FACE DETECTION AND CAMOUFLAGE BREAKING BY DIRECT CONVEXITY ESTIMATION

Ariel Tankus, Yehezkel Yeshurun, and Nathan Intrator

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

1 Introduction

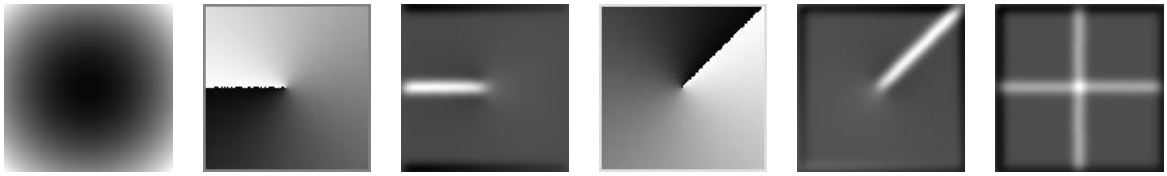
Edge detection was, so far, the core of most state of the art techniques for attentional mechanisms as well as face detection[1,2] . This excludes some recent works which utilize neural networks[3,4] , color histograms[5,6] , or shape statistics[7,8] for face detection. Though one cannot disregard their advantages, edge maps sustain severe flaws such as: sensitivity to changes in illumination, strong effect of surrounding objects, and inability to delineate objects in a cluttered scene. We overcome these problems of edge-based approaches by a novel attentional operator which detects smooth three dimensional convex or concave objects in the image. The operator is robust to face orientation, scale, and illumination, and is capable of detecting the subject in a strongly textured background. It is employed for face detection, namely to detect the eyes and hair, from which the scale of the face can be inferred. The operator answers the above problems as a whole, demands a relatively short running time, and its robustness leads to reliable results.

2 Y-Arg: Attentional Operator for Detection of Convex Regions

We next define our attentional mechanism. We refer to faces as three dimensional objects with convex and concave regions, and take advantage of this structure.

2.1 Defining the Argument of Gradient

Let us estimate the *gradient map* of image $I(x, y)$ by:
$\nabla I(x, y) \approx ([D_\sigma(x)\, G_\sigma(y)] * I(x, y), [G_\sigma(x)\, D_\sigma(y)] * I(x, y))$, where $G_\sigma(t)$ is the one dimensional Gaussian with zero mean and standard deviation σ, and $D_\sigma(t)$ is the



(a) (b) (c) (d) (e) (f)

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)** derivation of (b) in 90°. **(d)** Rotation of (b), so that the discontinuity ray is at 45° from the positive x-axis. **(e)** Derivation of (d) in 315°. **(f)** Response of D-Arg, the isotropic operator.

derivative of that Gaussian. We turn the Cartesian representation of the intensity gradient into a *polar* representation. The *argument* (also denoted "phase", and usually marked by $\theta(x,y)$), is defined by:

$$\theta(x,y) = \arg(\nabla I(x,y)) = \arctan\left(\frac{\partial}{\partial y} I(x,y) \ , \ \frac{\partial}{\partial x} I(x,y) \right)$$

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: $\arctan(t) : [-\infty, \infty] \mapsto \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. While the term "phase" is widely used in the literature[9,10] , we use it in a completely different manner: the term "phase" in this article refers to the argument of the intensity gradient.
The attentional mechanism is simply the derivative of the argument map with respect to the y-direction: $\frac{\partial}{\partial y}\theta(x,y) \approx [G_\sigma(x)\, D_\sigma(y)] * \theta(x,y)$. We denote $\frac{\partial}{\partial y}\theta(x,y)$ as *Y-Arg*.

2.2 Mathematical Formulation of Y-Arg Reaction to Paraboloids

The projection of concave and convex objects can be estimated by paraboloids, since paraboloids are arbitrarily curved surfaces[11] . 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 first order derivatives of the paraboloid are: $\frac{\partial}{\partial x} f(x,y) = 2a(x-\epsilon)$ and $\frac{\partial}{\partial y} f(x,y) = 2b(y-\eta)$. The gradient argument is therefore: $\theta(x,y) = \arctan(b(y-\eta), a(x-\epsilon))$. Deriving it with respect to y yields: $\frac{\partial}{\partial y}\theta(x,y) = \frac{ab(x-\epsilon)}{a^2(x-\epsilon)^2 + b^2(y-\eta)^2}$. However, this derivative 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 fact that for a paraboloid, $\frac{\partial}{\partial y}\theta(x,y) \to \infty$ at the negative ray of the x-axis, while continuous at the rest of the plane can be clearly seen in Fig. 2.2(c) (we define our coordinate system at the horizontal and vertical axes of the sphere).

2.3 D-Arg: The Isotropic Variant

We also define an isotropic variant of Y-Arg, whose reaction is to all axes of the paraboloid, rather than merely the negative part of the x-axis. We do so by rotating

the gradient argument by:

$$\theta_\alpha(x,y) = \begin{cases} \theta(x,y) + (\pi - \alpha), & \text{if } \theta(x,y) + (\pi - \alpha) \leq \pi \\ \theta(x,y) + (\pi - \alpha) - 2\pi, & \text{if } \theta(x,y) + (\pi - \alpha) > \pi \end{cases}$$

so the ray of discontinuity of Y-Arg is transformed to a ray from the origin forming an angle of α radians with the positive part of the x-axis. We then derive the rotated argument of gradient in the direction: $\alpha - \frac{\pi}{2}$ (or: $\alpha + \frac{\pi}{2}$), to get the response to the ray of discontinuity (see Fig. 2.2(d),(e)). Repeating this rotation with angles: $\alpha = 0°, 90°, 180°, 270°$ and summing their responses result in an isotropic operator called: *D-Arg* (see Fig. 2.2(f)). It is evident that D-Arg is more general than Y-Arg, but as we shall show, Y-Arg is effective and robust for face detection.

3 Features of Y-Arg

Two dimensional objects of constant albedo form a linear gray-level function, and are usually of no interest (for example, walls). It can be easily shown, that the Y-Arg attentional mechanism has zero response to planar objects. In addition, it can be shown that the response of Y-Arg to edges of planar objects is finite, and is therefore smaller than its response to paraboloids. Another provable characteristic of Y-Arg is its linear dependence on scale. We now present several invariants of Y-Arg, followed by a practical discussion and demonstration from real-life scenes.

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. Denote 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:

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

We denote $f^0 = f(x_0,y_0)$, $f^0_x = f_x(x_0,y_0)$, $f^0_y = f_y(x_0,y_0)$. The argument of 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:

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

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

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

Original image: $I(x,y)$ $\log(\log(\log(I)))$ $\exp(\exp(\exp(I))$

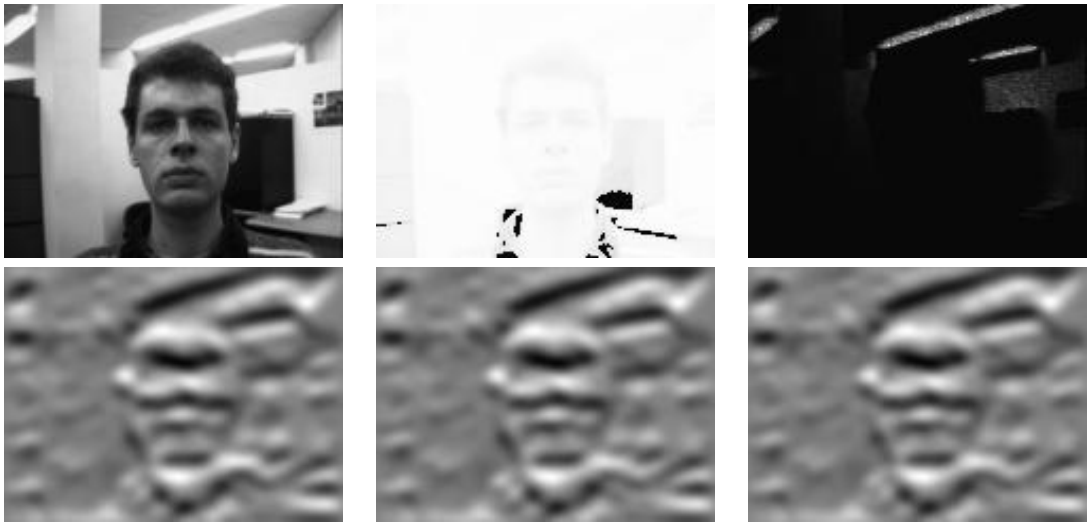


Figure 2: Top row: The original image $\mathrm{I}(x,y)$ is compared to $\log(\log(\log(\mathrm{I}(x,y))))$ and $\exp(\exp(\exp(\mathrm{I}(x,y))))$. Y-Arg is invariant under log and exp. **Bottom row:** Y-Arg. Similarity among Y-Arg of original image and Y-Args of transformed images is obvious.

□

Let us rephrase theorem 1 in the following manner:

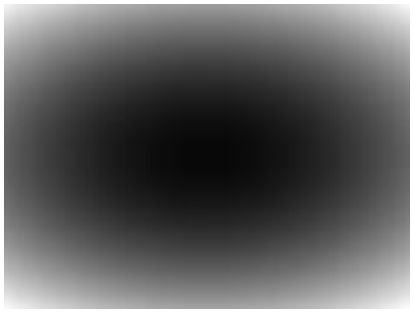
> *Y-Arg is invariant under any derivable monotonically increasing (in the strong sense) transformation of the gray-level function.*

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. Y-Arg is also invariant under linear combinations (with positive coefficients) and compositions of these functions, since such combinations are also derivable and strongly monotonically increasing. The functions mentioned above and their combinations are common in image processing for lighting improvement. This implies that Y-Arg is invariant under a large variety of lighting conditions. Figure 3 demonstrates Y-Arg invariance to $\log(\log(\log(z)))$ and $\exp(\exp(\exp(z)))$ in a real-life scene.
In view of Y-Arg invariants, the suggested model is not only a paraboloidal gray-levels detector, but also a detector of any derivable (strongly) monotonically increasing transformation of paraboloids. This makes Y-Arg particularly attractive for usage in various scenes in which the environment is unknown before hand.

4 Face Detection Using Y-Arg

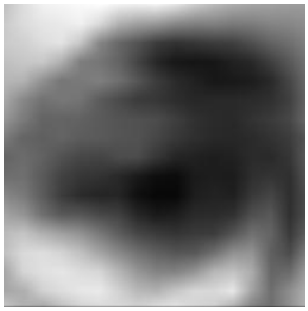
4.1 Approximation by Paraboloids

One of the underlying ideas of the theoretical model is the estimation of the gray-levels describing convex and concave objects, in our case—the eyes and hair, using paraboloids or a derivable monotonically increasing transformation (in the strong sense) of paraboloids. Figure 4.1 shows such a synthetic paraboloid along with a magnified eye. The eye gray-levels are similar to those of a paraboloid. The Y-Arg

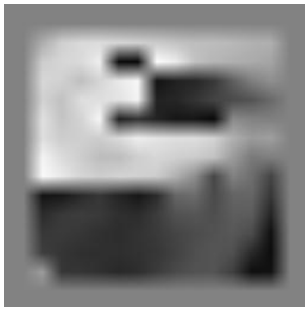


synthetic paraboloid.

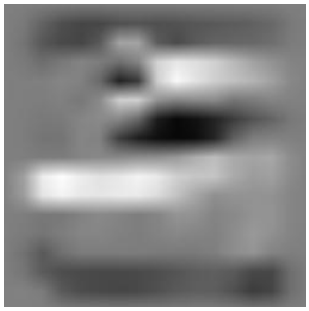
original image.



right eye (zoom).



gradient argument.



Y-Arg.

Figure 3: The eye exhibits strong similarity to the artificial paraboloidal gray-levels: $I(x, y) = 10x^2 + 30y^2$. The gradient argument of the eye is similar to that in Fig. 2.2. A clear response of Y-Arg at the negative part of the x-axis is observed.

of the eye strongly reacts to the x-axis; this behavior resembles that of Y-Arg of paraboloids.
Figures 4.1, 4.1,4.1, demonstrates the robustness to three factors: illumination direction, scale, and orientation of the head, respectively. Mirrored auto-correlation serves to detect the face, i.e., choose the window with the best cross correlation between left and right halves (mirroring one of them) among all possible window positions; the window is of the same height as the image. Y-Arg robustness to illumination, scale, and orientation is mainly due to the fact that $\frac{\partial}{\partial y}\theta(x, y) \to \infty$ for paraboloids, which is a very stable feature. The areas of strong Y-Arg response enable the heuristic detection of the face scale.

Figure 4: Robustness to lighting. Illumination comes from a single point light source. Each row relates to the corresponding azimuth: 90°, 60°, 30°, 0°, −30°, −60°, −90°. Detection by mirrored auto-correlation is marked.

Figure 5: Robustness to scale. Several scales of the face. The Y-Arg image strongly reacts to the eyes and hair regions, regardless of the scale. Largest face is about 6 times larger than smallest face. Detection by mirrored auto-correlation is marked.

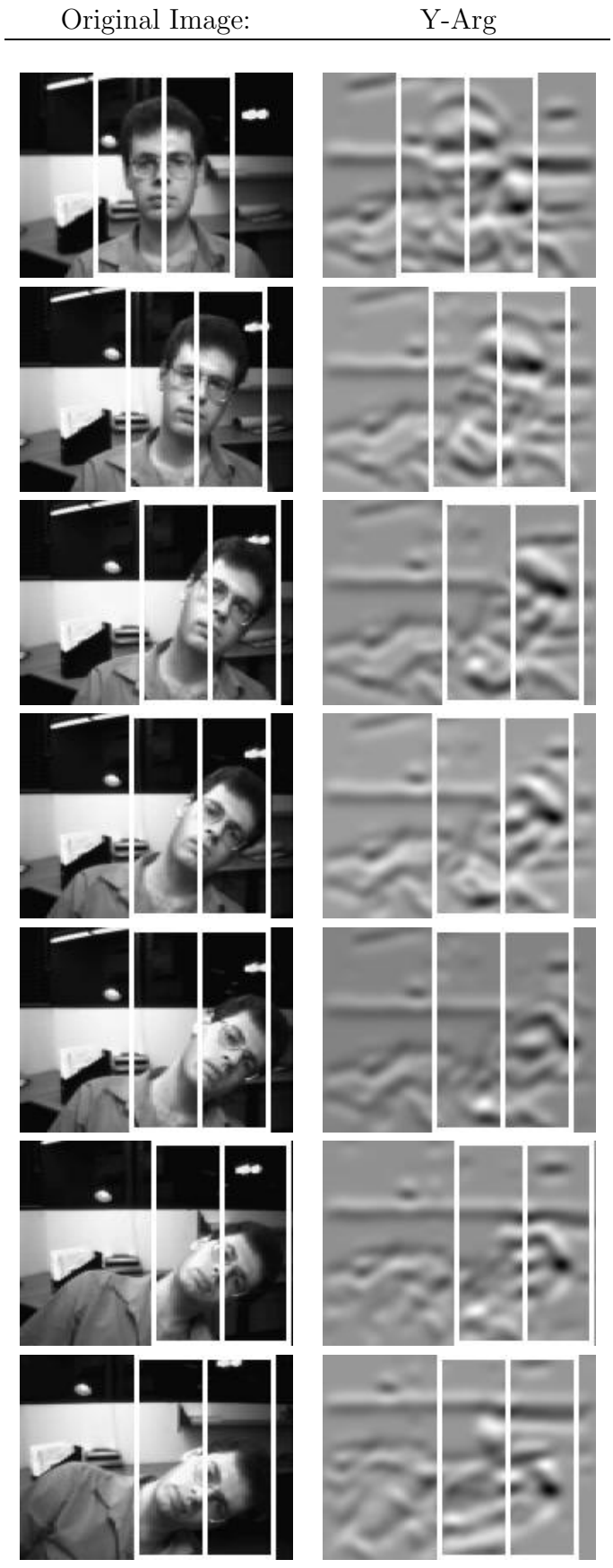


Figure 6: Robustness to orientation. Capability to detect oblique faces. Y-Arg strongly reacts to the eyes and hair, even though the face is slanted. Detection by mirrored auto-correlation is marked.

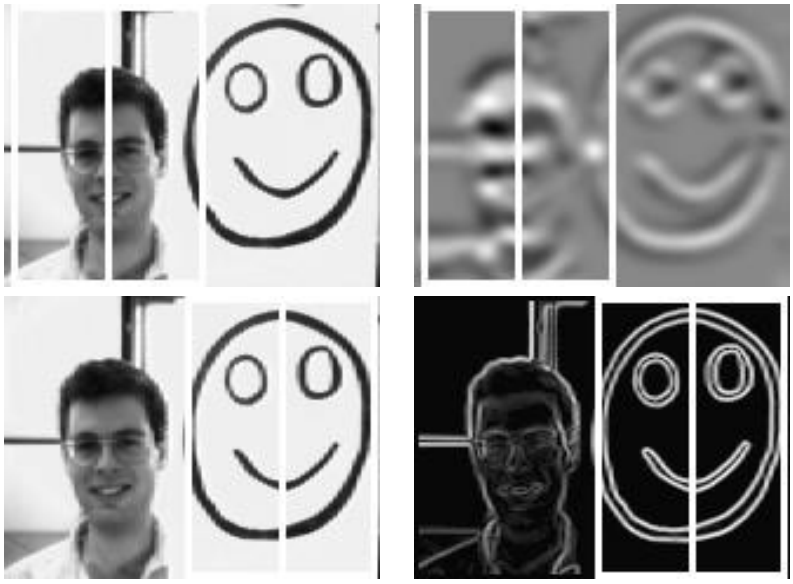


Figure 7: A face vs. smiley. Edge-based methods locate the larger object, which is the flat smiley. Y-Arg detects the three dimensional face, although it is smaller.

4.2 Superiority of Y-Arg on Edge Detection

In this section, we briefly delineate the results of an extensive comparison between Y-Arg and the edge map (taken as gradient modulus). Following the operation of each method, mirrored auto-correlation attempts to detect the face.

1. *Reaction to 3D Objects*: Y-Arg detects 3D objects, so the Y-Arg of a smiley (2D object) is relatively low, as opposed to edge methods (Fig. 4.2).

2. *Insensitivity to Strong Edges*: Sharp color changes are likely to appear due to different object colors (or albedo) (Fig. 4.2). These variations lead to strong edges, which distract edge-based methods from the subject. Y-Arg does not react strongly to sharp changes, but rather, to gradual changes of intensity of the kind exhibited by the eyes and hair[12] .

3. *Robustness to Lighting*: In Fig. 4.2(a), the background is better lit than the subject. Y-Arg detects the subject, while edge methods, the background. The improvement of the image by the logarithm (Fig. 4.2(b)) makes edge-based methods too detect the subject. Note, that the decision to apply this specific (log) function was made by a human. Y-Arg robustness to illumination releases the automatic face detector from the need to decide which illumination it is facing.

4. *Stability in Textured Background*: The existence of texture in an image makes the task of discriminating the subject from the background very hard. The difficulties emanate from the large amount of edges covering a substantial image area, and the periodicity of the (usually symmetric) pattern composing the texture. As Fig. 4.2(e) shows, Y-Arg is much more robust than edge based methods, and is capable of separating the face from dominant textures.

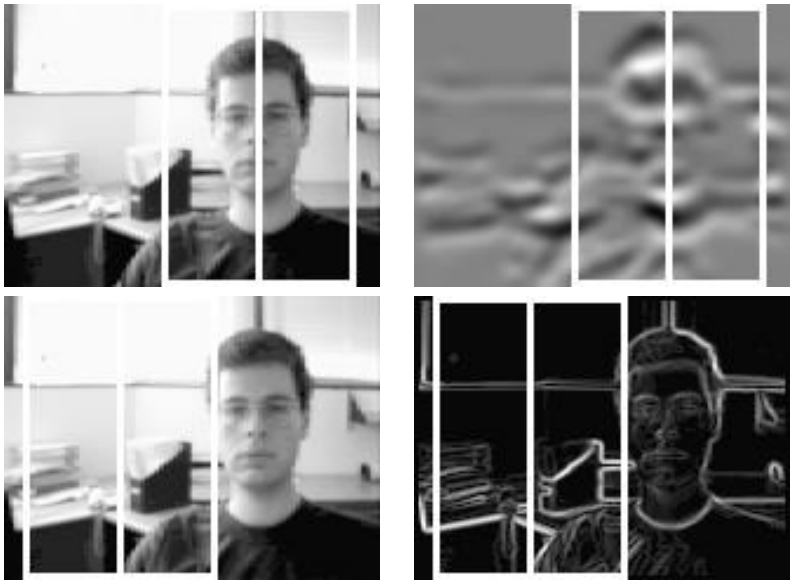
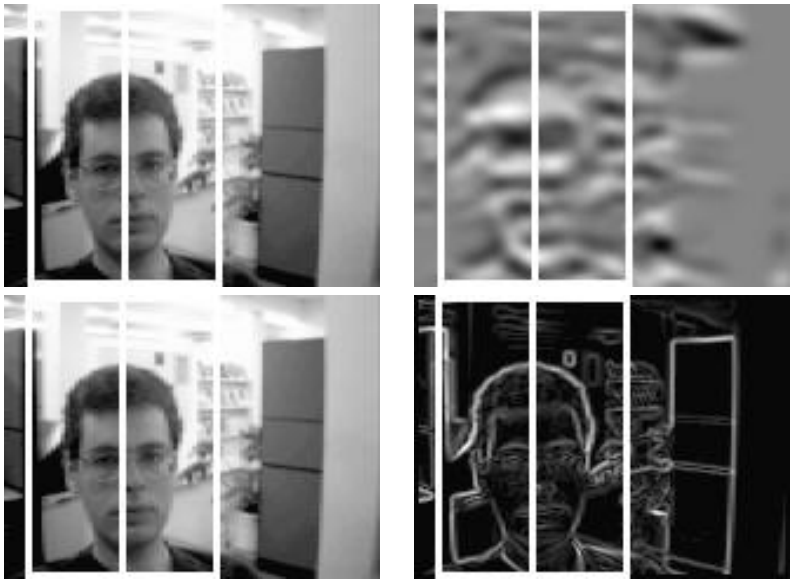


Figure 8: Small objects with strong edges divert edge-based methods from the real subject. Y-Arg reacts to gradual variations rather than to sharp color changes.

(a)



(b)

Figure 9: **(a)** The background is better lit than the face. Y-Arg is capable of detecting the face despite the poor illumination. **(b)** Illumination improved by applying logarithm. Edge methods as well as Y-Arg now detect the subject.

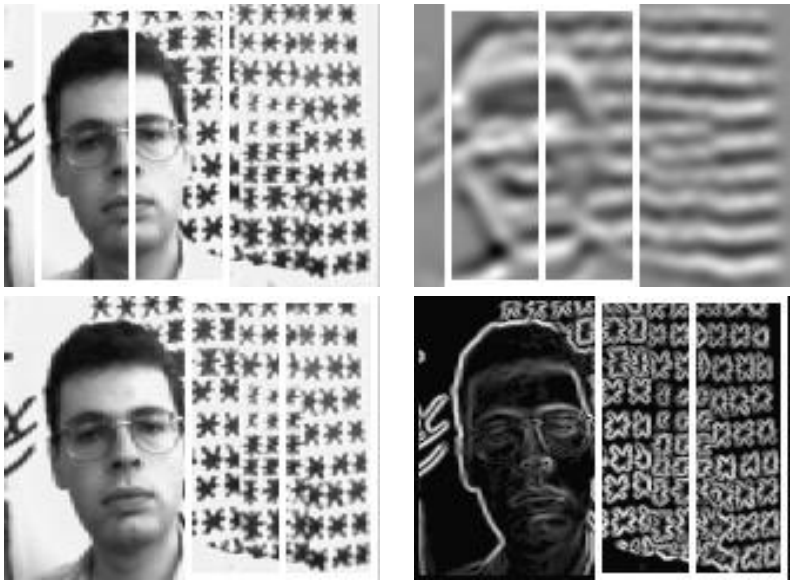


Figure 10: A texture of asterisks. Facial edges look negligible near texture edges. In the Y-Arg map, facial regions attain higher values than textural areas.

Figure 11: A squirrel on the background of leaves covering the ground, under the shades of a tree. A human viewer is usually disguised by the camouflage.

5 Camouflage breaking

The stability of D-Arg under various conditions (illumination, scale, orientation, texture) makes it suitable for camouflage breaking. This section demonstrates one camouflage example out of several difficult scenes where D-Arg succeeded in breaking camouflage.

The camouflage of many animals is based on mimicking the environment. Both the color and the texture of fur of the camouflaged animal fit its habitat. Figure 5 is an example of animal camouflage. In this image, the squirrel stands on a ground covered with leaves, under the shades of a tree. Figure 5 shows the response of an edge-based attentional operator, the radial symmetry transform[2] to Fig. 5. The strong camouflage yields a strong response to a large image area, making it impossible to detect the quirrel from the symmetry map. No single image region can be isolated using the symmetry transform in this case. D-Arg, on the other hand, exploits the convexity of the squirrel (especially - its uniformly colored belly) to detect the squirrel. D-Arg has two strong peaks, both correspond to the squirrel.

6 Conclusions

We introduce a novel attentional operator (Y-Arg) for detection of regions emanating from smooth convex or concave three dimensional objects. We use it to detect the eyes and hair, and thus, the face. Y-Arg is proved invariant under any derivable (strongly)

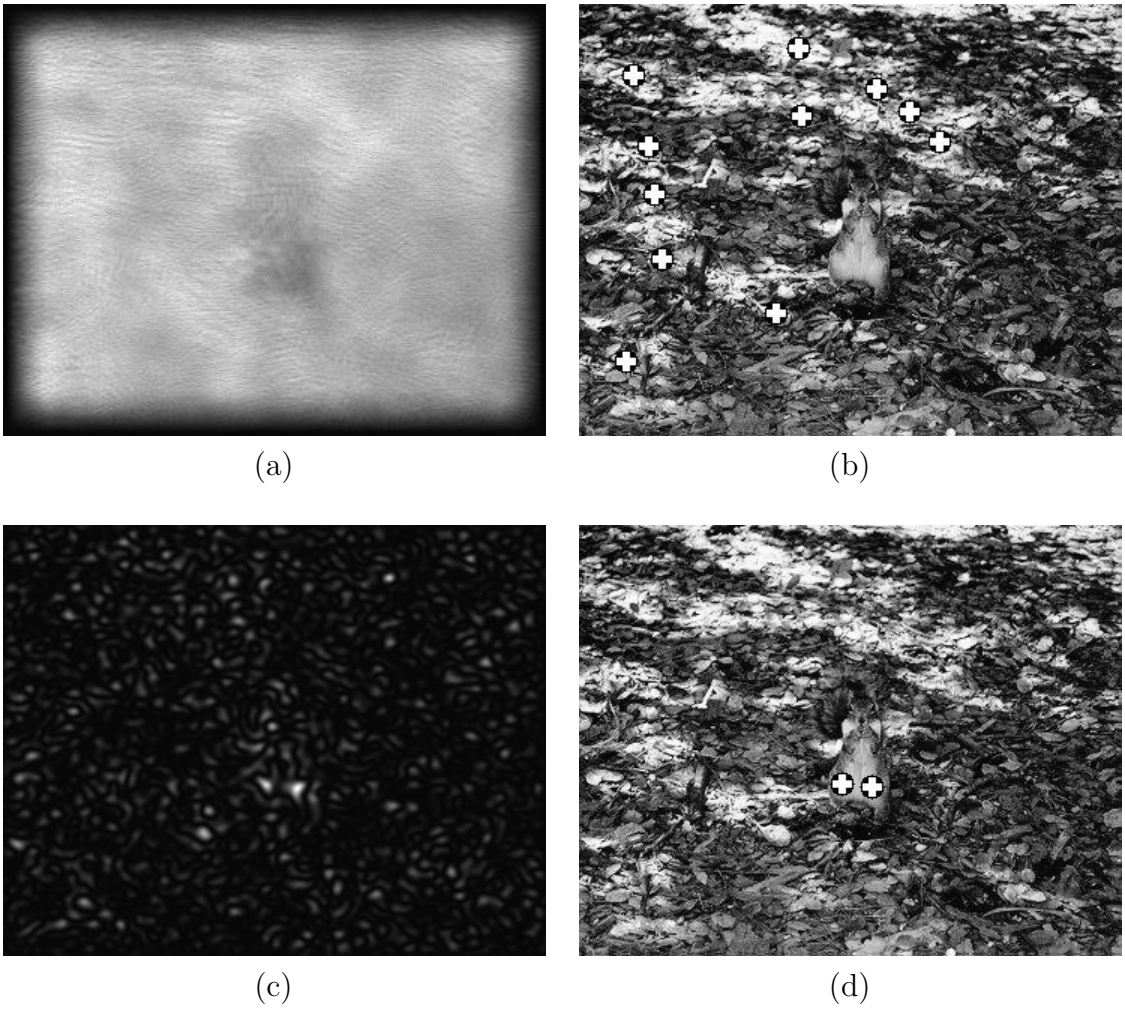


Figure 12: Detection of the squirrel of Fig. 5. (a) Radial symmetry map (radius=30). (b) Detection by redial symmetry (90% threshold). (c) D-Arg^2. (d) Detection by D-Arg (only 70% threshold are required for detection).

monotonically increasing transformation of the image gray-levels, which practically means robustness to illumination changes. Robustness to orientation and scale is also described. The operator is *not* based on edge maps, and thus free of their flaws (e.g., Y-Arg is robust in dominant textures). An extensive comparison with edge-based methods is depicted. Finally, we have demonstrated the usage of D-Arg (an isotropic variant of Y-Arg) for camouflage breaking in highly textured environments.

References

[1] Arnaud Jacquin and Alexandros Eleftheriadis. Automatic location tracking of faces and facial features in video sequences. In Bichsel,[13] pages 142–147.

[2] Daniel Reisfeld, Haim Wolfson, and Yehezkel Yeshurun. Context free attentional operators: the generalized symmetry transform. *International Journal of Computer Vision*, pages 119–130, 1995.

[3] K-K. Sung and Tomaso Poggio. Example-based learning for view-based human face detection. In *Proceedings of Image Understanding Workshop*, volume II, pages 843–850, Monterey, Canada, November 1994.

[4] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade. Human face detection in visual scenes. will appear in: Advances in Neural Information Processing Systems 8.

[5] Ying Dai and Yasuaki Nakano. Extraction of facial images from complex background using color information and SGLD matrices. In Bichsel,[13] pages 238–242.

[6] Bernt Schiele and Alex Waibel. Gaze tracking based on face-color. In Bichsel,[13] pages 344–349.

[7] M.C. Burl, T.K. Leung, and P. Perona. Face localization via shape statistics. In Bichsel,[13] pages 154–159.

[8] Baback Moghaddam and Alex Pentland. Maximum likelihood detection of face and hands. In Bichsel,[13] pages 122–128.

[9] S. Fischer and Joseph Bigün. Texture boundary tracking with gabor phase. In Gunilla Borgefors, editor, *Proceedings of the 9th Scandinavian Conference on Image Analysis*, pages 877–884, Uppsala, Sweden, 1995.

[10] David J. Fleet and Allan D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, pages 77–104, 1990.

[11] S.W. Zucker, M.S. Langer, L.A. 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*, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.

[12] Hans Peter Graf, Tsuhan Chen, Eric Petajan, and Eric Cosatto. Locating faces and facial parts. In Bichsel,[13] pages 41–46.

[13] Martin Bichsel, editor. *Proceedings of the First International Workshop on Automatic Face- and Gesture-Recognition*, Zurich, Switzerland, June 1995.