# SAMI: SAliency based Metrics of Identification for object concealment evaluation

Julien Gosseaume Kidiyo Kpalma Joseph Ronsin
Institut d'Electronique et de Telecommunications de Rennes
Universite Europeenne de Bretagne, France
INSA, IETR, UMR 6164, F-35708 RENNES
{julien.gosseaume,kidiyo.kpalma,joseph.ronsin}@insa-rennes.fr

## ABSTRACT

We propose original metrics for estimation of detection and identification of an object in an image. SAMI (SAliency based Metrics of Identification) gives a *detection* score, called *D_score*, and an *identification* score, called *I_score*, for a Region Of Interest (ROI), basically the footprint area of the object. The contribution of this paper is attractive since SAMI is basically a simple easy-to-implement heuristic method based on existing image processing techniques and some intuition-based postulates. SAMI has initially been conceived to estimate the performance of SCOTT, a "Synthesis COncealment Two-level Texture" algorithm. However, a direct derived application of such metrics could be the evaluation of saliency algorithms for object segmentation: the best saliency algorithm would be the one with the highest SAMI *D_score* for a given object. Another possible application could be the use of SAMI *inside* a saliency algorithm, to compute a dense modified saliency map, in which each pixel has the SAMI *D_score* corresponding to its neigborhood (used as ROI). Such a resulting map would be more robust to saliency noise from small spots.

## Keywords

Concealment evaluation, concealment metrics, object identification, object segmentation, saliency map, Human Visual System.

## 1 INTRODUCTION

To evaluate the detection and the identification of an object, it is necessary to compute metrics in a given Region Of Interest (ROI), basically the footprint area of this object. Of course in an application without any prior knowledge, the ROI is unknown.

For the evaluation of the detection, one could immediately think of a saliency map, which represents the ability of each pixel of an image to catch one's visual attention. However to our knowledge, to modelize visual attention, so far all saliency algorithms do not compute such specific objective metrics [Bor13a, Itt98a, Itt99a, Itt05a, Tor06a, Xu14a]. Then to estimate the global saliency impact of an object, the raw values have to be processed to answer the question: can one detect the object? Futhermore, a saliency map does not give any information about the structural appearance of the object in the scene: can one identify the object?



SAMI, for "SAliency based Metrics of Identification", is a first answer to both problems by computing a detection score, called *D_score*, for the evaluation of the detection, and an identification score, called *I_score*, for the evaluation of the identification. Only the "identification" is mentioned in the name SAMI, since the detection is implicit: an object has to be detected to be identified. The input data are the test image, a mask containing the ROI (basically the footprint area of the object, or a bounding box), and the ground truth *edges* of the image. The goal of SAMI is then to compute objective metrics of the *ability* of an observer to detect and identify a given object.

Initially, SAMI is the best friend of SCOTT: "Synthesis of COncealment Two-level Texture" [Gos14a]. SCOTT is a concealment algorithm, initially designed to reduce visual pollution caused by manmade equipment [Maz13a]: it creates a texture, of any size, faithful to the visual environment of a target, and this from only two samples from this environment: one for the *macro*-texture concealment texture, one for the *micro*-texture concealment texture. Such a texture can then be printed on a plastic film and sticked over the equipement to make it fuse with its environment. Then SAMI allows to improve SCOTT algorithm by adjusting its parameters to obtain better SAMI scores.

A direct derived application of SAMI could be the evaluation of saliency algorithms designed for an object segmentation, using only SAMI detection score *D_score*. The *D_score* could then be used as "simple" metrics of relative performance: the best segmentation would merely be the one with the highest SAMI *D_score* for a given object.

Another application of SAMI could be the computation of a robust dense visual attention map. In an application of detection based on a saliency map, the highest value is always selected as the center of visual attention. However, such a consideration is sensitive to noise in the saliency map and does not take into account the area of objects: a sole pixel, with the highest value, would be selected while an object, with a little smaller average saliency but in a greater neighboring area, would actually be the most salient object of the scene. We could process the whole saliency map with SAMI, by using a shifting spatial window as ROI, and instead of taking the maximum raw saliency value, take the highest value of the resulting dense map.

In section 2, SCOTT is briefly described to understand the problem solved by SAMI. For more details about SCOTT, see [Gos14a]. In section 3, SAMI is described with both the *D_score* and the *I_score*. In section 4 are some results of SAMI used for the improvement of SCOTT. In section 5, some other applications of SAMI are discussed. Finally an overview of the future work is given in section 6.

## 2 SCOTT: SYNTHESIS OF CONCEALMENT TWO-LEVEL TEXTURE

SCOTT [Gos14a], for "Synthesis of COncealment Two-level Texture", is an algorithm designed to compute a texture and map it onto an object so that it can visually integrate its environment [Gos14a]. SCOTT synthesizes a concealment texture, which is both generic and visually faithful to that environment, from texture samples of a visual environment: one for the *macro*-texture concealment texture, one for the *micro*-texture concealment texture (Figure 1).

SCOTT can be used for a lot of applications, like the reduction of visual pollution caused by manmade equipments (antenna, electrical cabinets, distributor boxes, repeater shelters, etc.), by giving the "polluants" an aesthetically more pleasing look [Maz13a]. First results of simulation show that SCOTT is efficient. However, so far we did not use any relevance feedback. In a process of concealment, the problem is the evaluation of *detection* and *identification* of the concealed object: SAMI is an answer to this problem.

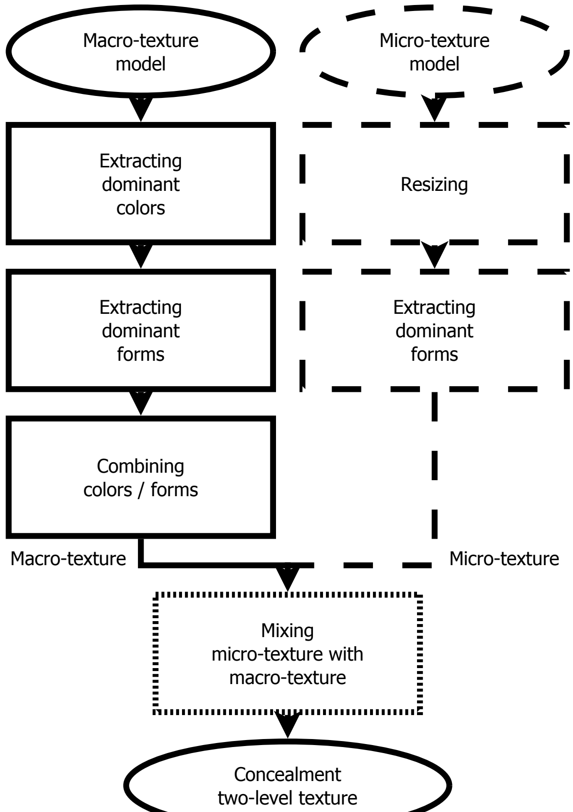



Figure 1: SCOTT is based on a "two-level" texture concept, mixing a micro-texture (dashed lines) with a macro-texture (solid lines), from two input models.

## 3 SAMI: SALIENCY BASED METRICS OF IDENTIFICATION

SAMI is basically a simple easy-to-implement heuristic method based on existing image processing techniques and some intuition-based postulates. SAMI computes objective *metrics* of the ability of an observer to detect and identify an object in a scene. By "objective" we mean here that such metrics are computed automatically, by reproducing the average observer SVH response, as they use the saliency map concept [Itt98a]. From a test image, a mask containing a Region Of Interest ROI (basically the footprint area of the object, or a bounding box), and the ground truth *edges* of the object, it computes two scores: a *detection* score, called *D_score*, for the evaluation of detection, and an *identification* score, *I_score*, for the evaluation of the identification (Figure 4).

Before computing the scores, two temporary data are computed for the whole SAMI process. First the saliency map of the test image. In SAMI, the saliency map is computed by [Itt98a], since it is the first, and so far the most used, biologically inspired saliency algorithm. Indeed, the goal of SAMI is to estimate the ability of an observer to detect and identify an object, then it is important that the whole process is based on

 ISBN 978-80-86943-71-8

the Human Visual System. Secondly, the ROI is dilated by a morphological operation, so that the area of the dilated ROI, the *d_ROI*, is *twice* that of the ROI. That is because an object can also modify the visual impact of its neighborhood, notably in terms of saliency. So the detection and identification of an object is not limited to its own area, but to the area around it (Figure 2) contained in the *d_ROI*.

(a) Original image (b) ROI (c) d_ROI (d) Image masked with *d_ROI*

Figure 2: For the entire SAMI process, the ROI (basically the footprint area of the object, or a bounding box) is processed to compute the *d_ROI*. For a given object (hydrant in (a)), the ROI is the footprint of the object (b). Once computed by morphological dilation, the *d_ROI* (c) is then not limited to the area of the object, but to the area around it (d).

## Detection score

The detection score *D_score* evaluates the *ability* of *detection* of an object in a visual environment.

**Postulate: the global saliency of an object is a function of the continuous areas of its highest values.**

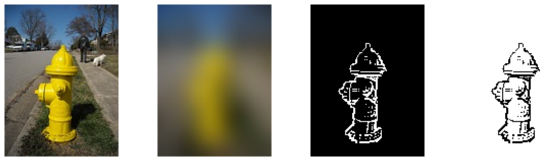
First, the saliency map is masked with the *d_ROI*. Then a mask is computed by thresholding the saliency map within the *d_ROI*. The threshold is the mean saliency of the area given by the logical *xor* between the ROI and the *d_ROI*, that is the part of the *d_ROI* around the ROI. The mask may be processed with an opening morphological operation to remove very small spots (which are unsignificant, according to our postulate). The saliency map is then masked with this last mask. The final *D_score* is then computed as the mean saliency of the remaining pixels.

The *D_score* is then an objective measure of the *ability* of *detection* of an object (Figure 4). This is an absolute score as it does not depend on any saliency value reference; then it has only a meaning when compared with the original image *D_score*, or other test image *D_scores*. Its range is between 0 and 100. Practically, through different scenarios, we have observed that the entire theoretical range is not used, then some statistical processing could be used to improve the meaning of the results (see Section 6).

## Identification score

The identification score *I_score* evaluates the *ability* of *identification* of an object in a visual environment.

**Postulate: the structure of an object stands out if its edges stands out (Figure 3).**



(a) Original image (b) Blurred image (c) Edges of the object (d) Inverse of the edges of the object

Figure 3: For a given image (a), if we observe only the color information (b), we cannot identify the main object in the scene (hydrant). What allows us to identify an object is its structure. According to our postulate, that structure stands outs if the edges stands out ((c) and (d)).

The computation of the *I_score* then requires the ground truth *edges* of the object.

First, we extract the edges of the test image, and we mask it with the *d_ROI*: this is the "test image edges". The extraction of the edges is realized by an edge extractor (e.g. Canny filter), not on the grayscale image, but on a weighted average of the normalized L* (lightness), C* (chroma) and h (hue) channels of the image. Indeed the L*C*h colorspace, derived from the L*a*b* colorspace, combines the advantges of both the L*a*b* and HSV (Hue, Saturation, Value) colorspaces [Bal12a]. A euclidian distance corresponds to a visual distance when merely switching from the L*C*h colorspace to the L*a*b* colorspace. And the components of the colors are clearly and intuitively separate in the L*C*h colorspace, like in the HSV colorspace. This way we can attach more importance to one particular component, like the *hue* channel, which we think may be more responsible for the visual sensation of edges. The pixels of the edges are weighted by their corresponding saliency values in the saliency map. This way, each edge pixel does not only stand out through a non-zero gradient in L*C*h colorspace, but also through the saliency of its neighborhood.

The idea behind the *I_score* is to compare the ground truth edges with the test image edges. Before comparing, both the ground truth edges and the test image edges may be processed with a morphological dilation. Indeed, the *d_ROI* in the test image may have the same structure as that of the ground truth edges, but its structure may have been affected by a small translation or scaling effect. Processing both the ground truth edges and the test image edges with a morphological dilation prevents from claiming that the structures are different while they are not (false negative).

 ISBN 978-80-86943-71-8

The final *I_score* is then computed, according to our postulate, as a combination of the cross-correlation of the image edges with the ground truth edges *and* with the *inverted* ground truth edges. It is comparable to a combination of the true positive rate and the false positive rate. All these operations are processed inside the *d_ROI*.

The *I_score* is then an objective measure of the *ability* of *identification* of an object (Figure 4). This is a relative score as it is a comparison to a ground truth reference; then it has a meaning by itself, as the rate of preservation of the original image structure (of course the *I_score* of the original image has no meaning since it is 100%). Its range is between 0 and 100. Practically, through different scenarios, we have observed that the entire theoretical range is not used, then some statistical processing could be used to improve the meaning of the results (see Section 6).

At the end, we have two objective metrics of the *ability* to detect and identify and object in a scene (Figure 4).

# 4 RESULTS

SAMI has initially been designed to evaluate the performance of SCOTT. The evaluation, of the results of SCOTT, proves that SCOTT does reduce the visual impact of objects [Gos14a]. It allows to improve SCOTT algorithm by adjusting its parameters to obtain better SAMI *D_score* and *I_score*.

In the evaluation of SCOTT, the name of the game is to compare an original image, containing a visually polluting object, and several test images, where the object is concealed by SCOTT.

The ground truth edges of the object, for the estimation of the identification, is merely obtained by extracting the edges of the original image (where the object is *not* concealed), the same way the edges are extracted from the test image (where the object *is* concealed) to compute the *I_score*.

From the results (Figure 5 and Figure 6), we can first conclude that SCOTT is efficient, since the scores are better (smaller in this case) with SCOTT textures. Secondly, we understand that the need here is to improve SCOTT algorithm by adjusting its parameters, since some concealment textures are clearly better than others.

Of course SAMI can also be used to evaluate general purpose texture synthesis algorithm, like *inpainting*.

# 5 OTHER APPLICATIONS

We think that SAMI could be used in two other applications: the comparison of saliency algorithms and the computation of visual attention map.

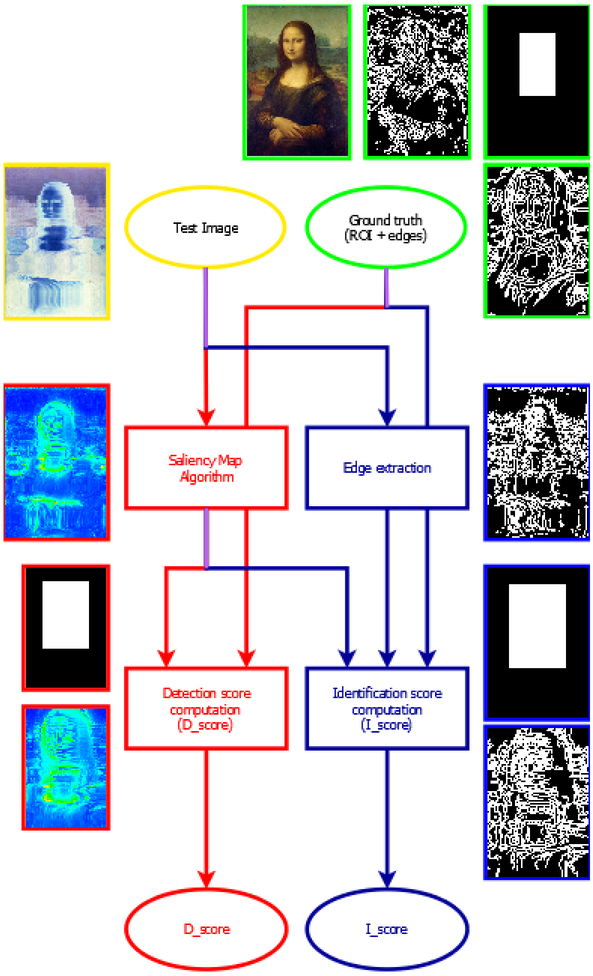


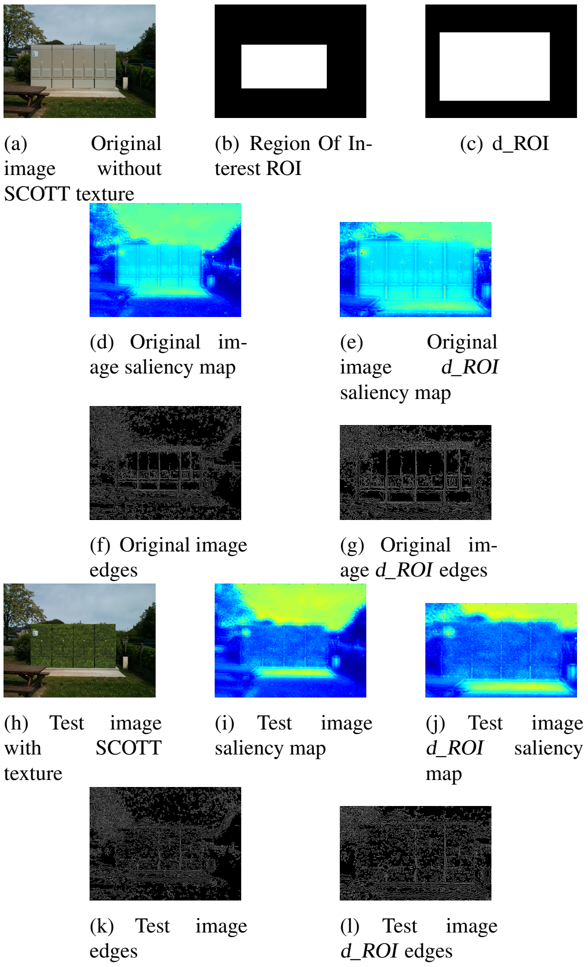

Figure 4: SAMI computes objective metrics of the ability of an observer to detect and identify a given object in a scene. From an input test image (distorted *Mona Lisa*, with yellow borders), the original image and a Region Of Interest ROI (both with green borders), SAMI computes two scores for this ROI: a detection score *D_score* and an identification score *I_score*. The *D_score* (red path) is computed from the saliency map of the test image ROI. The *I_score* is computed as a comparison between the ground truth edges (computed from the original image, both with green borders) and the test image edges, weighted by the test image saliency. The *D_score* is 11.02%, and the *I_score* is 14.86%. These scores confirm that the ROI is salient compared to the rest of the painting (Leonardo da Vinci would certainly agree), and the test image ROI structure does not match that of the original version, since it has been distorted.

## Comparison of saliency algorithms

SAMI could be used to compare saliency algorithms in an application of object segmentation. Indeed the best algorithm would then be the one with the "best" (the highest in this case) SAMI *D_score* for this object.

In this application, SAMI should be adapted as follows: there would be no more *d_ROI*, only the ROI is used

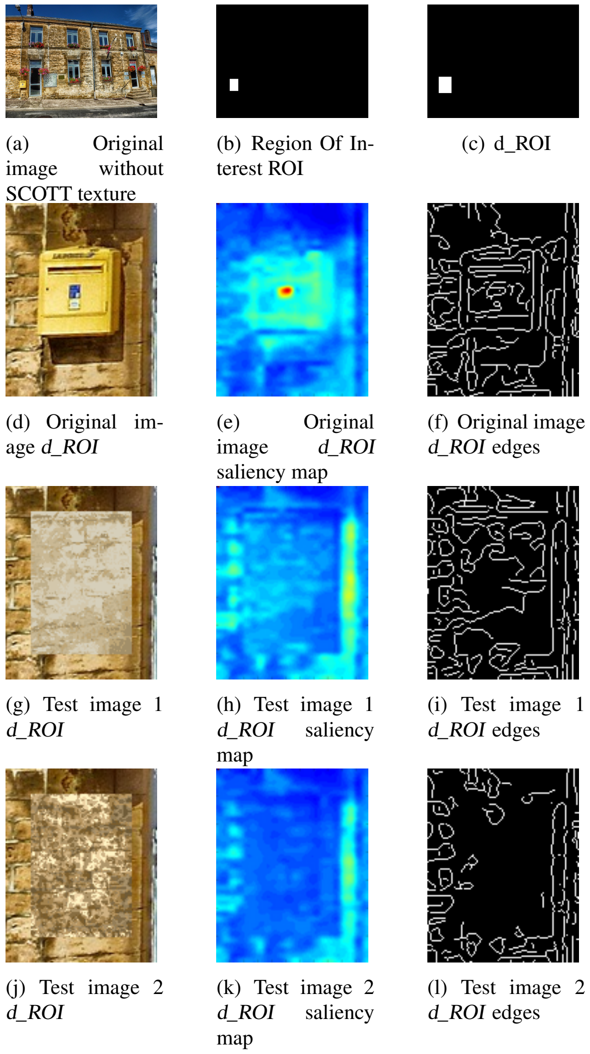
 ISBN 978-80-86943-71-8



(a) Original image without SCOTT texture
(b) Region Of Interest ROI
(c) d_ROI
(d) Original image saliency map
(e) Original image *d_ROI* saliency map
(f) Original image edges
(g) Original image *d_ROI* edges
(h) Test image with SCOTT texture
(i) Test image saliency map
(j) Test image *d_ROI* saliency map
(k) Test image edges
(l) Test image *d_ROI* edges

Figure 5: SAMI has initially been designed to evaluate the performance of SCOTT, a concealment algorithm. In this example, we have a concrete example of visual pollution caused by a repeater shelter without (a) and with (h) a SCOTT concealment texture. The object is contained in a bounding box used as Region Of Interest ROI (b), from which is computed the dilated ROI *d_ROI* (c). SAMI compute the detection score *D_score* from the saliency map of the test image (i) in the ROI (j); this *D_score* can be compare to that of the original image (d) in the ROI (e). The identification score *I_score* is computed from the edges of both the original image and the test image ((f) and (k), respectively) in the ROI ((g) and (l), respectively). In this simulation, the concealment texture has been mapped on the object in a way that it best fits its 3D shape. The SAMI results prove that SCOTT is effecient by making the polluting object less salient and by breaking its visual structure: the *D_score* of the original image is 12.63%, while that of the test image is 6.34%; the *I_score* of the test image is 18.12%.

(no dilation of the ROI), and then the *D_score* threshold should merely be computed as the mean saliency of the not-ROI area.



(a) Original image without SCOTT texture
(b) Region Of Interest ROI
(c) d_ROI
(d) Original image *d_ROI*
(e) Original image *d_ROI* saliency map
(f) Original image *d_ROI* edges
(g) Test image 1 *d_ROI*
(h) Test image 1 *d_ROI* saliency map
(i) Test image 1 *d_ROI* edges
(j) Test image 2 *d_ROI*
(k) Test image 2 *d_ROI* saliency map
(l) Test image 2 *d_ROI* edges

Figure 6: SAMI allowed us to improve SCOTT, by evaluating its resulting images, that is the images with SCOTT textures, and making the SAMI scores better (smaller). In this example, we have a visual environment (a), with an object we want to conceal (d): the mailbox. By using a bounding box as the ROI (b), from which is computed the dilated ROI *d_ROI* (c), SAMI compares the ROI saliency maps and edges, in the original image ((d), (e) and (f)), in the same image with a concealment texture from the original version of SCOTT "test image 1" ((g), (h) and (i)), and in the same image with a concealment texture from an improved version of SCOTT "test image 2" ((j), (k) and (l)). These results prove that we can improve SCOTT by evaluating its results with SAMI, and take the best (smallest) scores. In this scenario, the *I_score* is not relevant since the concealment textures do not fit the shape of the object: they are just superimposed on the 2D plane of the image. Here are the scores: the *D_score* of the original image is 12.72%, while that of the test image 1 is 7.05% and that of the test image 2 is 4.50%.

 ISBN 978-80-86943-71-8

For example, if we compare Itti [Itt98a] saliency algorithm with a more recent one like MSSS [Ach10a], we find out that the second one would be better for an application of saliency based object segmentation, since its SAMI *D_score* is higher than Itti's (Figure 7): 22.87 against 17.82. Indeed, visually, the target object is more highlighted in MSSS saliency map than in Itti saliency map.

### Computation of visual attention map

Another possible application could be to include SAMI *in* a saliency algorithm, by processing the whole saliency map with SAMI to compute a visual attention map, as a dense SAMI *D_score* map.

Indeed, in a visual attention modeling process [Bor13a, Itt98a, Itt99a, Itt05a, Tor06a, Xu14a], we simply use the raw values of the saliency map: the maximum would define the first point of focus, then the second maximum would define the second point of focus, and so on. However, such a process would not take into account the size of the saliency spots, and it is our intuition that the global saliency of a region is a function of the continuous areas of its highest values. That is why SAMI could be useful to improve such a process.

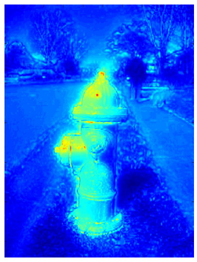
To do so, SAMI would merely process the saliency map with a shifting neighborhood window as ROI (no more *d_ROI*). Then a *D_score* would be computed for each pixel of the saliency map. Mathematically, in this case, SAMI is a non-linear function which moves the saliency values away from each other: the small values become smaller and the high values become higher; it is like increasing the local contrast. The results (Figure 8), realized with the saliency algorithm Itti [Itt98a], prove that the resulting map highlights better a salient object in a scene, by removing secondary objects in the background.

## 6 FUTURE WORK

First, the meaning of the results could me improved by statistically processing the SAMI scores. Indeed we have observed that the theoretical ranges of the score are not fully used, notably because the theoretical maximum (100) of the functions cannot be reached. By studiying the practical useful ranges of these scores, we could compare the results better by clipping and stretching the results.

Secondly, even if the results of evaluation by SAMI show that such an evaluation is relevant, and even if SAMI does allow us to improve SCOTT algorithm by adjusting its parameters, SAMI has not been validated so far. In other words, we also need a subjective evaluation for the objective evaluation! Since by "objective" we mean "HVS-inspired automatically computed", immediate future work will be to validate SAMI by comparing the obtained results with data provided by test subjects.

(a) Original image object

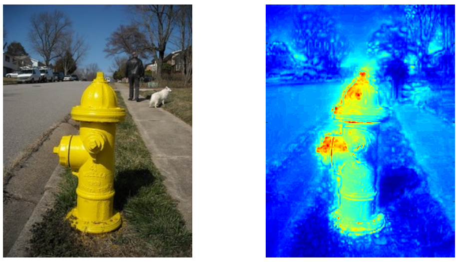


(b) Original image saliency map: Itti

(c) Original image saliency map : MSSS

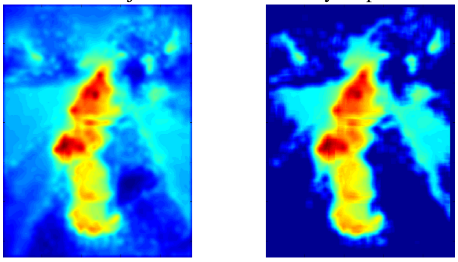
Figure 7: SAMI could be used to compare saliency algorithms for an application of object segmentation. The best saliency algorithm is the one which highlights the most the object, that is the one with the best (highest) *D_score* for this object (that is the ROI and there is no more *d_ROI*). In this example with a very salient object in a image (a), SAMI can compare the saliency map of two algorithms: Itti (b) and MSSS (c). According to SAMI *D_score*, MSSS is the best for an application of object segmentation, as it highlights more the salient object with the highest *D_score*: 22.87% against 17.81%.

But the main future work will be to combine both SCOTT and SAMI in a feedback loop (Figure 9). One of the most influential parameters of SCOTT is the criterion for the selection of the input samples as models to synthesize a concealment texture faithful to the visual environment: *macro*-texture model and *micro*-texture model. So far, this selection is done manually (Figure 6), according to what the user wants the concealment texture to look like; then the criterion of selection is purely subjective. But it could be very beneficial to the algorithm that this selection is automatic. First, SAMI would extract, from an image of a given visual environment, the possible pairs of input SCOTT models (*macro*-texture model and *micro*-texture model), as the least salient samples in the environment for example. Then for each pair of models, SCOTT would synthesize a concealment texture. A simulation module would then simulate the mapping of each concealment texture onto the object to conceal. From the renderings of the simulation, SAMI would then evaluate the results by computing both the *D_score* and the *I_score*. Finally, the concealment texture with the best SAMI scores would be selected as the SCOTT concealment



(a) Original image with salient object

(b) Original image saliency map: Itti



(c) Original image saliency map: Itti and low-pass filter

(d) Original image ROI saliency: Itti and SAMI

Figure 8: Another possible application could be to include SAMI *in* a saliency algorithm, by processing the saliency map to compute a visual attention map, as a dense SAMI *D_score* map. Indeed, SAMI acts as a non-linear function which moves the saliency values away from each other: the small values become smaller and the high values become higher; it is like increasing the local contrast. This way the most salient object is more highlighted while the secondary objects are removed. If we take the same example as Figure 7 (a), with the saliency algorithm Itti, we see that some secondary objects are salient, while the main object (hydrant) is not really highlighted (b). However, with SAMI, only the hydrant has a high value over a big area (d). This way the hydrant is much more highlighted. One could argue that SAMI has the same effect as a mere low-pass filter. But the result of a low-pass filter (b) shows that, even if the hydrant is more highlighted, the saliency value of the background is bigger on average. Besides, if we focus on the main object (that is the ROI and there is no more *d_ROI*), the *D_score* is the best (highest) for the algorithm Itti *with* SAMI: 48.85% for Itti, 62.78% for Itti filtered by a low-pass filter, and 63.56% for Itti *with* SAMI.

texture to use. One problematic is to find the criterion to select the best SAMI scores. Indeed, for each score the smaller the better, but it is possible that no SCOTT concealment texture has both the best *D_score* and *I_score*. This criterion may then depend on the user setting and/or the visual environment itself.

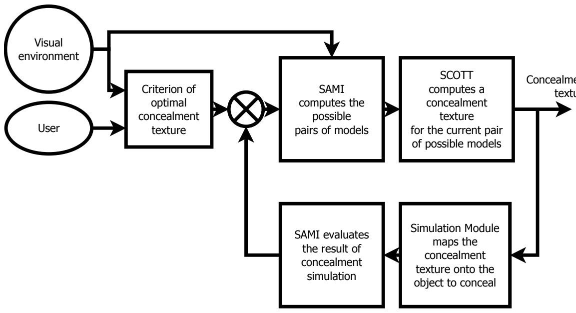



Figure 9: Future work will be to combine both SCOTT and SAMI in a feedback loop. One of the most SCOTT influential parameters is the selection of the input sample as models to synthesize a concealment texture faithful to the visual environment: *macro*-texture model and *micro*-texture model. It could be very beneficial to the algorithm that this selection is automatic.

## 7 CONCLUSION

SAMI provides original metrics of *detection* and *identification* of an object in an image, based on the saliency map of this image. To do so, SAMI computes two complementary scores, a detection score, called *D_score*, and an identification score, called *I_score*, which evaluate the detection and the identification of the object, respectively. SAMI has been initially designed to evaluate SCOTT, a concealment texture algorithm synthesis, to complete our concealment algorithm package. This way we improve SCOTT and we are considering incorporating SAMI *inside* SCOTT itself, to make automatic the selection of SCOTT input models. A first other possible application of SAMI is the evaluation of saliency algorithms for an application of object segmentation. Finally, SAMI could even be included in a saliency algorithm, by processing the whole saliency map with a shifting window, then computing a visual attention map as a dense SAMI *D_score* map.

## 8 REFERENCES

[Ach10a] Achanta, R., and Susstrunk, S. Saliency Detection using Maximum Symmetric Surround. In Proceedings of 2010 IEEE 17th International Conference on Image Processing, pp.2653-2656, Hong Kong, 26-29 September 2010.

[Bal12a] Baldevbhai, P.J., and Anand, R.S. Color Image Segmentation for Medical Images using L*a*b* Color Space. IOSR Journal of Electronics and Communication Engineering (IOSRJECE) , Vol.1, No.2, pp.24-45, May-June 2012.

[Bor13a] Borji, A., and Itti, L. State-of-the-Art in Visual Attention Modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.35, No.1, pp.195-207, January 2013.

[Maz13a] Dandumont, P. Question No. 26298 de M. Francois de Mazieres a Mme la ministre de la culture et de la communication, 14 May 2013. http://questions.assemblee-nationale.fr/q14/14-26298QE.htm

[Gos14a] Gosseaume, J., Kpalma, K., and Ronsin, J. The World vs. SCOTT : Synthesis of Concealment Two-level Texture. In Proceedings of the 9th International Conference on Computer Vision Theory and Applications, Vol.1, pp.257-264, Lisbon, Portugal, 5-8 January 2014.

[Itt98a] Itti, L., Koch, C., and Niebur, E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.20, No.11, pp.1254-1259, November 1998.

[Itt99a] Itti, L., Koch, C. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research, Vol.40, pp.1489-1506, 2000.

[Itt05a] Navalpakkam, V., and Itti, L. Modeling the influence of task on attention. Vision Research, Vol.45, pp.205-231, 2005.

[Tor06a] Torralba, A., Oliva, A., Castelhano, M.S., and Henderson,J.M. Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search. Psychological Review, Vol.113, No.4, pp.766-786,2006.

[Xu14a] Xu, J. Hierarchical Bayesian Modelling of Visual Attention. In Proceedings of the 9th International Conference on Computer Vision Theory and Applications, Vol.1, pp.347-358, Lisbon, Portugal, 5-8 January 2014.