Approach to Invariant Object Detection

Teodor Mandziy

Department of Computational Methods and Systems for Information Transformation
Physico-Mechanical Institute, Lviv
teodor_mandziy@ipm.lviv.ua

Abstract

A new approach to invariant object detection is proposed. A general case solution for invariance to required set of transformation is considered on the task of image object detection.

1. Introduction

Up to date there are many approaches to image object detection and pattern recognition tasks. Majority of those approaches aims to solve the problem for a certain subset of objects and images.

There exist many global object detection methods that are able efficiently detect objects on input images [1]. Problems arise when some sort of transformation or distortion is introduced to the target object in input image. In a presence of such transformations as scale, rotation, projection, shape variations etc the problem rises to a new level of algorithmical and computational complexity. Presence of such transformations makes conventional methods more than useless. Existing adaptations of those methods to such transformations [1] are usually restricted to some small subset of them. On the other hand there exist methods that are able to efficiently model object appearance under some transformations on input image given the approximal initial position of that object in input image [2,3,4]. Those two approaches to object detection and object modeling exist and develop separately.

Absence of some kind of "holistic" approach to object detection is a concern of this paper. It presents an attempt to combine existing object detection and object modeling techniques to produce a new and to a degree general approach to invariant object detection.

2. Invariant object detection

In [5] was proposed an approach to contour based object detection invariant to shape variation of target object. Under shape variation it is understood the possible plausible variations of the target object shape that can be reproduced by its mathematical model. The main idea of the method consisted in efficient correlational picture sum generation of an input image with all possible target object shape variations. The practical applicability of the approach is valid only for a certain class of objects that satisfy introduced in [5] smoothness assumption. The smoothness assumption states that small changes in shape (between object model and input image) should cause small changes of correlational peaks for all possible target object shape variations. Practical value of this method exists only for the objects with relatively small overall variance of shape changes and number of obstacles on input image. It also shows low tolerance for scale and rotation changes. The main reason for aforementioned restrictions is accumulation of correlational noise to the level of useful correlational signal. So naturally to overcome these drawbacks the way of amplifying useful correlational signal in overall sum is required.

2.1. Direct MAX computation

Without loss of generality let us consider proposed invariant object detection approach in terms of image object detection. The goal for invariant image object detection is to detect a target object on an input image I invariant to certain set of possible transformations of target object on I. Also without loss of generality for the set of transformation choose target object appearance changes, scale s and rotation φ. Let $M(b,s,\varphi)$ be a mathematical model of target object with some parameter vector b responsible for appearance changes (for instance shape, texture, illumination etc), scale s and rotation φ. Let $C(I,M(b,s,\varphi),x,y)$ be some similarity measure that measures similarity of an input image $I(x,y)$ with target object model $M(b,s,\varphi)$ at (x,y).

The objective is to detect object of interest $M(b,s,\varphi)$ with arbitrary parameter vector b on arbitrary input image $I(x,y)$ regardless to affine transformation (scale s and rotation φ) of target object on input image.

So basically described task of invariant object detection in general can be represented as following:

$$C^{inv}(x,y) = \max_{b,s,\varphi} \{C(I,M(b),s,\varphi,x,y)\}, \qquad (1)$$

where $C^{inv}(x,y)$ is some invariant similarity measure, $C(I,M(b,s,\varphi),x,y)$ is a similarity measure sensitive to affine transformations (s,φ) and appearance changes b. In fact in form of (1) can be represented any object detection algorithm (with correction on a set of transformation). The difference is in the way a particular algorithm solves $max\{C(I,M)\}$. In general case this task falls into optimization theory where $max\{C(I,M)\}$ is formulated in terms of some conventional optimization technique (lest squares, dynamic programming, gradient based methods etc). $C(I,M(b,s,\varphi),x,y)$ is a complex function of many variables and local minimums. Thus optimization of (1) is complex and generally unsolvable task.

But there is a way to represent solution for (1) not as optimizational but as strictly computational task. To do that one should substitute co called maximum norm for $max\{C(I,M)\}$ in (1). Analytical representation of maximum norm for integrable function f is the following:

$$\|f\|_n = \lim_{n\to\infty}\left(\int_D f^n d\mu\right)^{\frac{1}{n}}, \tag{2}$$

Substitution of (1) into (2) gives the following:

$$C^{inv} = \lim_{n\to\infty}\left(\iiint_D C(I, M(b,s,\varphi), x, y)^n \, dbdsd\varphi\right)^{\frac{1}{n}}, \tag{3}$$

So basically such problem formulation brings object detection task down to “simple” integration of similarity measure $C(I,M(b,s,\varphi),x,y)$ over a set of model parameters b and affine transform parameters s and φ.

2.2. Practical difficalties

Even though theoretically (3) can be used for general object detection and recognition tasks, it is crucial for practical reasons to build proper analytical model $M(b,s,\varphi)$ and similarity measure $C(I,M(b,s,\varphi),x,y)$. The main purpose of that is practical integrability of (3) and simplicity of final result.

One of the biggest drawbacks of that approach is computational complexity. There are few reasons for that. The first reason is representation of input image and model of target object. Generally speaking representation of input image I and target object model $M(b,s,\varphi)$ would be in a form of superposition of their parts. Thus computation complexity will grow polynomially with the growth of parameter n in (3).

In practice n does not go to infinity. n is chosen depending on type of object of interest and input image $I(x,y)$ content, to be sufficiently large enough to separate useful correlational peaks from noise ones. The second reason is integration of n -th power of $C(I,M(b,s,\varphi),x,y)$. Depending on the type of object and chosen transformation set this function definitely would be multivariable. Commonly such function would require numerical integration at least over part of variables. And under given conditions numerical integration of functions of many variables is computationally heavy task considering high number of such integration operations required.

2.3. Basic experimental results

In this section basic results obtained for described above approach are demonstrated. To make computation as simple as possible triangle was chosen as target object. Contour image of triangle was modeled by ASM [2,5] with one-dimensional parameter vector b (figure 1). For similarity measure correlation measure was chosen. Computational results are sown on figure 2. The difference between second ($n=1$) and third ($n=3$) columns of figure 2 shows that proposed approach allows to significantly amplify useful correlational signal. In result it is possible more accurately locate the object on an input image.

3. Conclusions

As one can see from presented experimental results the proposed approach can be successfully used for object

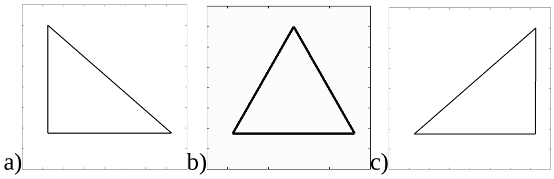

Figure 1: ASM generated triangle shape samples: a) $b=-1$; b) $b=0$; c) $b=1$.

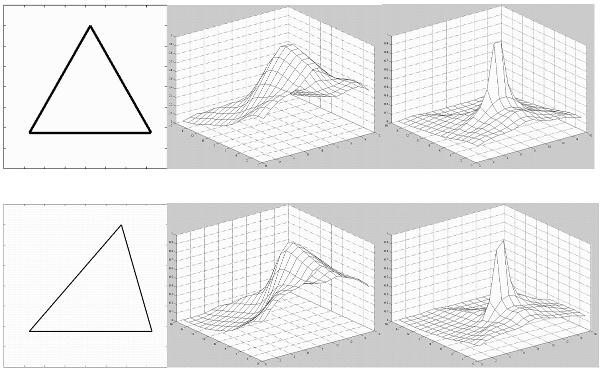

Figure 2: Triangle detection results: first column shows input images (top corresponds to object with $b=0$, bottom- $b=0.5$); second and third columns show shape invariant C^{inv} ($n=1$ and $n=3$ consequently).

detection invariant to certain set of transformations. Even though shown results show the invariance only to shape variations the approach is general and theoretically can be used for any type of transformation. Generality of the approach allows user to choose suitable similarity measure and means for object modeling. The cost for simplicity and generality of this technique is its computational complexity.

4. References

[1] Brunelli R., “Template Matching Techniques in Computer Vision: Theory and Practice”. – Wiley, 2009.

[2] Cootes T. F., Taylor C. J., Cooper D. H. and Graham J., Active shape models - their training and application // *Computer Vision and Image Understanding*.- Jan. 1995.- 61(1).-P.38–59.

[3] *Cootes T.F., Edwards G.J. and Taylor C.J.,* Active Appearance Models // *Proc. Fifth European Conf. Computer Vision*, H. Burkhardt and B. Neumann,eds.- 1998.-vol. 2.- P. 484-498.

[4] *Blanz V., Vetter T.*, A Morphable Model for the Synthesis of 3D Faces // *SIGGRAPH'99 Conference Proceedings*.- 1999.-P. 187-194

[5] Mandziy T. "Model-Based Correlational Object Detection", *The Tenth All-Ukrainian International Conference on Signal/Image Processing and Pattern Recognition UkrObraz’ 2010, 2010, p.* 133-136.