Face Detection by Direct Convexity Estimation

Ariel Tankus, Hezy Yeshurun and Nathan Intrator

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

Abstract. We suggest a novel attentional mechanism for detection of smooth convex and concave objects based on direct processing of intensity values. The operator detects the regions of the eyes and hair in a facial image, and thus allows us to infer the face location and scale. Our operator is robust to variations in illumination, scale, and face orientation. Invariance to a large family of functions, serving for lighting improvement in images, is proved. An extensive comparison with edge-based methods is delineated.

1 Introduction

Edge detection was, so far, the core of most state of the art techniques for attentional mechanisms as well as face detection (see [4], [6]). This excludes some recent works which utilize neural networks ([9],[7]), color histograms ([2],[8]), or shape statistics ([1],[5]) 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-Phase: 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 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: $\arctan(t) : [-\infty,\infty] \mapsto \left[-\frac{\pi}{2},\frac{\pi}{2}\right]$. 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-Phase.*

2.2 Mathematical Formulation of Y-Phase Reaction to Paraboloids

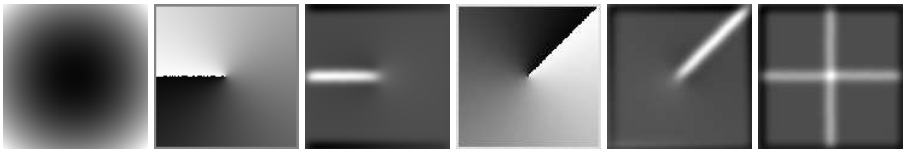
The projection of concave and convex objects can be estimated by paraboloids, since paraboloids are arbitrarily curved surfaces (see [10]). Our mathematical formulation refers to a general paraboloid of the form: $f(x,y) = a(x-\epsilon)^2 + b(y-\eta)^2$, where $a > 0$, $b > 0$ are constants, and (ϵ,η) 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. 1(c) (we define our coordinate system at the horizontal and vertical axes of the sphere).

2.3 D-Phase: The Isotropic Variant

We also define an isotropic variant of Y-Phase, 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-Phase 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



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

Fig. 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-Phase, the isotropic operator.

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. 1(d),(e)). Repeating this rotation with angles: $\alpha = 0°, 90°, 180°, 270°$ and summing their responses result in an isotropic operator called: *D-Phase* (see Fig. 1(f)). It is evident that D-Phase is more general than Y-Phase, but as we shall show, Y-Phase is effective and robust for face detection.

3 Features of Y-Phase

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-Phase attentional mechanism has zero response to planar objects. In addition, it can be shown that the response of Y-Phase to edges of planar objects is finite, and is therefore smaller than its response to paraboloids. Another provable characteristic of Y-Phase is its linear dependence on scale. We now present several invariants of Y-Phase, 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:

$$g_x(x_0, y_0) = T'(f(x_0, y_0))f_x(x_0, y_0)$$
$$g_y(x_0, y_0) = T'(f(x_0, y_0))f_y(x_0, y_0)$$

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

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

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

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

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:

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

□

Let us rephrase theorem 1 in the following manner:

> *Y-Phase 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-Phase is invariant, for example, under linear transformations, positive powers (where $f(x, y) > 0$), logarithm, and exponent. Y-Phase 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-Phase is invariant under a large variety of lighting conditions. Figure 2 demonstrates Y-Phase invariance to $\log(\log(\log(z)))$ and $\exp(\exp(\exp(z)))$ in a real-life scene.

In view of Y-Phase 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-Phase particularly attractive for usage in various scenes in which the environment is unknown before hand.

4 Face Detection Using Y-Phase

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 3 shows such a synthetic paraboloid along with a magnified eye. The eye gray-levels are similar to those of a paraboloid.

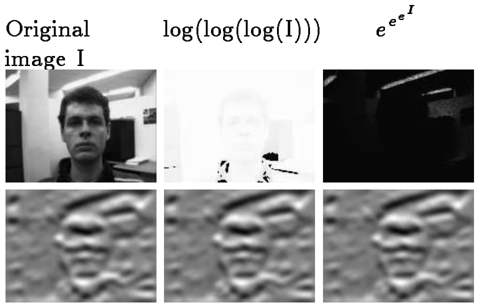


Fig. 2. Top row: The original image $\mathrm{I}(x, y)$ is compared to $\log(\log(\log(\mathrm{I}(x, y))))$ and $\exp(\exp(\exp(\mathrm{I}(x, y))))$. Y-Phase is invariant under log and exp. **Bottom row:** Y-Phase. Similarity among Y-Phase of original image and Y-Phases of transformed images is obvious.

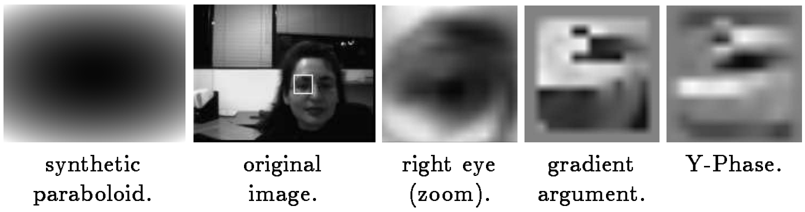


Fig. 3. The eye exhibits strong similarity to the artificial paraboloidal gray-levels: $I(x, y) = 10x^2 + 30y^2$. The gradient argument of the eye is similar to that in Fig. 1. A clear response of Y-Phase at the negative part of the x-axis is observed.

The Y-Phase of the eye strongly reacts to the x-axis; this behavior resembles that of Y-Phase of paraboloids.

Figure 4 demonstrates the robustness to three factors: illumination direction, scale, and orientation of the head. 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-Phase robustness to illumination, scale, and orientation is mainly due to the fact that $\frac{\partial}{\partial y}\theta(x, y) \rightarrow \infty$ for paraboloids, which is a very stable feature. The areas of strong Y-Phase response enable the heuristic detection of the face scale.

4.2 Superiority of Y-Phase on Edge Detection

In this section, we briefly delineate the results of an extensive comparison between Y-Phase 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-Phase detects 3D objects, so the Y-Phase of a smiley (2D object) is relatively low, as opposed to edge methods (Fig. 5(a)).

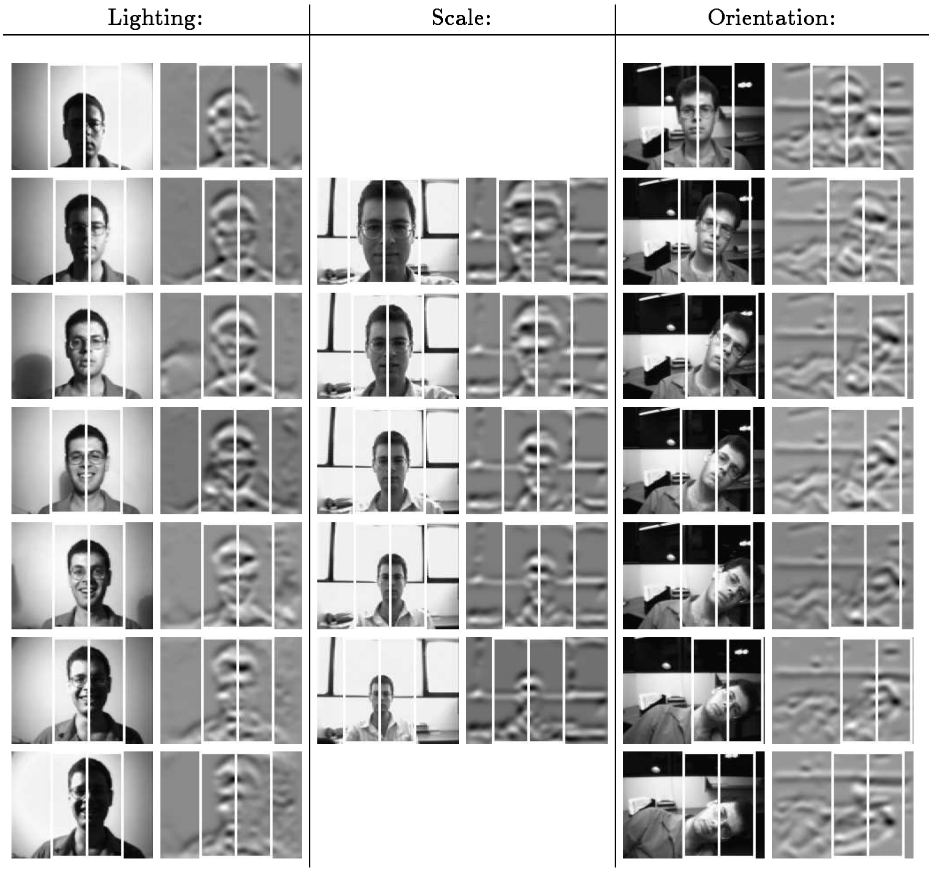


Fig. 4. Robustness to lighting, scale, and orientation. Each column contains the original image (left) and its Y-Phase (right); detection by mirrored auto-correlation marked on all images. **Lighting:** Illumination comes from a single point light source. Each row relates to the corresponding azimuth: $90°$, $60°$, $30°$, $0°$, $-30°$, $-60°$, $-90°$. **Scale:** Several scales of the face. The Y-Phase image strongly reacts to the eyes and hair regions, regardless of the scale. Largest face is about 6 times larger than smallest face. **Orientation:** Capability to detect oblique faces. Y-Phase strongly reacts to the eyes and hair, even though the face is slanted.

2. *Insensitivity to Strong Edges*: Sharp color changes are likely to appear due to different object colors (or albedo) (Fig. 5(b)). These variations lead to strong edges, which distract edge-based methods from the subject. Y-Phase does not react strongly to sharp changes, but rather, to gradual changes of intensity of the kind exhibited by the eyes and hair. (see also [3]).
3. *Robustness to Lighting*: In Fig. 5(c), the background is better lit than the subject. Y-Phase detects the subject, while edge methods, the background. The improvement of the image by the logarithm (Fig. 5(d)) makes edge-

based methods too detect the subject. Note, that the decision to apply this specific (log) function was made by a human. Y-Phase 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. 5(e) shows, Y-Phase is much more robust than edge based methods, and is capable of separating the face from dominant textures.

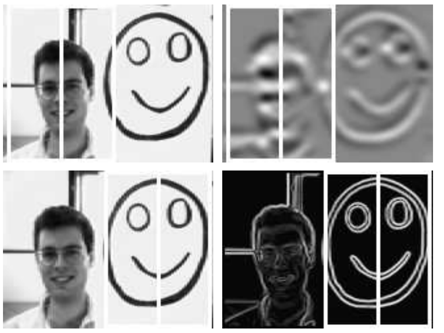
5 Conclusions

We introduce a novel attentional operator (Y-Phase) 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-Phase 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 described. The operator is *not* based on edge maps, and thus free of their flaws (e.g., Y-Phase is robust in dominant textures). An extensive comparison with edge-based methods is depicted.

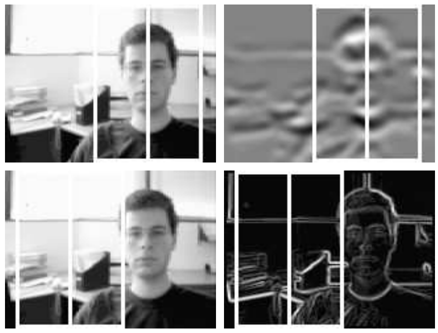
References

1. M.C. Burl, T.K. Leung, P. Perona. *Face Localization via Shape Statistics*, Proc. of the Intl. Workshop on Auto. Face- and Gesture-Recognition, Zurich, 1995.
2. Y. Dai, Y. Nakano. *Extraction of Facial Images from Complex Background Using Color Information and SGLD Matrices*, Proc. of the Intl. Workshop on Auto. Face- and Gesture-Recognition, Zurich, 1995.
3. H.P. Graf, T. Chen, E. Petajan, E. Cosatto. *Locating Faces and Facial Parts*, Proc. of the Intl. Workshop on Auto. Face- and Gesture-Recognition, Zurich, 1995.
4. A. Jacquin, A. Eleftheriadis. *Automatic Location Tracking of Faces and Facial Features in Video Sequences*, Proc. of the Intl. Workshop on Auto. Face- and Gesture-Recognition, Zurich, 1995.
5. B. Moghaddam, A. Pentland. *Maximum Likelihood Detection of Face and Hands*, Proc. of the Intl. Workshop on Auto. Face- and Gesture-Recognition, Zurich, 1995.
6. D. Reisfeld, H. Wolfson, Y. Yeshurun. *Context Free Operators: The Generalized Symmetry Transform*, Intl. Journal of Computer Vision, Vol. 14, 1995.
7. H.A. Rowley, S. Baluja, T. Kanade. *Human Face Detection in Visual Scenes*, will appear in: Advances in Neural Information Processing Systems 8.
8. B. Schiele, A. Waibel. *Gaze Tracking Based on Face-Color*, Proc. of the Intl. Workshop on Auto. Face- and Gesture-Recognition, Zurich, 1995.
9. K-K. Sung, T. Poggio. *Example-Based Learning for View-Based Human Face Detection*, in Proc. from Image Understanding Workshop, Monterey, CA, Nov. 1994.
10. S.W. Zucker, M.S. Langer, L.A. Iverson, P. Breton. *Shading Flows and Scenel Bundles: A New Approach to Shape from Shading*, ECCV '92, Springer-Verlag.

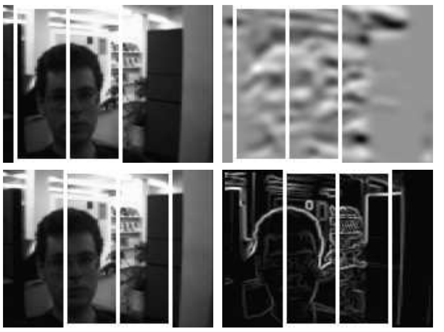
This article was processed using the LaTeX macro package with LLNCS style



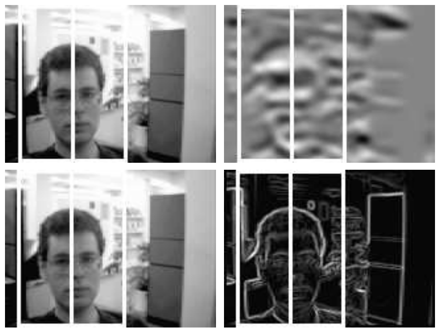
(a) A face vs. smiley. Edge-based methods locate the larger object, which is the flat smiley. Y-Phase detects the three dimensional face, although it is smaller.



(b) Small objects with strong edges divert edge-based methods from the real subject. Y-Phase reacts to gradual variations rather than to sharp color changes.



(c) The background is better lit than the face. Y-Phase is capable of detecting the face despite the poor illumination.



(d) Illumination improved by applying logarithm. Edge methods as well as Y-Phase now detect the subject.

(e) A texture of asterisks. Facial edges look negligible near texture edges. In the Y-Phase map, facial regions attain higher values than textural areas.

Fig. 5. Comparison between Y-Phase and edge-based methods. Y-Phase is more robust, and performs better than edge-based methods.