Model-Based Correlational Object Detection

Teodor Mandziy

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

Abstract

New correlation-based approach for object detection is proposed. Method for varying shape object detection is developed. Promising results were obtained by the proposed method on synthetic images.

1. Introduction

Object detection is one of the hardest problems in computer vision. It is virtually impossible to extract superior in general approach from variety of existing up to date object detection methods. This is caused by the complexity of the given task. Any particular approach is developed and can be considered as superior only for a certain class of objects. Existing object detection methods can be divided on two classes. Those are feature- and template-based techniques. Feature-based methods represent an object image by a set of features and corresponding them spatial relations, thus they neglect certain amount of information about an object. On the other hand, template-based methods use complete image of an object, but majority of those methods are unable to deal effectively enough with variations in shape or texture, otherwise it becomes very computationally expensive.

This paper deals with template-based paradigm. It considers template to be dynamical model of given object image and gives computationally efficient solution to the object detection by matching with such a dynamical template.

2. Dynamical template matching

This section formulates the problem of template-based object detection. It also reveals existing problems of this approach connected with stationarity of template image. There is proposed an approach for efficient correlational object detection, which is based on dynamical template matching. Under dynamical template we understand an object image template that is able to change its shape depending on some parameters. It is shown how such technique can be done in computationally efficient way.

2.1. Correlational template matching

In general template matching consists of comparison of input image I with template T in order to find coordinates (x,y) of the best match [1]. In general, any suitable metric M can be chosen as the degree of matching. One of the most practical and common used metric is a sum of squared distances (SSD):

$$M(n,m)=\sum_{i,j}\left(I(i,j)-T(i-n,j-m)\right)^2 . \qquad (1)$$

Straightforward utilization of SSD can be computationally expensive so it is more convenient to use cross-correlation as a kind of fast SSD approximation

$$\sum_{i,j}\left(I(i,j)-T(i-n,j-m)\right)^2 \cong -\sum_{i,j} I(i,j)T(i-n,j-m) \qquad (2)$$

It follows from decomposition:

$$\sum_{i,j}\left(I(i,j)-T(i-n,j-m)\right)^2 = \sum_{i,j}\left(I(i,j)^2 - 2I(i,j)T(i-n,j-m)+T(i-n,j-m)^2\right) \qquad (3)$$

Sum over $T(i-n,j-m)^2$ is an energy of template which is disregarded as a constant. Sum over $I(i,j)^2$ is energy of an input image under the template and is neglected under assumption to be a slow changing term. The sum over $I(i,j)T(i-n,j-m)$ is a definition of a so called cross-correlation.

Correlation has a few desirable properties for template matching task. The main two advantages are its robustness to noise and comparatively low computational cost in frequency domain. Correlation of two functions in spatial domain is a simple inverse transform of product of their Fourier spectrums:

$$f * g = F\overline{G} \qquad (4)$$

where $*$ denotes correlation, F and $\overline{G}$ are Fourier spectrum of f and complex conjugate of Fourier spectrum of g, respectively.

Correlational methods were successfully used in object detection, image registration, image recognition, stereo reconstruction and so on.

The problems arise when one tries to detect objects with complex shape and texture variations. To detect such a complex object it is required to match input image with all possible variations of object shape and texture. To account all of those variations can very computationally complex thus impractical task.

2.2. Dynamical template object detection

Suppose $T(b)$ is a template image, where b is a parameter which responsible for shape and texture variations in template object. Straightforward approach to detect such an object on input image I would be to correlate it with a set of templates

$\{T(i\Delta b) \mid i \in [-M;M]\}$ that covers all possible variations in object appearance. But such number of matchings of input image with different variations of template, in general, is very computationally heavy task.

Under assumption of smoothness computational cost of this task can be trade on accuracy of a method. Let us assume that small changes of parameter vector b cause small changes in correlation picture. The assumption suggests that correlation pictures of two templates that differ on some small Δb with an input image I do not qualitatively dissimilar, but slightly differ only in amplitude, position and width of the correlational peaks.

Based on the smoothness assumption and given a set of correlation pictures $\{C_i\}$ corresponding to the set of templates $\{T(i\Delta b) \mid i \in [-M;M]\}$ we can assume that summation over correlation pictures set $\{C_i\}$ does not changes qualitative picture of cumulative correlogram C^{cum}. Qualitatively steady C^{cum} means that positions $\left(x_j^{max}, y_j^{max}\right)$ of all main peak maximums are not changed. Although the relative values amplitudes of those peaks can be different. Now by using the property of cross-correlation

$$f * (g + h) = f * g + f * h, \tag{5}$$

instead of summation over $\{C_i\}$ we can first sum over all templates $\{T(i\Delta b) \mid i \in [-M;M]\}$ and only than correlate the result with input image

$$\sum_i \{C_i\} = \sum_i I * T(i\Delta b) = I * \left(\sum_i T(i\Delta b) \right). \tag{6}$$

With such approach computational cost of cross-correlation for dynamical template is equal to cross-correlation with regular template. All the computation complexity lies on the creation of sum over a set $\{T(i\Delta b) \mid i \in [-M;M]\}$. Advantage in this case is that sum over $\{T(i\Delta b) \mid i \in [-M;M]\}$ is computed only once during training stage. So detection process per se remains low cost.

2.3. Efficient computation of template sum

The key moment in this approach is to be able efficiently generate sum over a set of template images $\{T(i\Delta b) \mid i \in [-M;M]\}$. Straightforward computation of $\{T(i\Delta b) \mid i \in [-M;M]\}$ for complex objects with multidimensional parameter vector b would be impractical. In case of having proper analytical description for $T(b)$, parameter vector b can be simply integrated out, what is equivalent to summation over $\{T(i\Delta b) \mid i \in [-M;M]\}$ when $\Delta b \to 0$.

In our opinion there are a few state of the art methods most suitable for object image generation. Those are active shape models ASM [2], active appearance models (AAM) [3] and morphable models (MM) [4]. After training, those methods are able to generate modeled object images with intrinsic shape and texture variations.

For computational simplicity here is regarded binary edge images of objects. Usage of binary edge images considerably simplifies computations and also provides certain invariance to brightness changes and lightning conditions.

Active shape models (ASM) were taken as a basis of object edge image modeling. On this stage piecewise linear approximation was used for mathematical description of object edges. ASM are statistical models of shape. They represent the object as a set $\chi = \{x_1, ...x_n, y_1, ...y_n\}$ of key point coordinates. The basic ASM consist of mean shape vector $\overline{\chi}$ and matrix P that holds information on allowed variations and restrictions on shape variation. To produce a new shape ASM uses the following equation

$$\chi = \overline{\chi} + Pb \tag{7}$$

where χ is a key point coordinate set of a new shape and b is a parameter vector of generated shape χ. This paper does not concerned with ASM training and usage so interested readers are referenced to [2] for more details on this subject.

But ASM mathematical model (7) cannot be used as it is to solve our task. The problem is that model takes as an input initial conditions (location and affine group of transformation), parameter b and (i, j) pixel of generated object and as an output give coordinates (x, y) and pixel intensity on output image. To be suitable for this particular task image generation method given the input initial conditions (location and affine group of transformation), parameter b and output image coordinates (x, y) should give as an output pixel intensity of generated object. To do this we integrate it into image resampling method. In our case object generation methods provide two important functions required for resampling. First is $f(x, y, b)$ - texture function, second is $m(u, v, b)$ - coordinate mapping function. Given $f(x, y, b)$ and $m(u, v, b)$ resampling can be written as follows

$$T(x, y, b) = \\ = \iint f(u, v) h\left(x - m_x(u, v, b), y - m_y(u, v, b)\right) du dv \tag{8}$$

where $f(u, v)$ is a mean texture function ($f(u, v) = f(u, v, 0)$), $h(x)$ is a low-pass filter and $m_x(t, b)$ and $m_y(t, b)$ are the mapping functions of x and y coordinates respectively. Not only there is no simple analytical solution to (8) but also numerical computation would be unreasonably expensive for a given task.

ASM provides coordinates of object key points and connecting them lines form a piecewise linear approximation of an object edge image. Given such a piecewise linear approximation integration in (8) from integral over an area reduces to an integral along a edge lines, what also eliminates $f(x, y, b)$ from under the integral (because of a binary edges $f(x, y, b)$ is equal to *1* only along the line of integration and *0* everywhere else)

$$T(x, y, b) = \sum_k \int_0^1 h\left(x - m_x^k(t, b), y - m_y^k(t, b)\right) dt \tag{9}$$

where k is an index of a edge line segment, t is a parameter in parametric representation of lines, $m_x^k(t,b)$ and $m_y^k(t,b)$ are piecewise linear approximations of mapping functions for x and y coordinates respectively. Piece-wise linear approximation for mapping function has the following form

$$\begin{aligned} m_x^k(t,b) &= \bar{x}_k^1 + \left(\bar{x}_k^2 - \bar{x}_k^1\right)t + {}^xP_k^1 b + \left({}^xP_k^2 - {}^xP_k^1\right)bt \\ m_y^k(t,b) &= \bar{y}_k^1 + \left(\bar{y}_k^2 - \bar{y}_k^1\right)t + {}^yP_k^1 b + \left({}^yP_k^2 - {}^yP_k^1\right)bt \end{aligned} \quad (10)$$

where $\left(\bar{x}_k^1, \bar{y}_k^1\right)$ and $\left(\bar{x}_k^2, \bar{y}_k^2\right)$ are the k^{th} line end points, $\left({}^xP_k^1, {}^xP_k^2\right)$ and $\left({}^yP_k^1, {}^yP_k^1\right)$ are the values of mapping parameters in $\left(\bar{x}_k^1, \bar{y}_k^1\right)$ and $\left(\bar{x}_k^2, \bar{y}_k^2\right)$ coordinates.

Fourier transform of a sum over $\{T(i\Delta b) \mid i \in [-M;M]\}$ is now given by

$$F(w,v) = \int_{-\lambda}^{\lambda} \left(\int\int T(x,y) e^{-2\pi i(xw+yv)} dxdy \right) db \quad (11)$$

By substituting (9) into (11) and changing order of integration and summation we get

$$\begin{aligned} F(w,v) &= \\ &= \sum_k \int_0^1 \int_{-\lambda}^{\lambda} F\left\{h\left(x - m_x^k(t,b), y - m_y^k(t,b)\right)\right\}(w,v) dbdt \end{aligned} \quad (12)$$

where $F\{f\}(w,v)$ denotes Fourier transform of a function f. By applying shift property of Fourier transform to (12) we get

$$\begin{aligned} F(w,v) &= \\ &= F\{h\}(w,v) \sum_k \int_0^1 \int_{-\lambda}^{\lambda} e^{-2\pi i\left(wm_x^k(t,b) + vm_y^k(t,b)\right)} dbdt \end{aligned} \quad (13)$$

where $F\{h\}(w,v)$ is a Fourier transform of low pass filter. There is no simple analytical solution to (13). So to make this task practical integration over t is reduced to simple summation.

$$\begin{aligned} F(w,v) &= \\ &= F\{h\}(w,v) \sum_j \int_{-\lambda}^{\lambda} e^{-2\pi i\left(wm_x^k(j,b) + vm_y^k(j,b)\right)} db \end{aligned} \quad (14)$$

where sum over j means summation over all edge points from $f(x,y)$. Basically what this reduction does is it considers $f(x,y)$ to be discrete image of object edges thus makes it more practical.

Now consider in (14) the integral over parameter vector b. Without loss of generality assume b to be one-dimensional vector

$$\begin{aligned} &\int_{-\lambda}^{\lambda} e^{-2\pi i\left(w\left(x_j + P_j^x b\right) + v\left(y_j + P_j^y b\right)\right)} db = \\ &= e^{-2\pi i\left(wx_j + vy_j\right)} \int_{-\lambda}^{\lambda} e^{-2\pi i\left(wP_j^x + vP_j^y\right)b} db \end{aligned} \quad (15)$$

where $\left(x_j, y_j\right)$ is a discrete coordinates of ${}_j{}^{th}$ edge point and P_j^x, P_j^y are values of mapping functions in ${}_j{}^{th}$ edge point for x and y coordinates respectively. Now integration of (15) gives the following

$$\int_{-\lambda}^{\lambda} e^{-2\pi i\left(wP_j^x + vP_j^y\right)b} db = \left. \frac{e^{-2\pi i\left(wP_j^x + vP_j^y\right)b}}{-2\pi i\left(wP_j^x + vP_j^y\right)} \right|_{-\lambda}^{\lambda} \quad (16)$$

As a result of integration we obtain an expression, which actually is nothing else but definition of a $sinc$ function through complex exponent up to constant 2λ.

Thus the final expression for Fourier transform of a sum over $\{T(i\Delta b) \mid i \in [-M;M]\}$ is now given by

$$\begin{aligned} F(w,v) &= \\ &= 2\lambda F\{h\}(w,v) \sum_j e^{-2\pi i\left(wx_j + vy_j\right)} sinc\left(2\pi\left(wP_j^x + vP_j^y\right)\right) \end{aligned} \quad (17)$$

As one can see computation complexity of $F(w,v)$ is not bigger from computational complexity of a standard discrete Fourier transform. $F(w,v)$ should be computed only once on learning stage. So given $F(w,v)$ object detection process is now straightforward and consist of three simple steps: a) computation of input image I Fourier transform F^I; b) multiplication of $F(w,v)$ with complex conjugate of F^I; computation of inverse Fourier transform of $F(w,v)\,\overline{F}^I$. As a result of those steps we obtain correlation picture C. Peaks on this correlational picture denote object of interest most probable locations.

3. Experimental results

Rentgenographic image of pipe weld was chosen as an object of interest. Depending on a relative position of pipe weld to source of x-rays radiation we get different ellipse-like shape images of pipe welds. For tasks of radiographic nondestructive testing it is important to be able to detect position of welds on radiographic images.

Developed approach was tested on synthetic images of pipe welds. Edges of welds were modeled by piecewise linear approximation of key points obtained by ASM training. Model was reduced to consist of only one parameter b. Reasonable variation range for this parameter was $b \in [-2,2]$. For testing, synthetic set of object images with different values of parameter b were generated. Before computation of (4) test image was blurred by gaussian-type filter mask. This is made to achieve more noiseless correlation picture and thus

more steady detection results. Examples of those generated images are gathered in test image shown on Fig. 1.

A number of occlusions in a form of objects with different shapes were added to image (Fig. 2) to complicate the task of object detection. After computation of (17) and (4) as a result we obtain a correlation picture shown at Fig 3. Correlational picture has a complex structure with many correlational peaks. Nevertheless, lots of those peaks can be filtered out by the absolute values of their amplitudes. Correlational peaks

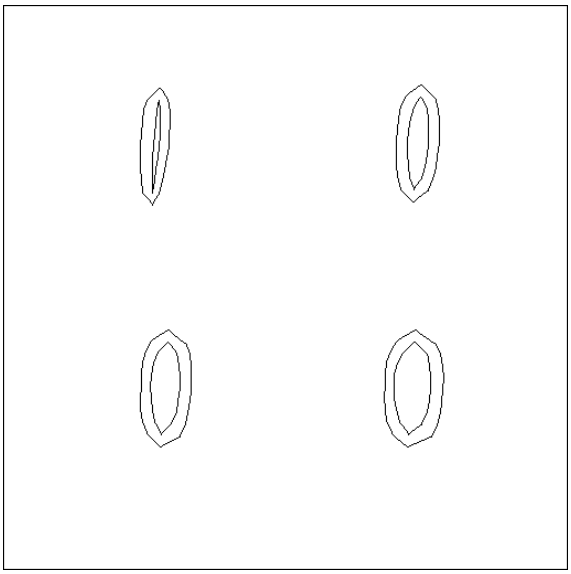

Figure 1: A set of synthetic test images with different parameter b. Top row from left to right: $b = -1.8$, $b = -1$; bottom row from left to right: $b = 1$, $b = 2$.

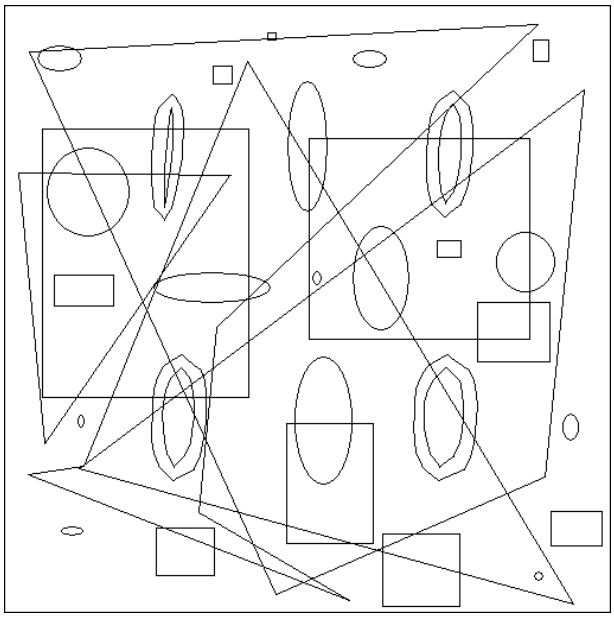

Figure 2: A set of synthetic test images with different parameter b in clutter environment.

for modeled object have considerably bigger values compared to added noise objects.

Three biggest maximum correlational peaks were found. Amplitude values of those peaks were considerably bigger among all the other peaks. This three peak locations with precision up to 93-98% correspond to the true location of modeled object. Even though correlational peak for fourth object (object with $b = 2$) indicates the true location of an input image it has much smaller amplitude than the other three. Such amplitude value for forth object is caused by parameter value b to be on the border of integration limits.

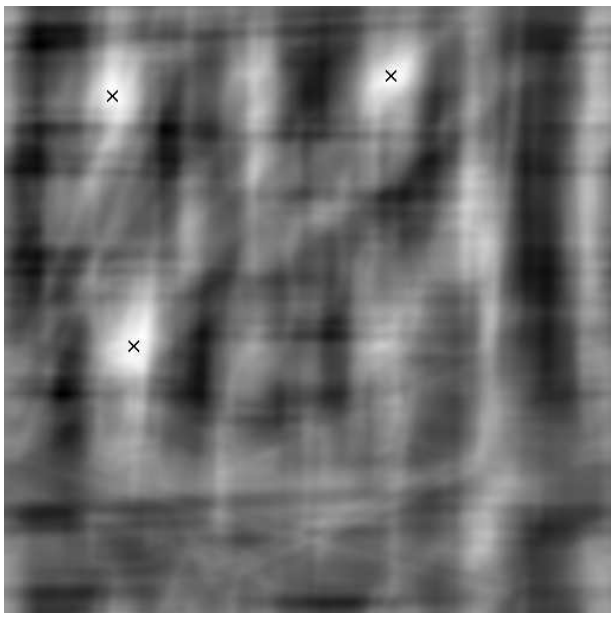

Figure 3: Correlation picture for test image Fig. 2.

Accuracy of a proposed method is satisfactory and can be improved by using more precise approximation used in a method. As experimental results show the proposed method can be successfully used for object detection of dynamical objects. The only requirement for such objects is fulfillment of smoothness assumption.

4. References

[1] R. Brunelli, *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*, 61(1):38–59, Jan. 1995.

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

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