency, scale and image description. *Intl. J. of Computer Vision*, 45(2):83–105.

Ko, B. C. and Nam, J.-Y. (2006). Automatic object-of-interest segmentation from natural images. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 45–48, Washington, DC, USA. IEEE Computer Society.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Intl. J. of Computer Vision*, 60:91–110.

Meur, O. L., Callet, P. L., Barba, D., and Thoreau, D. (2006). A coherent computational approach to model bottom-up visual attention. *IEEE T-PAMI*, 28(5):802–817.

Mikolajczyk, K. and Schmid, C. (2004). Scale & affine invariant interest point detectors. *Intl. J. of Computer Vision*, 60(1):63–86.

Moosmann, F., Larlus, D., and Jurie, F. (2006). Learning saliency maps for object categorization. In *ECCV International Workshop on The Representation and Use of Prior Knowledge in Vision*.

Navalpakkam, V. and Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45(2):205–231.

Pauwels, E. J. and Frederix, G. (1999). Finding salient regions in images: nonparametric clustering for image segmentation and grouping. *Computer Vision and Image Understanding*, 75(1/2):73–85.

Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *Intl. J. of Computer Vision*, 37(2):151–172.

Sun, Y. and Fisher, R. (2003). Object-based visual attention for computer vision. *Artificial Intelligence*, 146:77–123.

Sun, Y., Fisher, R. B., Wang, F., and Gomes, H. M. (2008). A computer vision model for visual-object-based attention and eye movements. *Computer Vision and Image Understanding*, 112(2):126–142.

Walther, D. and Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19:1395–1407.