

USING MORPHLET-BASED IMAGE REPRESENTATION FOR OBJECT DETECTION

V.S. Gorbatsevich, Yu. V. Vizilter

State Research Institute of Aviation Systems (GosNIIAS),
125319, 7, Viktorenko str., Moscow, Russia – (viz,gvs)@gosniias.ru

ICWG III/VII

KEY WORDS: Mathematical morphology, object detection, Wavelets, Wavelet analysis, object segmentation

ABSTRACT:

In this paper, we propose an original method for objects detection based on a special tree-structured image representation – the trees of morphlets. The method provides robust detection of various types of objects in an image without employing a machine learning procedure. Along with a bounding box creation on a detection step, the method makes pre-segmentation, which can be further used for recognition purposes. Another important feature of the proposed approach is that there are no needs to use a running window as well as a features pyramid in order to detect the objects of different sizes.

1. INTRODUCTION

1.1 Selective object detection algorithms

Nowadays the problem of selective object detection in an image is getting an increasing attention. Such detection is highly demanded for optimization of recognition algorithms based on Deep Learning, which provides high quality results but poses significant computational load. The optimization can be achieved by making a preliminary hypothesis about objects location using some fast methods and then applying complex, time-consuming recognition algorithms only for the hypothesis approval. Selective object detection algorithms can be divided into two groups: machine learning and non-machine learning-based algorithms. The first one group requires a learning dataset with the large amount of images and ground-truth answers. The second one usually employs empirical assumptions about general features of the objects of interest.

This paper proposes a new algorithm for selective object detection based on morphological image analysis methods. The developed algorithm uses image descriptions and wavelet-based analysis proposed in (Vizilter 2015) and morphological methods developed by Yu.P. Pyt'ev(Pyt'ev 1993). The algorithm is non-machine learning-based and uses trees of morphlets descriptors. The problem of selective object detection is reduced to the search of a sub-tree with a given properties in the tree of morphlets for a given image. It gives possibility to perform multi-scale object detection without using running window search or features pyramid generation. The other benefit of the proposed method is a direct extraction of accurate boundaries of object of interest.

The algorithm was evaluated using PASCAL VOC 2007 dataset. The evaluation proved high efficiency of the algorithm compared to the other selective object detection algorithms.

1.2 Related work

A large number of algorithms for object detection were developed up to the date. Boosting methods are widely used for classifier creation since classical work by Viola-Jones. Haar-

like features and their modifications (Viola 2001) can be used for effective boosting, as well as more complex features like SURF(Jianguo 2013), HOG(Dalal 2005), etc.

The forest of decision rules trees, generated by features of special types, are proposed in (Schulter 2014). On the other hand, in (Sudhakar 2015) a deep convolutional network is used directly to search objects(faces) in the image. The algorithm is considered as the state-of-the-art for face detection, but due to high computational complexity, it cannot be used in real-time.

In (Li 2015) the problem of object(face) detection is solved by a cascade of the deep convolutional networks. The complexity of the networks topology increases for higher levels of the cascade.

A interesting approach is used in (Karianakis 2015), however a boosting is applied to the filters of the first level of the deep convolutional network.

A different way of capturing the closed boundary characteristic of objects rests on using superpixels as features is used in (Felzenszwalb 2004).

EdgeBox (Zitnick 2015) method is based on empirical rules, generated using maps of oriented gradients for a running window.

In the contrast to methods mentioned above, the presented method is based on image description using tree of morphlets proposed in (Vizilter 2015).

2. PYT'EV MORPHOLOGICAL METHODS

2.1 Image tessellation, Pyt'ev morphology and image-shape comparison

A brief description of Pyt'ev morphology (Pyt'ev 1993) is presented below. Let $f(x,y)$ be a function of image intensity:

$$f(x,y): \Omega \rightarrow R, \Omega \subset R^2, \quad (1)$$

where

Ω - region of an image,
 R - set of real numbers,
 R^2 - image plane.

Images are considered as elements of a Hilbert space $L^2(\Omega)$. Using such notation, we can define a morphological description of image shape. In the base Pyt'ev morphology images are treated as piecewise continuous functions:

$$f(x,y) = \sum_{i=1,\dots,n} f_i \chi_{F_i}(x,y), \quad (2)$$

where

n - number of regions in frame tessellation \mathbf{F} of frame Ω into connected non-overlapped regions of constant intensity;

$\mathbf{F}=\{F_1,\dots,F_n\}$; $\mathbf{f}=(f_1,\dots,f_n)$ - vector of scalar intensity values of corresponding image regions;

$\chi_{F_i}(x,y) \in \{0,1\}$ - support function of i -th intensity region:

$$\chi_{F_i}(x,y) = \{1, \text{ if } (x,y) \in F_i; 0 - \text{ otherwise}\}. \quad (3)$$

Hence a set of images of similar shapes for tessellation \mathbf{F} is a convex and closed subspace $F \subseteq L^2(\Omega)$:

$$F = \{f(x,y) = \sum_{i=1,\dots,n} f_i \chi_{F_i}(x,y), \mathbf{f}_F \in R^n\}. \quad (4)$$

In practice methods of Pyt'ev morphology are usually applied to intensity centred images. To centre the image a constant mean value is subtracted from each pixel. Hence shapes of centred images are considered.

2.2 Morphology of localized shapes

The main difficulty that arises with the application of Pyt'ev morphology methods to the problem of object detection is that they are not designed for comparison of images captured with different field of view (different frames of different supports). However objects of interest are usually captured using different sensors with different frame size. Hence, it is possible to make a transition from classical Pyt'ev morphology to morphology of localized shapes presented in (Vizilter 2015). It can be done by introducing the concepts of localized shapes and shape supports.

Let us call a support of an image f (two-dimensional function) a set of points on a plane at which the function f is not equal to null:

$$\Omega_f = \Omega(f) = \{(x,y) \in R^2: f(x,y) \neq 0\}. \quad (5)$$

Then a support of shape F is a maximal (by inclusion) support of images covered by that shape. It is obvious that supports of image tessellation \mathbf{F} of non-centered form F and centered form F coincide:

$$\Omega_{\mathbf{F}} = \Omega_F = \Omega_{F^*}. \quad (6)$$

Let f and g be images with localizations Ω_f и Ω_g correspondingly. Let $P_\Omega g(x,y)$ be called a projection of image on a support:

$$P_\Omega g(x,y) = \{g(x,y), \text{ if } (x,y) \in \Omega \cap \Omega_g; 0, \text{ if } (x,y) \in (\Omega \cap \Omega_g^c \setminus \Omega)\} \quad (7)$$

The projector P_Ω can be considered as a localization operator of a function on the support Ω . Then *the projection of any localized image onto the shape of other localized image* could be defined as a combination of a projection onto the shape and a projection onto a support:

$$g_{\#}(x,y) = P_{\#} P_{\Omega_{\#}} g(x,y) = \{ \sum_{i=1,\dots,k} g_{\#i} \chi_{\Omega_{\#i}}(x,y), \text{ if } (x,y) \in \Omega_{\#} \cap \Omega_g; 0, \text{ if } (x,y) \in (\Omega_{\#} \cap \Omega_g^c \setminus \Omega_{\#}) \}, \quad (8)$$

where $\Omega_{\#} \cap \Omega_g \setminus \Omega_{\#}$ - a complement $\Omega_{\#} \cap \Omega_g$ to $\Omega_{\#}$. If a localization of the functions coincide, such definition of the projection is equal to the definition in the classical Pyt'ev morphology. If localizations are different, the projection of shape F is obviously belongs to shape F and has a formal support $\Omega_{\#}$. In the region of intersection of supports, the projection is calculated as defined above. In the region where function g doesn't have any information, the projection is supplemented with zeroes.

An important property of such description of a shape is that any shape could be considered as a direct sum of its constituent localized shapes. On the other hand a concept of hierarchy of localized shapes could be defined:

A localized shape \mathbf{G} with a support Ω_G is called a *child shape* (a *subshape*) of a shape $\mathbf{F}=\{F_1,\dots,F_n\}$ with a support Ω_F , if $\exists i: \Omega_G = F_i$.

The shape \mathbf{F} with respect to the shape \mathbf{G} is called a *parent shape*.

Such hierarchy of shape could be always represented as a tree graph. Let us define such graph as a *tree of localized shapes*. Nodes of such tree represent localized shapes from the hierarchy. Edges of the tree connect only parent shapes and child shapes.

Using such approach it is possible to define the problem of object detection on a given image using morphological methods. The object detection problem became equivalent to a search of a subtree in a tree of localized shapes. Such subtree represents a shape with a given properties.

3. MORPHLETS

In (Vizilter 2015) a new class of tree-structured morphological descriptors of image shape ("tree-based morphlets") was proposed. Morphlets are considered as Haar-like wavelets - the system of differential operators, satisfying the following requirements:

- Wavelets are two-dimensional piecewise constant functions with zero mean;
- Supports of wavelets of smaller scale always belong to the regions of constant values of larger scale wavelets;
- System of Haar-like wavelets corresponds to the tree of image tessellations (mosaic shapes).

Such Haar-like wavelets, in contrast to the classical Haar wavelets, do not have a "parent" wavelet, the property that imposes no constraints on the shape of each wavelet. On the other hand, Haar wavelets are obviously can be treated as Haar-like wavelets.

Due to the absence of constraints imposed by the "parent" wavelet, shape can be described as a system of Haar-like wavelets, and, as it was shown in (Vizilter 2015), the tree of

localized mosaic shape corresponds to the tree of Haar-like wavelets. Trees of Haar-like wavelets, which shape of local basis on each level can be different and determined adaptively by considering a brightness distribution of some image or group of images for the support of the current tree level, are called trees of morphlets, whereas Haar-like wavelets, which compose them, are called morphlets.

Thus, the discussed above methodology of object detection as search for the subtree in the tree of localized shapes can be applied to trees of morphlets.

4. MORPHLETS BASED ON HISTOGRAM SEGMENTATION (OTSU-TREES)

In this paper, we use morphlets, which based on histogram segmentation by Otsu method. The basic idea is to use the Otsu binarization to create receptive fields of morphlets. The algorithm for constructing Otsu-tree is the following:

Input data:

I – image

Output data:

T – tree-structured description

Variables:

T – tree of morphlets

$node = \{node.\Omega, node.\Omega_P, node.\Omega_N\}$ – morphlet

$node.\Omega$ – support of morphlet

$node.\Omega_P$ – positive receptive field

$node.\Omega_N$ – negative receptive field

Parameters:

S_{thr} – minimal square of a segment

Initialization:

$T = \{\{I, 0, 0\}\}$ – add an element $node.\Omega = I$, $x.\Omega_P = 0$, $x.\Omega_N = 0$ to T as a root

Steps:

Step 1. Select a new element $x \in T$: $x.\Omega_P = x.\Omega_N = 0$

Step 2. Create binary image using Otsu threshold $x.\Omega$ and retrieve connected regions: $\Omega \rightarrow \mathbf{B}; \{\mathbf{B}^+, \mathbf{B}^-\}$, where \mathbf{B}^+ – the set of connected regions corresponding to “white” segmentation (above the threshold), \mathbf{B}^- – to “black” segmentation (below the threshold).

Step 3. Create the current morphlet x :

$$x.\Omega_P = \bigcup_{B_i^+ \in \mathbf{B}^+} B_i^+$$

$$x.\Omega_N = \bigcup_{B_i^- \in \mathbf{B}^-} B_i^-$$

Step 4. For the each connected area $b \in \mathbf{B}$, such that its square $S(b) > S_{thr}$, create a new element n : $n.\Omega = b$, $n.\Omega_P = n.\Omega_N = 0$ and add this element in the tree T to the element x .

Step 5. Proceed to Step 1.

The proposed algorithm allows one to create a morphlet description for an arbitrary greyscale image. In the same way, it is possible to synthesize an algorithm for creation of morphlet description based on any other histogram binarization method.

5. OBJECT DETECTION

The problem of object detection is reduced to the problem of a subtree search in a tree of morphlets with specific characteristics (examples of detected objects are shown on Fig. 1). Such characteristics can be the following:

- Minimal and maximal complexity – the number of levels in subtree.
- Minimal and maximal squares of a root morphlet support.
- Geometrical parameters of the root morphlet support, such that the direction of the principal axes, aspect ratio of a bounding box, etc.

Unlike the traditional methods of objects detection that use a running window and an image pyramid, the proposed approach gives possibility of finding objects of different sizes as well as with complex shapes (such as roads, bridges, extended objects, and so on).



Figure 1. Using the tree of morphlets approach for objects detection in images

It is necessary to mention that the support of the root morphlet for the corresponding subtree also contains information about object shape, and, therefore, can be used for segmentation and bounds retrieving (see Fig. 2). This property makes the method much more effective in case of objects with non-rectangular shape.

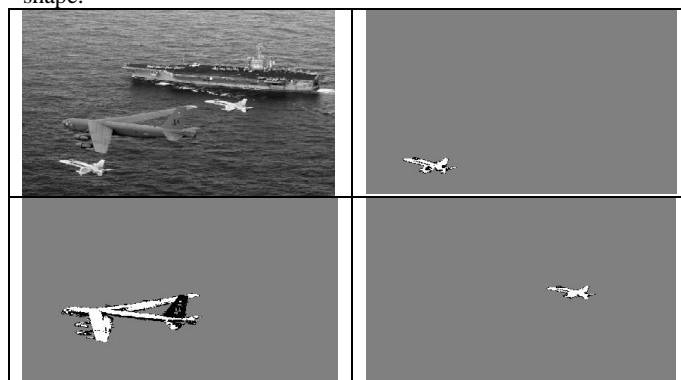


Figure 2. Examples of image segmentation.

6. EXPERIMENTS

Experiments were performed using Pascal VOC 2007 image datasets with the following parameters:

- Minimal and maximal complexity: 4 – 100
- Maximal and minimal squares of a root morphlet support: 0.95 — 0.05;
- Geometrical parameters of the root morphlet support, such that the direction of the principal axes, aspect ratio of a bounding box are not considered.

The table below shows the results of the experiments for the proposed method and for popular methods of the same type (object detection without learning algorithm).

Number of windows	2	10	27	100
Proposed method	0.15	0.44	0.57	0.68
OD(Dalal 2005)		0.19		0.44
MS(X. Hou 2007)		0.31		0.6
SS(Felzenszwalb 2004+ Alexe 2012)		0.35		0.68
MS+CC+SS(Alexe 2012)		0.41		0.71
Edge boxes 70 (Zitnick 2015)				0.4
Edge boxes 50 (Zitnick 2015)				0.76
Edge boxes 90 (Zitnick 2015)				0.64

Table 1. The results of the experiments

It can be seen from the table 1 that in case of small number of windows the proposed method provides the good results in the group(even in case of only 2 windows), and even in case of the number of windows equal to 100, concedes only to MS+CC+SS and Edge box 50 method. It should be noted that the proposed method does not employ any color information and any shape hypothesizes (like rectangle in case of sliding window).

7. CONCLUSIONS

To sum up, the proposed method for selective objects detection based on trees of morphlets provides a robust search of objects even in case of minimal prior knowledge about their type and shape. The method is convenient to use for the search of large objects of complex shape in various scales. It is necessary to mention that the method does not employ a running window and image or feature pyramids. This property provides the robust search of objects of complex shape as well as possibility of testing hypotheses about their size and shape. Additionally, the method does not require the analysis of color information, which can also be useful in some tasks.

REFERENCES

Vizilter, Yu., 2015. Morphlets: A new class of tree-structured morphological descriptors of image shape. *Computer Optics Journal*, 39(1), pp. 101-108

Pyt'ev, Yu., 1993. Morphological Image Analysis. *Pattern Recognition and Image Analysis*, 1993, 3(1), pp. 19–28.

Felzenszwalb, P., 2010. Object detection with discriminatively trained part based models. In: *IEEE TPAMI* 2010, 32(9), pp. 1627–1645.

Hou, X., 2007. Saliency Detection: A spectral residual approach. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, Minnesota, USA, pp. 1-8.

Alexe, B., 2012. Measuring the objectness of image windows. In: *IEEE TPAMI* 2012, 34(11) pp. 2189-2202.

Viola, P., 2001. Rapid object detection using a boosted cascade of simple features. In: *IEEE Conference on Computer vision and pattern recognition(CVPR)*, Kauai, HI, USA, pp. 511-518.

Jianguo, L., 2013. Learning SURF Cascade for Fast and Accurate Object Detection. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, Portland, Oregon, USA, pp. 3468-3475.

Dalal, N., 2005. Histogram of Oriented Gradients for Human Detection. In: *IEEE Conference on Computer vision and pattern recognition (CVPR) 2005*, Fort Collins, Colorado, USA, pp. 886-893.

Haoxiang, L., 2015. A Convolutional Neural Network Cascade for Face Detection. In: *IEEE Conference on Computer vision and pattern recognition (CVPR) 2015*, San Francisco, CA, USA, pp. 5325-5334.

Karianakis, N., 2015. Boosting Convolutional Features for Robust Object Proposals. In: *IEEE Conference on Computer vision and pattern recognition (CVPR) 2015*, San Francisco, CA, USA, , pp. 2524-2532.

Felzenszwalb, P., 2004. Efficient graph-based image segmentation. *International Journal of Computer Vision* 2004, 59(2), pp. 167–181.

Zitnick, C., 2015. Edge boxes: Locating object proposals from edges. In: *Proc European Conference on Computer Vision (ECCV) 2015*, Zurich, Switzerland, pp. 391-405.

Schulter, S., 2014. Accurate Object Detection with Joint Classification-Regression Random Forests. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2014*, Columbus, Ohio, USA, pp. 923-930.

Farfadi, S. S., 2015. Multi-view Face Detection Using Deep Convolutional Neural Networks. In: *Proc. International Conference on Multimedia Retrieval (ICMR) 2015*, Shanghai, China, pp.643-650.