Detection of Regions of Interest and Camouflage Breaking by Direct Convexity Estimation

Ariel Tankus Yehezkel Yeshurun

Tel-Aviv University
Department of Computer Science
Ramat-Aviv 69978, Israel
{arielt,hezy}@math.tau.ac.il

Abstract

Detection of Regions of Interest is usually based on edge maps. We suggest a novel non-edge-based mechanism for detection of regions of interest, which extracts 3D information from the image. Our operator detects smooth 3D convex and concave objects based on direct processing of intensity values. Invariance to a large family of functions is mathematically proved. It follows that our operator is robust to variation in illumination, orientation, and scale, in contrast with most other attentional operators. The operator is also demonstrated to efficiently detect 3D objects camouflaged in noisy areas. An extensive comparison with edge-based attentional operators is delineated.

1. Introduction

Automatic detection of regions of interest in images (also referred to as "Attentional" algorithms) is usually based on edge maps. Some mechanisms use the points of high curvature of the edge map as their interest points [11]. Corner detection [9] is another means of detection of interesting points in the image. The common feature of all these mechanisms and many others (see [10]) is that directly or indirectly the attentional mechanisms are based on edge maps. Even more recent attentional mechanisms such as the generalized symmetry transform [4] receive the edge map as their input. This excludes some works which utilize color [2] or motion [7] as attentional mechanisms, and direct gray value processing for anchor points in object recognition [6]. Though one cannot disregard their advantages, edge maps sustain severe flaws such as:

- **Sensitivity to illumination:** Strong illumination might change the strength of the edges, and thus, the region of interest.
- **Scale:** Existing attentional operators search for interesting objects of a predetermined scale. Multi-scale approaches (e.g. [3], [8]), are computationally heavy. They are not a neat solution as they apply a heuristic to a scale-dependent method.
- **Strong effect of the surroundings:** The outline edges are not intrinsic to the object, but are affected from background objects.
- **In cluttered/textured scenes:** Edge-based methods are unable to separate different objects.

Our main goal is, thus, to develop a novel attentional mechanism, that will overcome the typical pitfalls of edge based methods. The operator we propose is applied to the intensity image, and responds to 3D convex or concave regions. The purpose of the suggested operator is the detection of regions of interest. The operator is robust to *illumination*, *scale*, and *orientation*. It is capable of focusing the attention in noisy environments; e.g, in a strongly textured background. It is even capable of breaking very strong camouflage of 3D objects, including camouflage which might delude a human viewer. The invariance and robustness features of the operator are mathematically proven. The application of the operator to real-life images demands a relatively short running time, and its robustness leads to reliable results.

2. Attentional operator for detection of convex regions

We next define the suggested attentional mechanism.

2.1. Defining the argument of gradient

Let us estimate the *gradient map* of image $I(x, y)$ by:

$$\nabla I(x, y) \quad \approx \quad ([D_\sigma(x)\, G_\sigma(y)] * I(x, y),$$

$$[G_\sigma(x)\, D_\sigma(y)] * I(x,y))$$

where $G_\sigma(t)$ is the 1D Gaussian with zero mean and standard deviation σ, and $D_\sigma(t)$ is the derivative of that Gaussian. We turn the Cartesian representation of the intensity gradient into a *polar* representation. The *argument* is defined by:

$$\begin{aligned}\theta(x,y) &= \arg(\nabla I(x,y)) \\ &= \arctan\Big(\tfrac{\partial}{\partial y}I(x,y)\ ,\ \tfrac{\partial}{\partial x}I(x,y)\Big)\end{aligned}$$

where the two dimensional arc tangent is defined by:

$$\arctan(y,x) = \begin{cases} \arctan(\frac{y}{x}), & \text{if } x \geq 0 \\ \arctan(\frac{y}{x}) + \pi, & \\ & \text{if } x < 0\ ,\ y \geq 0 \\ \arctan(\frac{y}{x}) - \pi, & \\ & \text{if } x < 0\ ,\ y < 0 \end{cases}$$

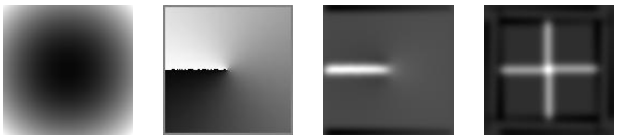
and the one dimensional $\arctan(t)$ denotes the inverse function of $\tan(t)$ so that: $\arctan(t) : [-\infty, \infty] \mapsto \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. The attentional mechanism is simply the derivative of the argument map with respect to the y-direction: $\frac{\partial}{\partial y}\theta(x,y) \approx [G_\sigma(x)\, D_\sigma(y)] * \theta(x,y)$. The attentional operator, $\frac{\partial}{\partial y}\theta(x,y)$, is denoted: *Y-Arg*.

The intuition for the operator is: We are looking for zero-crossings of the argument of the intensity gradient. The zero-crossings are singled out using the derivation of the arctan, since at the negative x-axis the arctan is discontinuous, and its derivative there tends to infinity. These zero-crossings enforce a certain range of angles on the argument of the intensity gradient. This range of angles ensures that the detected object is either convex or concave.

2.2. Y-Arg response to paraboloids

The projection of concave and convex objects can be estimated by paraboloids (Fig. 1(a)), since paraboloids are arbitrarily curved surfaces (see [12]). Our mathematical formulation refers to a general paraboloid of the form: $f(x,y) = a(x-\epsilon)^2 + b(y-\eta)^2$, where $a > 0\,, b > 0$ are constants, and (ϵ, η) is the center of the paraboloid. The gradient argument is therefore: $\theta(x,y) = \arctan(b(y-\eta), a(x-\epsilon))$ (See Fig. 1(b)). The derivative of the gradient argument exists in the whole plane except for the ray: $\{(x,y)\,|\,y = \eta \text{ and } x \leq \epsilon\}$. At this ray, $\theta(x,y)$ has a first order discontinuity (in the y-direction), so its derivative there tends to infinity. The strong (infinite) reaction at that ray appears in Fig. 1(c).

The reader should note, that the argument of the gradient is being used in Computer Vision for a long time.



(a) (b) (c) (d)

Figure 1. ***(a)*** **The spheric gray-levels:** $I(x,y) = 10x^2 + 10y^2$**.** ***(b)*** **The argument of gradient of (a). The discontinuity ray is at** $180°$ **from the positive** x**-axis.** ***(c)*** **Y-Arg of (a).** ***(d)*** **Response of D-Arg, the isotropic operator.**

Hough transform [1], for example, uses the argument of the gradient to reduce the space of parameters when searching straight lines or circles in the image. Image improvement by argument-based methods is also known, and can be found in [5]. Nonetheless, the novel idea in the suggested operator is not the argument of the intensity gradient, but rather the usage of the discontinuity ray formed by the argument of the intensity gradient. The key idea for the detection scheme we describe lies in this discontinuity ray, which is detected by derivation of the gradient argument, to receive a strong response to that ray. The search for zero-crossings in the argument of the intensity gradient is very stable.

2.3. D-Arg: The isotropic variant

The strong reaction of Y-Arg at the negative part of the x-axis is caused when $\theta(x,y)$ changes sharply from high values (approx. π) to low values (approx. $-\pi$). This need not happen only on paraboloidal gray-levels: cropping a small strip around the negative part of the x-axis from the spheric gray-levels would still produce high values, since $\frac{\partial}{\partial y}\theta(x,y) \to \infty$ there. In order to avoid Y-Arg dependence on convexity orientation, we define an isotropic operator; i.e., an operator that would strongly react to *all* convexity orientations. A general way of doing so would be to rotate the original image by $\pi - \alpha$ degrees, calculate Y-Arg for the rotated image, and rotate the result back to the original angle (by $\alpha - \pi$ degrees). We call the result of this process: α-Arg. Our isotropic operator is defined to be the sum of α-Arg for the angles: $\alpha = 0°, 90°, 180°, 270°$. We call this operator: *D-Arg*. Figure 1(d) shows the D-Arg of the paraboloidal gray-levels. As expected, a strong reaction appears in all axes.

3. Features of D-Arg

We next consider several features of Y-Arg. The same features hold for D-Arg too, by definition.

Planar objects of constant albedo form linear gray-level

functions, and are usually of little interest (e.g., walls). The following properties can be shown:

- Y-Arg has zero response to planar objects.
- The response of Y-Arg to edges of planar objects is finite, and is, consequently, smaller than its response to paraboloids.
- Y-Arg depends on the image scale linearly.

The 2D convexity of 2D objects (e.g. a circle in the plane) is the convexity of their edges referred to as contours. As we saw above, D-Arg does not strongly respond to edges, but rather to paraboloidal gray levels. In this manner, the convexity detected by D-Arg is *three dimensional*. Thus, D-Arg exploits the 3D information concealed in the image.

Nonetheless, the most important property of Y-Arg is:

Y-Arg is invariant under any derivable strongly monotonically increasing transformation of the gray-level function.

The following theorem proves the claim mathematically:

Theorem 1 *Let $f(x,y)$* [*the original gray-level function*] *be a derivable function at each pixel (x_0,y_0) with respect to x and y.*
Let $T(z)$ [*the transform*] *be a function derivable at point $z_0 = f(x_0,y_0)$, whose derivative there is positive in the strong sense. The composite function: $g(x,y) = T(f(x,y))$* [*the transformed gray-level function*].
The y-derivatives of the gradient arguments of $f(x,y)$ and $g(x,y)$ at point (x_0,y_0) are identical:

$$\frac{\partial\theta_g(x_0,y_0)}{\partial y} = \frac{\partial\theta_f(x_0,y_0)}{\partial y}$$

Proof: By the *chain rule*, the composite function: $g(x,y) = T(f(x,y))$ is derivable with respect to both x and y at point (x_0,y_0), and its derivatives are:

$$\begin{aligned} g_x(x_0,y_0) &= T'(f(x_0,y_0))f_x(x_0,y_0) \\ g_y(x_0,y_0) &= T'(f(x_0,y_0))f_y(x_0,y_0) \end{aligned}$$

Let $f^0 = f(x_0,y_0)$, $f^0_x = f_x(x_0,y_0)$, $f^0_y = f_y(x_0,y_0)$. The argument of the gradient at point (x_0,y_0) can be written as:

$$\theta_g(x_0,y_0) = \arctan\left(T'(f^0)f^0_y, T'(f^0)f^0_x\right)$$

Since we have required that $T'(f^0) > 0$, the point $\left(T'(f^0)f^0_x\ ,\ T'(f^0)f^0_y\right)$ lies in the same quarter of the plane as point (f^0_x, f^0_y). It follows that:

$$\begin{aligned} \theta_g(x_0,y_0) &= \arctan\left(T'(f^0)f^0_y, T'(f^0)f^0_x\right) \\ &= \arctan\left(f^0_y, f^0_x\right) \\ &= \theta_f(x_0,y_0) \end{aligned}$$

The last equation states that the argument of the intensity gradient is invariant under the transformation T. Deriving the gradient argument with respect to y preserves this invariance:

$$\frac{\partial\theta_g(x_0,y_0)}{\partial y} = \frac{\partial\theta_f(x_0,y_0)}{\partial y}$$

□

The practical meaning of the theorem is that Y-Arg is invariant, for example, under linear transformations, positive powers (where $f(x,y) > 0$), logarithm, and exponent. It is also invariant under compositions and linear combinations (with positive coefficients) of these functions. These functions are common in image processing for lighting improvement. This implies that Y-Arg is robust to a large variety of lighting conditions. Figure 2 demonstrates Y-Arg invariance to $\log(\log(\log(I)))$ and $\exp(\exp(\exp(I)))$ in a real-life scene.

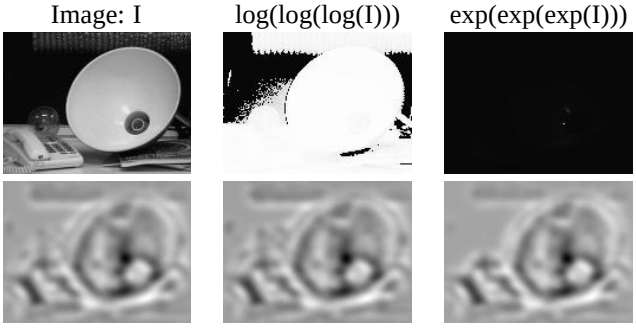


Figure 2. Invariance to derivable strongly monotonically increasing transformation of the gray-level function. *Top row:* The original image $I(x,y)$ is compared to log(log(log($I(x,y)$))) and exp(exp(exp($I(x,y)$))). *Bottom row:* D-Arg. The similarity between the D-Arg of the original image and the D-Args of the transformed images is obvious.

3.1. Regions of interest

The goal of this work is to overcome the problems inherent to edge-based detection. Edge maps fails to convey the necessary information for the detection of the subject when the background contains many strong edges. This kind of background has two disadvantages: First, the background distracts the attention from the subject when an edge-based technique is employed. Second, backgrounds full of edges are imitated by camouflage. The camouflage extends the edges of the natural environment by artificial edges on the cover (or clothes) of the concealed object. We cope with

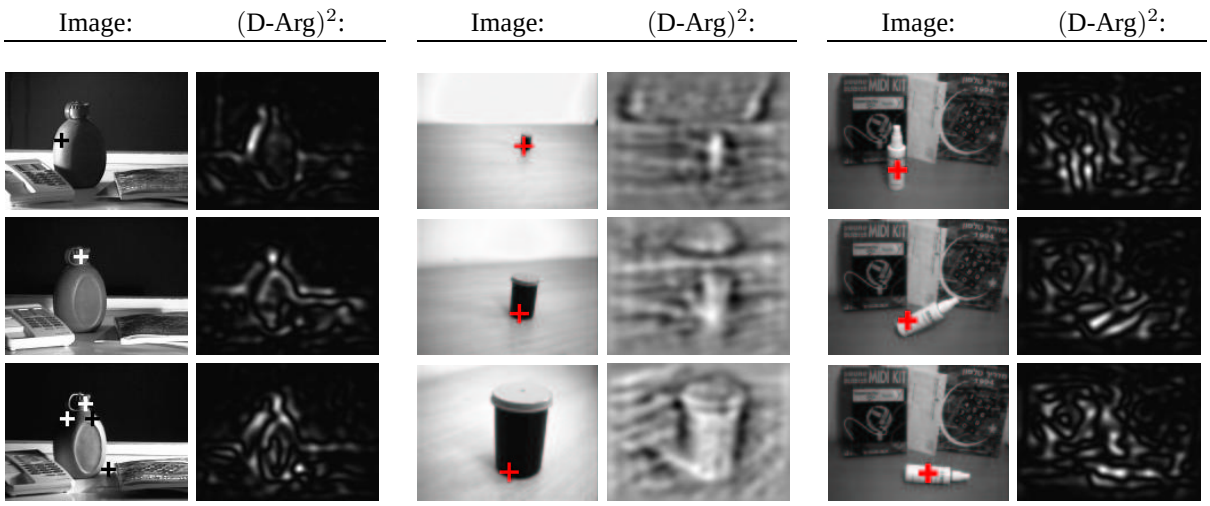


Figure 3. ***Left:*** **Robustness to illumination directions. D-Arg strongly reacts to the canteen, and is independent of the lighting direction. Each row corresponds to an illumination direction with an azimuth of $-90°$, $0°$ or $90°$, respectively.** ***Middle:*** **Robustness to scale. A film box in 3 different scales. The maximal D-Arg reaction is marked on each original image.** ***Right:*** **Robustness to orientation. D-Arg strongly reacts to the cylindric bottle, independent of its orientation.**

these problems by the D-Arg operator, which detects regions of interest characterized as smooth 3D convex and concave objects. By using three dimensional convexity, D-Arg can handle the above-mentioned problems even in very noisy images.

4. D-Arg robustness demonstration

D-Arg is robust to the following three factors: illumination, scale, and orientation. The robustness to the last two factors is gained mainly due to the fact that $\frac{\partial}{\partial y}\theta(x,y) \rightarrow \infty$ at the negative x-axis of paraboloids, which is a very stable feature. Scale and orientation variations preserve this divergence to infinity. Robustness to illumination changes has been proved in detail in Sect. 3. For all examples, the same dimensions for the environments were used: 2×2 pixels for the first-order derivation, and 14×14 pixels for the derivation of $\theta(x,y)$.

Robustness to illumination. In Fig. 3, a single point light source illuminates the scene. The light source is placed in seven positions which are equidistant from the subject, to form the azimuths $-90°$, $0°$, $90°$, with respect to the line connecting the subject and the camera. In each of the images, the regions of the highest D-Arg values are those of the canteen.

Robustness to scale. In Fig. 3, a cylinder-like black film box is photographed in 3 different scales. Each image is accompanied by its D-Arg. The film box is detected by D-Arg in each of the scales. It should be noted that throughout this article the environments for calculating D-Arg (and similarly, Y-Arg) are fixed. The robustness to scale is maintained since $\frac{\partial}{\partial y}\theta(x,y) \rightarrow \infty$ at the negative x-axis of paraboloids, which is a very stable feature. Regardless of the scale of the paraboloid, a small environment around its center would cause an infinite response.

Robustness to orientation. The D-Arg operator is isotropic by definition. Figure 3 demonstrates its robustness to the orientation of the maximal convexity of the object. In all orientations the bottle has been correctly and consistently detected as the region of interest in the image. Again, as $\frac{\partial}{\partial y}\theta(x,y) \rightarrow \infty$ at the negative x-axis of paraboloids, any orientation of the paraboloid which is not exactly perpendicular to the x-axis would result in an infinite response of Y-Arg. By definition, D-Arg overcomes the perpendicular extreme case as well.

4.1. Superiority of D-Arg on edge-based detection

We delineate the results of an extensive comparison between D-Arg and edge maps. We confront D-Arg with another context-free attentional mechanism: the radial symmetry operator (defined in [4]). This operator seeks for a generalized notion of radial symmetry in the edge map of the original image. As we would see, D-Arg performs better than edge-based methods in a large variety of situations.

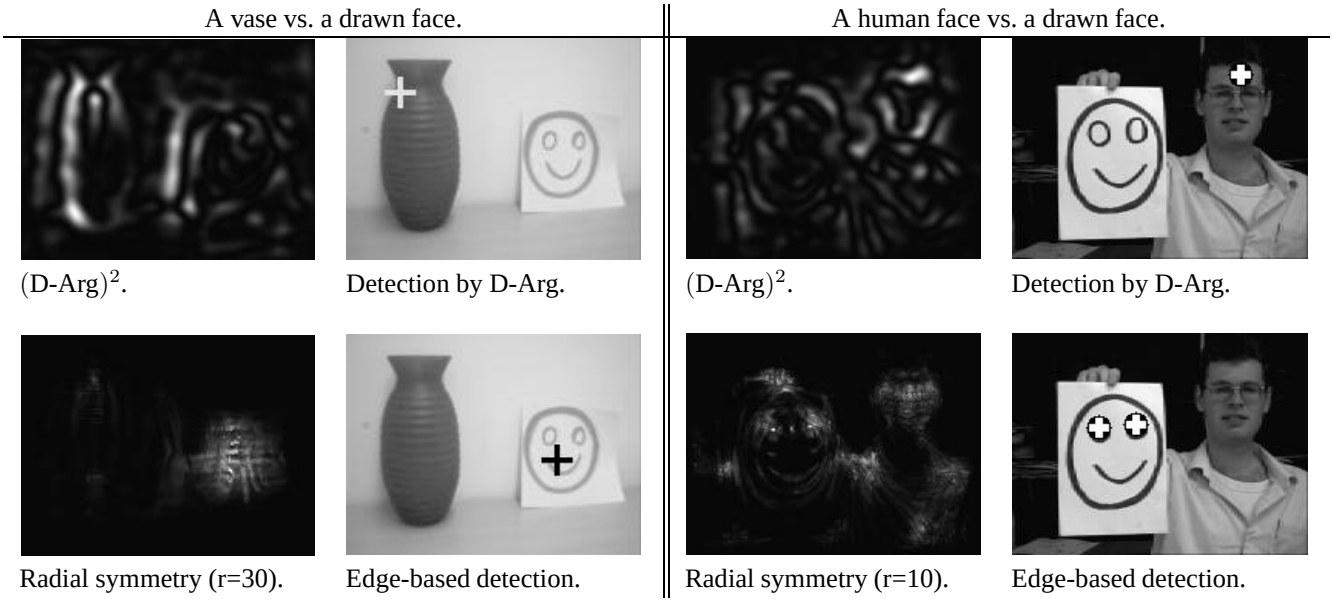


Figure 4. Reaction to 3D objects. Edge-based detection locates the drawn "Smiley", though it is a flat drawing. D-Arg detects the vase or the human face, as they are 3D convex objects.

Reaction to 3D objects. Smooth 3D convex objects exhibit a gradual change in gray levels, so their edges are weak. Therefore, in Fig. 4 the edges produced by the "Smiley" face predominate the edge maps, despite the presence of the vase or the human face, respectively. The drawn face is much more prominent in the edge map, and is therefore detected by edge-based methods. D-Arg, on the other hand, strongly reacts to 3D convex objects, while maintaining a relatively low reaction to 2D objects. This makes the response of D-Arg to a 2D drawn "Smiley" face relatively low in comparison with its response to the detected three dimensional vase or human face.

Stability in textured backgrounds. The existence of texture in an image makes the task of separating the subject from the background very hard. Figure 5 illustrate D-Arg robustness in dominant textures. Since the texture is 2D (i.e. drawn on a flat paper), D-Arg reacts relatively low to texture. The edge map contains a significant image *area* covered with edges. This contrasts with our intuition that edges describe the contours of objects, rather than areas. Areas covered with edges prevent attentional operators from distinguishing between different objects.

5. Camouflage breaking

The stability of D-Arg under various conditions (illumination, scale, orientation, texture) makes it suitable for usage as a camouflage breaker.

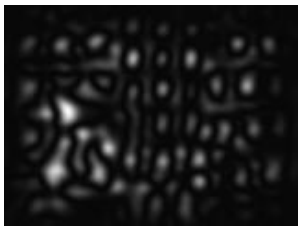
A camouflage approach taken by hunters in woods is mimicking the environment by clothes with stripes whose distribution is as close as possible to the stripes formed by the trees. This camouflage is exactly of the kind which deceives edge-based detection of regions of interest. The convexity of the gray-levels of the hunter is much stronger then that of the background. The usage of stripes does not conceal the convexity of the hunter, and thus D-Arg is capable of detecting the hunter, as can be seen in Fig. 6.

Another type of camouflage, used especially by the military, is achieved by covering most of the body or the equipment under branches of trees (i.e, edges) common in the area of operation. The camouflaged forces are detectable by means of convexity, as can be seen in Fig. 7: a machine-gun position manned by two soldiers is camouflaged by branches of trees. Edge-based techniques fail on this image. D-Arg detects the soldiers, as their (uncovered) faces are 3D and convex.

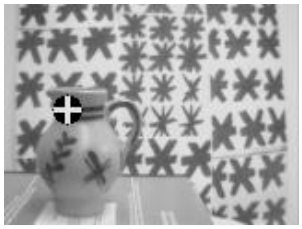
6. Conclusions

Currently, the vast majority of attentional operators are based on edge maps. Nevertheless, many images show that edge-based information is not enough to detect camouflaged objects, or objects in cluttered images. This implies, that a wide range of images cannot be successfully handled by edge-based methods. To answer these cases, we have proposed the use of three dimensional information. We have introduced a novel attentional operator, D-Arg, for the

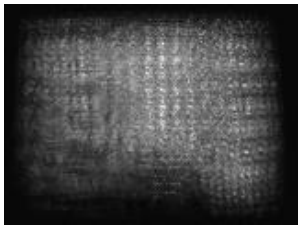
A vase on a highly textured background.



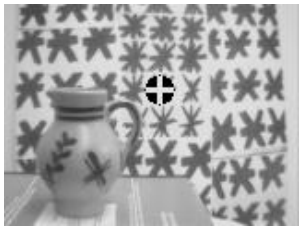
$(\text{D-Arg})^2$.



Detection by D-Arg.

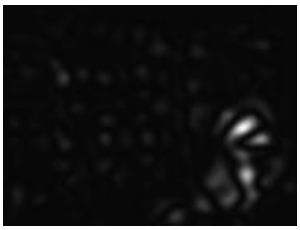


Radial symmetry (r=30).



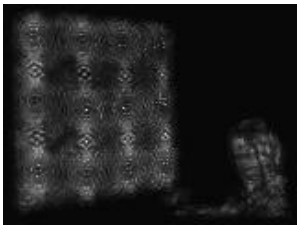
Edge-based detection.

A human face on a highly textured background.



$(\text{D-Arg})^2$.

Detection by D-Arg.

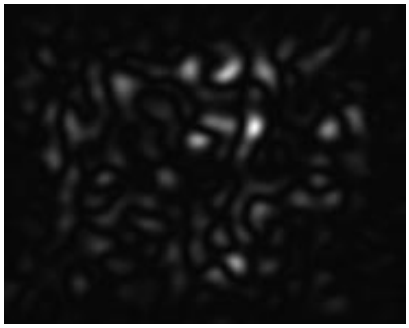


Radial symmetry (r=10).

Edge-based detection.

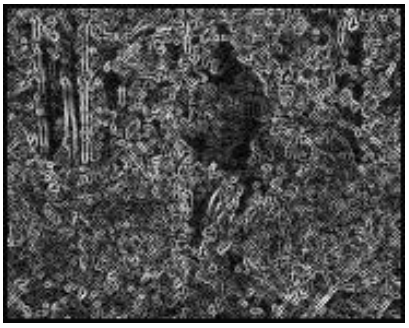
Figure 5. Highly textured environments. D-Arg detects the vase or the human face being a 3D object, despite the strong dominant texture.

Original image.

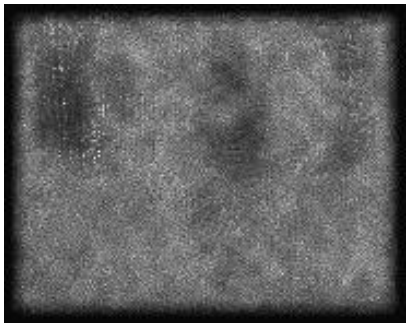


$(\text{D-Arg})^2$.

Detection by D-Arg.



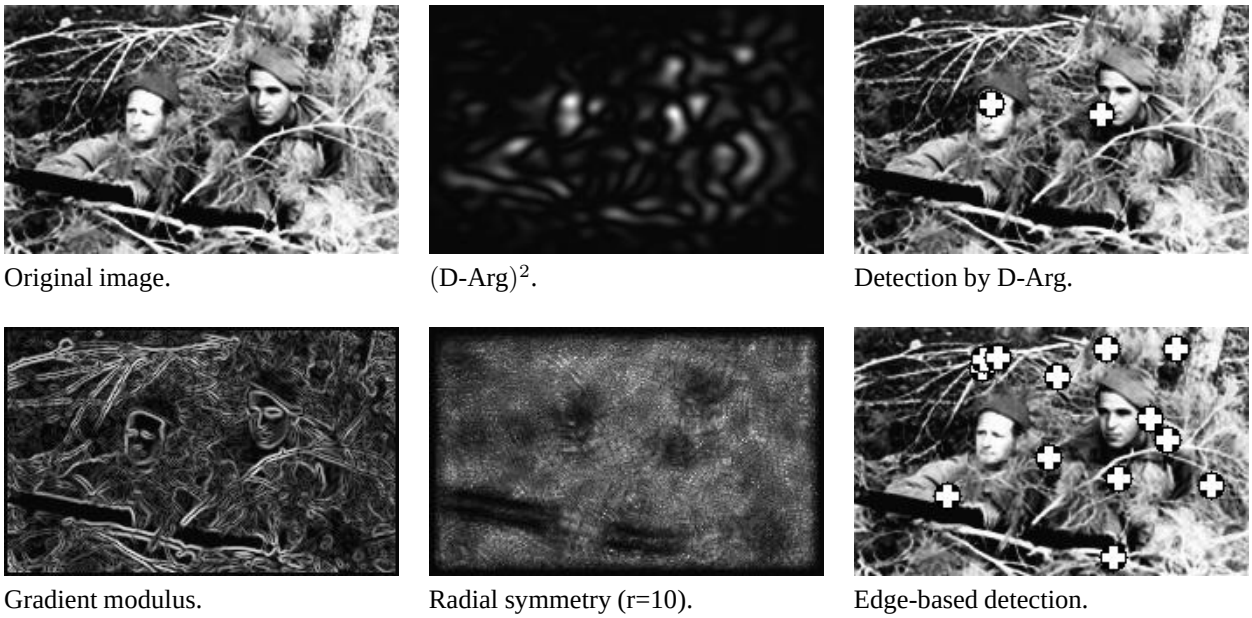
Gradient modulus.



Radial symmetry (r=10).

Edge-based detection.

Figure 6. A camouflaged hunter in the woods. The natural environment and the hunter's camouflage unite to form a uniform-looking edge map. D-Arg detects the subject based on convexity information.



Original image. $(\text{D-Arg})^2$. Detection by D-Arg.

Gradient modulus. Radial symmetry (r=10). Edge-based detection.

Figure 7. Camouflaged soldiers. The camouflage trees disguise edge-based detection schemes. Edge-based detection fails to detect any specific region of interest. Nevertheless, the soldiers faces are 3D and convex, and thus lead to their detection by D-Arg.

detection of regions emanating from smooth convex or concave three dimensional objects. The D-Arg attentional operator should be used as a complementary method to edge-based techniques, to answer those cases where edge-based schemes are doomed to fail. The suggested operator is *not* based on edge maps, and is thus free of their flaws (e.g., it is robust in dominant textures). D-Arg is proved invariant under any derivable strongly monotonically increasing transformation of the image gray-levels, which practically means robustness to illumination changes. Robustness to orientation and scale is also shown. D-Arg serves as a detector of regions of interest in real-life images. An extensive comparison between D-Arg and a typical edge based method is depicted. This comparison exhibits the advantages of D-Arg over edge-based techniques in a wide range of scenes, and suggests that it is highly efficient in many real-world applications.

References

[1] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice-Hall Inc., New Jersey, USA, 1982.

[2] B. A. Darper, S. Buluswar, A. R. Hanson, and E. M. Riseman. Information acquisition and fusion in the mobile perception laboratory. In *Proceedings of the SPIE*, volume 2059, pages 175–187, 1993.

[3] T. Lindberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11:283–318, 1993.

[4] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context free attentional operators: the generalized symmetry transform. *International Journal of Computer Vision*, 14:119–130, 1995.

[5] J. C. Russ. *The Image Processing Handbook*. CRC Press, 2nd edition, 1995.

[6] C. Schmid and R. Mohr. Combining greyvalue invariants with local constraints for object recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 872–877, 1996.

[7] P. Torr and D. Murray. Statistical detection of independent movement from a moving camera. *Image and Vision Computing*, 11:180–187, 1993.

[8] B. Vasselle, G. Giraudon, and M. Berthod. Following corners on curves and surfaces in scale space. In J. Eklundh, editor, *ECCV '94*, pages 109–114, Sweden. Springer-Verlag.

[9] Z. Xining, R. M. Haralick, and V. Ramesh. Corner detection using the map technique. In *Proceedings of the International Conference on Pattern Recognition*, volume 1, pages 549–552, Jerusalem, Israel, 1994.

[10] Y. Yeshurun. Attenional mechanisms in computer vision. In V. Cantoni, S. Levialdi, and V. Roberto, editors, *Artificial Vision*, chapter 2, pages 43–52. Academic Press Inc., 1997.

[11] Y. Yeshurun and E. Schwartz. Shape description with a space-variant sensor: Algorithm for scan-path, fusion, and convergence over multiple scans. *IEEE T-PAMI*, 11(11):1217–1222, 1989.

[12] S. Zucker, M. Langer, L. Iverson, and P. Breton. Shading flows and scenel bundles: A new approach to shape from shading. In G. Sandini, editor, *Second European Conference on Computer Vision '92*, Italy, 1992. Springer-Verlag.