УДК 004.932

T. S. Mandziy

# CORRELATIONAL OBJECT DETECTION BASED ON ACTIVE SHAPE MODELS

New correlation-based approach for object detection is proposed. Method for varying shape object detection is developed. Proposed method showed promising results on synthetic images.

**Keywords**: *object detection, template matching, correlation.*

Запропоновано новий підхід для кореляційного виявлення об’єктів. Розроблено метод виявлення об’єктів змінної форми. Тестування методу на синтезованих зображеннях продемонструвало перспективність методу.

**Ключові слова**: *виявлення об’єктів, порівняння з еталоном, кореляція.*

Object detection is one of the hardest problems in computer vision. It is virtually impossible to extract superior approach from the variety of existing 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 tasks or 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 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 the majority of those methods are unable to deal efficiently enough with variations in shape or texture of an object, otherwise they become very computationally expensive.

This paper deals with template-based paradigm. It considers template to be dynamical (dynamical not with respect to time but to possible shape variation) model of given object image and gives computationally efficient solution to the object detection by matching with such a dynamical template.

**Dynamical template matching.** This section formulates the classical problem of template-based object detection. It also reveals existing problems of this approach related to low tolerance to object distortions. There is proposed an approach for efficient correlation-based 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 its model parameters. It is also shown how such technique can be fulfiled in computationally efficient way.

***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]. Generally speaking, 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):

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

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

$$\sum_{i,j} (I(i,j) - T(i-n, j-m))^2 \cong -\sum_{i,j} I(i,j) T(i-n, j-m) \, . \tag{2}$$

It follows from decomposition:

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

Sum over $T(i–n, j–m)^2$ is energy of the template which is disregarded as it is a constant term. 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:

$$f * g = FG^* , \qquad (4)$$

where $*$ denotes correlation, $F$ and $G^*$ 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, etc.

The problems arise when one tries to detect objects with possible presence of complex shape and texture variations. The detection of such a complex object requires the match of input image with templates of all possible variations of object shape and texture. Consideration of all of those variations can be very computationally complex thus impractical task.

***Dynamical template object detection***. Suppose $T(b)$ is a template image, where $b$ is a parameter vector responsible for shape variations of 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 input image matchings 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 $\Delta b$ 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 correlograme $C^{cum}$. Qualitatively steady $C^{cum}$ means that positions $(x_j^{\max}, y_j^{\max})$ of all main peak maximums are not changed. Although, relative amplitude values of those peaks can be different. Now by using the property of cross-correlation:

$$f * (g + h) = f * g + f * h , \qquad (5)$$

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

$$C^{cum} = \sum_i \{C_i\} = \sum_i I * T(i\Delta b) = I * \left( \sum_i T(i\Delta b) \right) . \qquad (6)$$

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 the detection process per se remains computationally efficient.

***Efficient computation of template sum***. The key moment in this approach is the efficient generation of 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 $b$ is 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 simple analytical description and computational simplicity here are 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. ASM are statistical models of shape. They represent the object as a set $x = \{x_1, ...x_n, y_1, ...y_n\}$ 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

$$x = \overline{x} + Pb, \tag{7}$$

where $x$ is a key point coordinate set of a new shape and $b$ is a parameter vector of generated shape $x$, $\overline{x}$ is a mean shape. This paper does not concerned with ASM training and utilization, so interested readers are referenced to [2] for more details on this subject.

ASM provides coordinates of object key points and connecting them line segments form a piecewise linear approximation of an object edge image. Now the main question is how to analytically represent the image of object given a set of key points $\{x_1, ...x_n, y_1, ...y_n\}$ generated by (7). As it turns out, it is much more convenient to represent the analytical description of a line in frequency domain than it is in spatial domain. Let us assume that $g(x, y)$ is an image of a line segment in spatial domain. Consider the line with a slope $a$ that goes through the origin and is bounded by the spatial limits $x \in [-X, X]$ and $y \in [-Y, Y]$:

$$y = ax. \tag{8}$$

Fourier transform $L(w, v)$ of $g(x, y)$ is the next:

$$L(w,v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) e^{-2\pi i(wx+vy)} dxdy. \tag{9}$$

By making the substitution $x=y/a$ with assumption that $X=Y$, and because $g(x, y)$ is equal to 1 only along the line $y = ax$ and 0 everywhere else the (9) transforms as follows:

$$L(w,v) = \frac{1}{a} \int_{-Y}^{Y} \int_{-Y}^{Y} e^{-2\pi i\left(w\frac{y}{a}+vy\right)} dydy = \frac{2Y}{a} \sin c\left(Y\left(\frac{w}{a}+v\right)\right). \tag{10}$$

The slope $a$ and the integration bound $Y$ for a segment of a line $l$ that connects points $(x_1^l, y_1^l)$ and $(x_2^l, y_2^l)$ are look like the following:

$$a = \frac{(y_2^l - y_1^l)}{(x_2^l - x_1^l)}, \quad Y = y_2^l - y_1^l. \tag{11}$$

Expression (10) generates frequency representations of the line segments with arbitrary slopes and lengths but with restriction that the line goes trough the origin. For translation of the line to a desirable position in spatial domain the translation property of the Fourier transform is used. According to this property the multiplication of $L(w,v)$ by $e^{\pi \mathrm{i}(w(x_1^l+x_2^l)+v(y_1^l+y_2^l))}$ will cause the spatial image to be shifted along $x$ and

$y$ coordinates by $\frac{(x_1^l + x_2^l)}{2}$ and $\frac{(y_1^l + y_2^l)}{2}$ respectively. Thus, substitution of (11) into (10) and application of translation property gives the analytical representation for arbitrary line segment in frequency domain:

$$L(w,v) = 2(x_2^l - x_1^l)\sin c(w(x_2^l - x_1^l) + v(y_2^l - y_1^l))\, e^{\pi i(w(x_1^l + x_2^l) + v(y_1^l + y_2^l))}\,, \qquad (12)$$

where $x_1^l$, $y_1^l$, $x_2^l$ and $y_2^l$ are computed by (7) and take the next form:

$$\begin{cases} x_1^l = \overline{x}_1^l + {}^lP_x^1 b\,, \\ y_1 = \overline{y}_1 + {}^lP_y^1 b\,, \end{cases} \quad \begin{cases} x_2^l = \overline{x}_2^l + {}^lP_x^2 b\,, \\ y_2 = \overline{y}_2 + {}^lP_y^2 b\,. \end{cases} \qquad (13)$$

Based on (12) we can write down the analytical expression for the frequency representation of an object edge image as a superposition of all its line segments:

$$F\{T(b)\}(w,v) = \sum_{l=1}^{n} 2(x_2^l - x_1^l)\sin c(w(x_2^l - x_1^l) + v(y_2^l - y_1^l))\, e^{\pi i(w(x_1^l + x_2^l) + v(y_1^l + y_2^l))}\,. \qquad (14)$$

Taking Fourier transform of (6) and assuming $\Delta b \to 0$ we get frequency representation of cumulative correlogram:

$$F_{C^{cum}} = F_I^* \int_{-\lambda}^{\lambda} F\{T(b)\}db = F_I F_T\,, \qquad (15)$$

where $[-\lambda, \lambda]$ is a range of allowable changes of parameter $b$ and $F_I^*$ is a complex conjugate of an input image Fourier transform.

Expression (14) is not analytically integrable with respect to $b$. The solution to $\int_{-\lambda}^{\lambda} F\{T(b)\}db$ can be reduced to computation of exponential integral functions. Unfortunately the size of the analytical representation of the integral in terms of exponential integral function is inconvenient to be published in present paper. But the original solution can be precisely replicated by using any symbolic integration packages for integration of (15) with respect to $b$.

Computed once $F_T$ than can be used for detection of modeled object in arbitrary input image. So given $F_T$ object detection process is now straightforward and consist of input image $I$ Fourier transform $F_I$ computation, multiplication of precomputed $F_T$ with $F_I^*$ and computation of inverse Fourier transform of $F^{-1}\{F_{C^{cum}}\}$. Final stage of detection procedure consist in correlational peaks location on $F^{-1}\{F_{C^{cum}}\}$. Peaks on this correlational picture denote object of interest most probable locations.

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

Developed approach was tested on synthetic images containing generated pipe welds-like shapes. 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 $b$. Practically 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 $F_I$ input 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*.

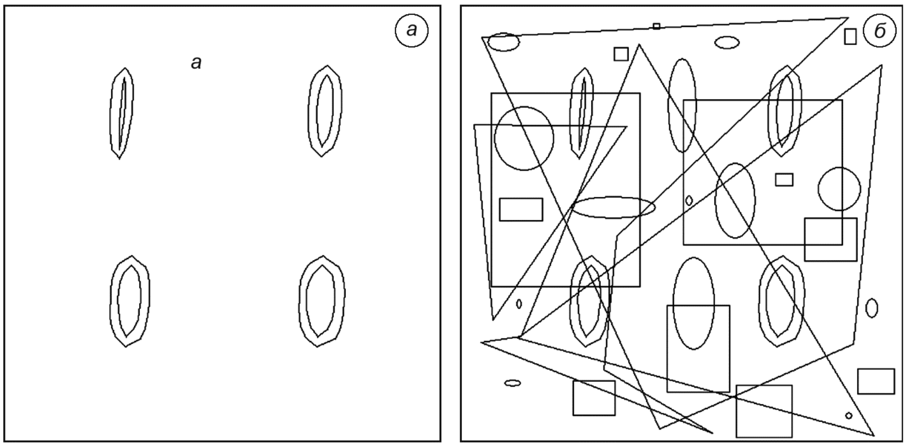

Fig. 1. *a* – a set of synthetic test images with different parameter *b* (top row from left to right: *b* = –1.8, *b* = –1, bottom row from left to right: *b* = 1, *b* = 2); *b* – a set of synthetic test images with different parameter *b* in clutter environment.

A number of occlusions in a form of objects with different shapes were added to original input image (Fig. 1*a*) to complicate the task of object detection. The cluttered version of input image is shown on Fig. 1*b*.

After computation of (15) and taking inverse Fourier transform of the result we obtain a correlation picture shown at Fig 2. 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 corresponding to true location of the modeled object have considerably bigger values compared to added noise objects.

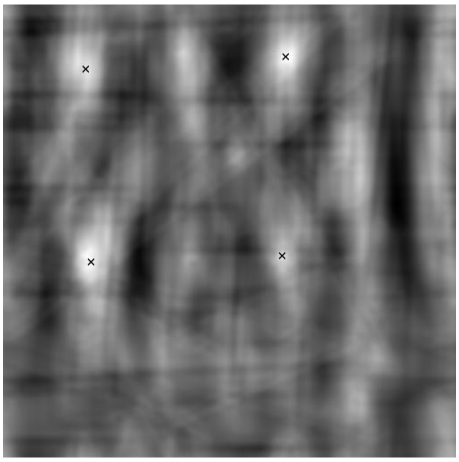

Fig. 2. Correlation picture for test image Fig. 1*b*.

Four biggest maximum correlational peaks were found. These peak locations with precision up to 93...98% correspond to the true locations of modeled object in input image.

Accuracy of the proposed method is satisfactory. As experimental results show the proposed method can be successfully used for detection of objects with varying shapes. The only requirement for such objects is the fulfillment of smoothness assumption.

1. *Brunelli R.* Template Matching Techniques in Computer Vision: Theory and Practice. – Wiley, 2009.
2. *Active* shape models – their training and application / T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham // 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 / Eds. H. Burkhardt and B. Neumann // Proc. Fifth European Conf. Computer Vision. – 1998. – Vol. 2. – P. 484–498.
4. *Blanz V., Vetter T.* A Morphable Model for the Synthesis of 3D Faces // SIGGRAPH'99 Conf. Proc. – 1999. – P. 187–194.

*Фізико-механічний інститут ім. Г. В. Карпенка*
*НАН України, Львів*

*Одержано*
*23.09.2010*